NAGAOKA UNIVERSITY OF TECHNOLOGY

DOCTORAL THESIS

Development of Mathematical Models for Allocation and Routing Problem of the Shared Taxi with Heuristic Algorithm

（ヒューリスティック・アルゴリズムを用いた乗合タクシーの配車・経路決定問題の数理モデルの開発）

Manik Mondal

Supervisor: Professor Kazushi Sano

A dissertation submitted in fulfillment of requirements for the degree of Doctor of Engineering

in

Department of Civil and Environmental Engineering

August – 2023

Development of Mathematical Models for Allocation and Routing Problem of the Shared Taxi with Heuristic Algorithm

A dissertation presented

by

Manik Mondal

to

Department of Civil and Environmental Engineering,

in partial fulfillment of the requirements

for the degree of Doctor of Energy and Environment Science

Supervised by **Professor Kazushi Sano**

Examination committees:	Prof. Kazushi Sano
	Prof. Hiroyuki Oneyama
	Dr. Yoko Matsuda
	Dr. Toshiya Matsukawa
	Dr. Teppei Kato
Nationality:	Bangladesh
Scholarship Donor:	Monbukagakusho (MEXT) scholarship, Government of Japan

Nagaoka University of Technology
Niigata, Japan - August 2023

‘‘The function of equation is to teach one to think intensively and to think critically. Intelligence plus character - that is the goal of true equation.’’

- *Martin Luther King*

"When once you have tasted flight, you will forever walk the earth with your eyes turned skyward, for there you have been, and there you will always long to return."

-Leonardo da Vinci

Abstract

Taxis are an essential means of transportation for addressing the movement requirements of the citizens of the local area. Additionally, effective transportation infrastructure is required to lower CO2 emissions in the context of climate change. Inefficient taxi services result in increased idle time as well as increased CO2 emissions. Therefore, optimizing taxi allocation can help reduce CO2 emissions and improve the sustainability of suburban transportation. The impact of the taxi market on daily life is significant, both in terms of taxi driver's salaries and social issues (e.g., travel demand satisfaction, and traffic congestion). The literature review conducted for this thesis revealed that shared mobility has become increasingly popular in recent years. Shared taxis have emerged as an important component of shared mobility, providing a cost-effective means of transportation for passengers. So, the efficient management of a fleet of shared taxis is challenging, particularly in terms of taxi allocations and routing cost optimization. Several studies have focused on developing mathematical models for fleet management and optimization, including the use of heuristic algorithms. The main objective of this thesis is to develop a mathematical model which would be used for shared taxi routing problems and taxi allocation. The shared taxi problem model deals with passenger demand, which is to find the shortest route of taxi routing decisions that can minimize the customers' total trip cost. On the other hand, the objective of the taxi allocation model is to minimize taxi idle time costs. This study was intended to maximize taxi allocation based on idle time costs, pickup delays, and passenger waiting time while minimizing CO2 emissions.

Firstly, this study focuses on taxi demand prediction for areas in Nagaoka City using Machine Learning Algorithms. To do so, several ride-related parameters have been studied such as the month of the year, day of the week, and time of the day. Moreover, weather-related variables, such as temperature, and weather conditions are used as predictors to estimate the demand level for taxis in urban areas. A suggested approach for predicting taxi demand in suburban areas involves the use of four well-known machine learning algorithms: Linear Regression, Decision Tree, Random Forests, and Gradient Boosting. Additionally, a Hybrid Machine Learning method has been proposed for this purpose.

Secondly, this thesis formulates a nonlinear shared taxi routing model for the riders considering the passengers' benefits and drivers. The model explores the concept of a shared taxi to minimize the total cost for passengers and the number of taxis utilized, while maximizing

customer satisfaction thereby increasing the taxi demand and driver's income. The problem is solved using the Branch and Cut algorithm employed for estimating the parameters. The branch and cut methodology are mathematical optimization techniques that are commonly used to solve combinatorial optimization problems. It is based on two algorithms, branch and bound and cutting planes. The branch and bound algorithm divide the problem into smaller subproblems, while the cutting plane algorithm adds additional constraints to eliminate suboptimal solutions. Together, these algorithms have been used to find the global optimum solution for a given problem. The computational results obtained by the Gurobi solver showed that the proposed mathematical model can find the shortest route of taxi routing decisions that can minimize the customers' total cost. In addition, it demonstrated the effect of various taxi-sharing practices, e.g., waiting time and fare decrease. Furthermore, the sensitivity of significant parameters such as time window, initial fare, and travel distance in the fixed fare is analyzed. Real taxi valid data of Sanjo city, Japan is used for this study.

Lastly, this study proposes a mathematical model of the taxi allocation optimization problem for minimizing taxi driver idle time costs, which helps to reduce CO2 emissions in suburban contexts. For solving the proposed model, we used instead of proposing three heuristic algorithms: greedy algorithm, Simulated annealing algorithm, and dynamic greedy algorithm. Further, a case study has been carried out to demonstrate the proposed methods and show its implication by using real taxi data of Nagaoka, Japan. For the case study, we first identified the taxi travel hotspots as potential locations for taxi spots by investigating the pickup and drop-off locations and times by analyzing GPS data of taxi. After that, we found the optimal taxi allocations by applying the proposed model and solution algorithms. To summarize, dynamic greedy heuristics successfully obtained excellent solutions in relatively short runtimes for the taxi issue, making strategic decisions and feasible choices for taxi markets to adopt. The case study application in this study demonstrates the potential for using data analysis, optimization of taxi allocations and the number of taxis, and to reduce approximately 81.84% of CO2 emissions in the transportation sector. Finally, sensitivity analysis was applied to validate the model for passenger demand. The sensitivity analysis showed that the total idle time cost increased if the demand increased. By strategically placing taxi spots and optimizing their routes, cities can reduce the number of vehicles on the road, decreasing harmful emissions and improving air quality. Therefore, this study suggests that it is essential for cities to invest in and prioritize taxi location optimization as a crucial step toward achieving sustainable transportation.

Keywords:

Taxi Demand Predication; Machine Leaning; Taxi Allocation; Shared Taxi Routing Model; Heuristic Algorithm; CO2 Emissions.

Acknowledgments

I would like to express my heartfelt gratitude to my supervisor, **Prof. Kazushi SANO**. His full support enabled me to work on my favorite topic under a supportive research environment.

I also appreciate his warm encouragement and critical comments, which helped me overcome many difficulties and anxieties during my research activities. In addition, his trust in me was a great motivation for me to finish my doctor's study in five years.

I would like to send my sincere appreciation to the committee members for their valuable feedback and suggestions to fulfill this dissertation. Many thanks are reserved for all colleagues in Urban Transport Engineering & Planning Laboratory at **Nagaoka University of Technology.**

I would like to convey my sincere appreciation to the members of my committee, Prof. Hiroyuki Oneyama, Dr. Yoko Matsuda, Dr. Toshiya Matsukawa, and Dr. Teppei Kato, for their time and expertise in reviewing the dissertation and providing useful feedback throughout the entire dissertation process.

Especially, sincere thanks to Dr. Chathura De Silva, Dr. Trinh Thanh Linh, Dr. Uditha, and Mr. Frank for his valuable comments and discussion on my research.

Dr. Chonnipa Puppateravanit (Rainny), thank you for always having your best smile when I was working on my thesis until late at night or during the weekends.

I would like also to thank the taxi company for the data provided, especially to Mr. Watari and Mr. Wang.

I would also like to express my sincere gratitude to the Japanese Government for awarding me with the Mobukagakusho scholarship for three years. Without this financial support, it would be impossible for me to give full commitment to research and complete it successfully.

Last but not least, I would like to express sincere thanks to my family for infinite motivation and unlimited support.

Abbreviation and Acronym

Terms	**Full words**
HML	**Hybrid Machine Learning**
GPS	**Global Positioning System**
SA	**Simulated Annealing**
LR	**Linear Regression**
RF	**Random Forest**
DT	**Decision Trees**
GB	**Gradient Boost**
KNN	**K-nearest Neighbors**
DARP	**Dial-a-ride-problem**
PDPTW	**Pickup and Delivery Problem with Time Windows**
MAPE	**Mean Absolute and Percentage Error**
RMSE	**Root Mean Squared Error**
MAE	**Mean Absolute Error**
MTDL	**Multi-Task Deep Learning**
ML	**Machine Learning**
LSTM	**Long Short-Term Memory**
MdAE	**Median Absolute Error**

Contents

Figures

Tables

Chapter 1. INTRODUCTION

One of the main challenges facing the taxi industry in Japan is the aging population of taxi drivers. Many taxi drivers in Japan are older, and there has been a decline in the number of young people entering the industry. Nowadays, most people are concentrated in urban areas. This has led to concerns about a potential shortage of taxi drivers in the future. According to research, more than 80% of the world's population will live in cities by 2030, (Lahariya (2008)). As a result, the number of vehicles, travel and upward mobility within cities will increase significantly. Therefore, the current capacity of a city's road network will confront operational issues that need to be addressed to encourage increased vehicular traffic in the urban environment. This chapter of the essay covers the background, research need, research questions, motivation, objective, and framework of the study.

1.1 Background

At present, taxis are an important part of supporting local public transport in major cities. Other public transport services such as trams, commuter rail, and urban buses (known as 'mass transit' systems) have some storage in some features. The general features of taxi services are door-to-door operation, 24-hour-a-day service, and single-passenger hire. But the more time a taxi driver spends seeking a new customer, the more fuel is consumed, and fewer people are transported. Inexperienced taxi drivers frequently don't know where to pick up new passengers since they are acquainted with taxi overtime and space limits. Future taxi demand information can help both new and experienced drivers satisfy the city's taxi demand promptly. Taxi demand prediction in a particular location can be used to create new business opportunities or new services. That’s why, passenger location is one of the most important to all taxi drivers, Vanichrujee (2017).

However, taxis are available in all sections of Japan, with the exception of isolated places, and are nearly entirely provided by private taxi firms and owner-driver cabs. There are around 6,400 taxi firms and 37,000 owner-driver cabs in the city, with a total fleet of over 230,000 cars. The overall number of taxi passengers is expected to be around 15 billion every year. According to Fukumoto *et al.* (2017), the market size of taxis in Japan is over 1.7 billion yen (JPY), which is around 60% more than the market size in 1992. Because the number of passengers is now declining, the Japanese taxi sector is experiencing several issues. The Japanese govt. and taxi companies are working together to address these fears, which are especially acute in regional cities and rural regions. The appropriate charge for shared taxis and taxi allocation in urban and suburban regions are presented in this thesis. As a result, it is required to provide a proposal that will help to revitalize the taxi business in the area.

There are two types of taxi businesses such as regular/general taxis and sharing taxis in Japan. Regular taxis in Japan are typically painted in a distinctive color, such as black or white, and are equipped with a meter to calculate the fare based on distance and time. A taxi transports passengers to and from their desired location. Regular taxis in Japan have a long history, dating back to the early 20th century. Initially, taxis were primarily used by wealthy individuals, but as the industry grew, they became more widely used by the public. Today, taxis are a common mode of transportation in Japan, particularly in urban areas where public transportation may be crowded or infrequent. Taxi drivers in Japan are known for their high level of professionalism and customer service and are expected to be knowledgeable about the city and provide helpful recommendations to passengers. Shared taxis, also known as "jumbo taxis" or "shuttle taxis", have been in operation in Japan since the 1970s. These taxis are larger than general taxis and can carry up to ten passengers. Shared taxis operate on fixed routes, picking up and dropping off passengers at designated locations. The introduction of shared taxis was a response to the high demand for public transportation in rural areas of Japan, where traditional buses and trains were often infrequent or unavailable. Shared taxis provide a more flexible and convenient option for passengers traveling to and from remote locations. Shared taxis have become increasingly popular in recent years, particularly in rural areas where public transportation options are limited. The Japanese government has also promoted the use of shared taxis as a means of reducing traffic congestion and improving environmental sustainability. However, regular taxis remain an important part of the transportation landscape in Japan, particularly in urban areas. One of the main challenges facing the shared taxi industry in Japan is the high cost of operation. Additionally, shared taxis require a larger vehicle and driver, which can increase operational costs. Another issue facing the shared taxi industry in Japan is the lack of public awareness and understanding of the service. Many people in Japan are not familiar with the concept of shared taxis and may not know how to use them, which can limit the demand for the service. Furthermore, the availability of shared taxis can be limited in some areas, particularly outside of major urban centers. This can make it difficult for passengers to find a shared taxi when they need one, particularly during peak travel times. In Japan, shared taxis have the advantage of being more cost-effective than general taxis for certain types of trips. By sharing the ride with other passengers, the fare per person can be lower than the fare for a general taxi ride. This can make shared taxis an attractive option for passengers traveling longer distances, particularly in rural areas where public transportation options are limited.

Another benefit of shared taxis is that they can be more convenient than traditional public transportation options such as buses and trains. Shared taxis operate on fixed routes and schedules, which can provide a reliable and flexible option for passengers who need to travel to and from remote locations or at off-peak times. The fare structure for shared taxis in Japan can also be a challenge. While shared taxis may be more cost-effective than general taxis for certain types of trips, the fare is usually fixed and can be higher than the fare for a similar trip on a bus or train. This can make shared taxis less attractive to price-sensitive passengers who are willing to sacrifice convenience for lower fares. Furthermore, the fare structure for shared taxis can be complicated, particularly for passengers who are not familiar with the system. In some cases, the fare may be based on the number of passengers, the distance traveled, or a combination of both. This can make it difficult for passengers to understand how much they will need to pay for the ride. At present, a shared taxi is becoming a travel mode. According to Chen *et al.* (2015), passengers and drivers agree that sharing a taxi with others is good for the environment, traffic congestion, and the participants. Taxi sharing can help the environment by reducing the usage of fuel and CO_2 emissions. For society, sharing taxis can meet the growing travel demand. Passengers may save money on their travel expenses, and it is a win-win situation. Drivers can make more money when there are more passengers in the cab, according to Huang *et al.* (2018). The price systems for taxi sharing are varied, and the majority of them are only appropriate in certain circumstances. The authors looked at four different taxi-sharing scenarios: (a) Passengers have the same origin and destination, (b) the same origin and different destinations, (c) the different origins and the same destination, (d) the various origins and different destinations.

A certain amount is charged while driving a taxi for a certain distance. After the relaxation of taxi regulations in 2002, each operator can freely decide to increase the amount of the charge (MLIT). In certain regions, only distance system fares are used when traveling in cabs on expressways. This is done to prevent excessive fares if taxis are stuck in traffic on the highway and passengers are unable to drop off them. When a taxi goes slower than a certain speed (10 km/h) or stops, the elapsed time is converted to distance using a formula, and the cost is added to the fare. Therefore, if there is traffic on the way to the destination, the fare is fairly expensive for the distance traveled. The fare determined by how long the passenger uses the taxi is called time charge care. This technique is often used by taxis to visit tourist attractions. Pre-determined fare without distance or time. This method is often employed by taxis transporting people to and from the airport.

The taxi industry in Japan is the increased competition from ride-hailing services such as Uber. While these services are still relatively new in Japan and face regulatory hurdles, they have the potential to disrupt the general taxi market by offering lower prices and greater convenience. Additionally, there is a perception among some consumers that taxi fares in Japan are relatively high compared to other countries, which can make the service less attractive to price-sensitive customers. Taxis provide door-to-door service as a highly flexible means of transportation, but this flexibility also introduces another significant challenge: taxi idle time. Idle time is a source of overhead for the operator. Idle time may be divided into two categories: moving idle time and stationary idle time. When taxi drivers move their idle time, they simply travel while looking for customers to pick up or return to their previous place. Taxi idle times occur for several reasons, the most common of which is a mismatch between supply and demand, as well as unevenly distributed demand (unless peak season) (taxi forecasting). Demand heterogeneity often occurs in urban areas due to high uncertainty when demand is unevenly distributed (context-aware), but in suburban areas, taxi demand is uncertain due to low per-taxi catchment area and low passenger density per square kilometer. As a result of all these changes, suburban taxi drivers have more idle time than city taxi drivers on daily trips. So, taxi businesses in suburban areas are facing a major problem, Hsu (2010). With such a wide catchment area, the main difficulty is taxi allocation. Without an optimized allocation strategy, the operator will have enough idle hours (context-aware). The study focused on this fleet allocation problem as we discovered the problems that operators regularly face due to it.

Shi and Lian (2016) studied the behavior of socially aware taxi passengers and used social welfare to maximize the taxi buffer size. They also looked at how the government regulates the number of taxis and taxi drivers by subsidizing taxis or imposing taxi surcharges. In a two-level choice model of customer-searching taxi drivers, Wong *et al.* (2014a) developed a hierarchical logit modeling technique. Taxi drivers can use the hierarchical logit method to interpret real situations. The model framework should be more suitable to describe the desired behavior of taxi drivers when visiting the taxi stand to look for customers. The authors identified factors that influenced empty taxi drivers' decisions to go to nearby taxi stands to look for passengers and wait for other empty taxi drivers. They talked about possible consequences of the proposed taxi legislation, such as adding additional full-time, daytime, and nightstands in various service areas, and enacting measures to increase taxi stand utilization rates. In a smart city, Askari *et al.* (2020) showed how to forecast taxi demand. They proposed a different model. The authors found that taxi demand in each region is influenced by neighborhood location as well as neighborhood attractions. Data sets for points of interest provide useful information during peak hours. To improve the basic prediction model, they looked at the influence of external factors. Miwa *et al.* (2013) analyzed data from Nagoya City in Japan to examine how taxi probe car devices might be assigned effectively. Despite the fact that taxis in Manhattan spend 33 percent of their time cruising for customers, even at rush hour, O’Keeffe *et al.* (2021) developed a model for the occurrence of passes by investigating vehicles and seeking to find a practical use for it. Stationary idle time refers to when a driver parks his or her automobile for a variety of reasons, such as taking a break or waiting for a client's request. A method where, for example, a user hires a taxi for a day and pays its driver the smallest amount of compensation as fare, which appears to be the same as the driver's day's sales, with less distance and time. This method is sometimes observed in private businesses with frequent customers. Higher fares apply for late hours which are between 22:00 (23:00 in some urban cities) and 5:00 the next morning. Usually, 20 to 30% of the regular fare is added. As 'warimashi' (extra fare) is printed in blue on the signaling lights, this time zone is also known as 'Aotan' (literally, a blue strip of paper). Winter rents are more expensive. Based on the fact that road conditions in Hokkaido, Tohoku, and Hokuriku become dangerous in winter.

Depending on the area, disabled persons are usually entitled to a 10% discount by showing their identification booklet for the physically disabled. Again, when a passenger uses a taxi for a certain distance, a certain amount of fare concession is given. Fare placement will be important to ensure the viability of taxi sharing. Fare distribution will affect participation. The first step is to think about equity issues. With changes in various criteria, such as increased travel distance due to roundabouts, increased travel time due to more lifts or stops, and comfort due to more people in less than one taxi, the method of allocating taxi-sharing costs should be different. The second step is to develop a simple pricing model that takes into account the number of passengers as well as the distance traveled.

1.2 Motivation

Taxis are the most readily available mode of transport in low-demand situations due to the speed, door-to-door service, privacy, comfort, 24-hour operation, and no parking fee. A taxi is a form of public transport that provides convenience in both private and public transport. Taxis can be used in a variety of ways, from private use and variable routes to shared use and predetermined routes. Due to the geographical dispersion of demand, empty taxis accumulate in places where they will not be used, leaving a shortage of taxis where they will be needed. Taxi companies do not make a profit even if they run empty taxis. Consequently, optimizing taxi operational costs is an important component of transport operators' planning and control activities. In local cities, population density is declining, and it is expected that the role of taxis in public transport will increase in the future due to the reduction or elimination of buses on certain routes.

A study of shared taxi trip costs in Japan helps us understand the factors influencing pricing and identify ways to make the industry more accessible and affordable. It can be influenced by a variety of factors, including operating costs, competition, and regulatory requirements. In exploring how the cost structure of shared taxis affects pricing, as well as the impact of competition from other transportation providers. This may involve examining pricing strategies that encourage ridership, such as dynamic pricing based on demand, or exploring the potential for subsidies or other financial incentives to promote shared taxi use. However, regular taxi and shared taxi fares remain high due to labor shortages.

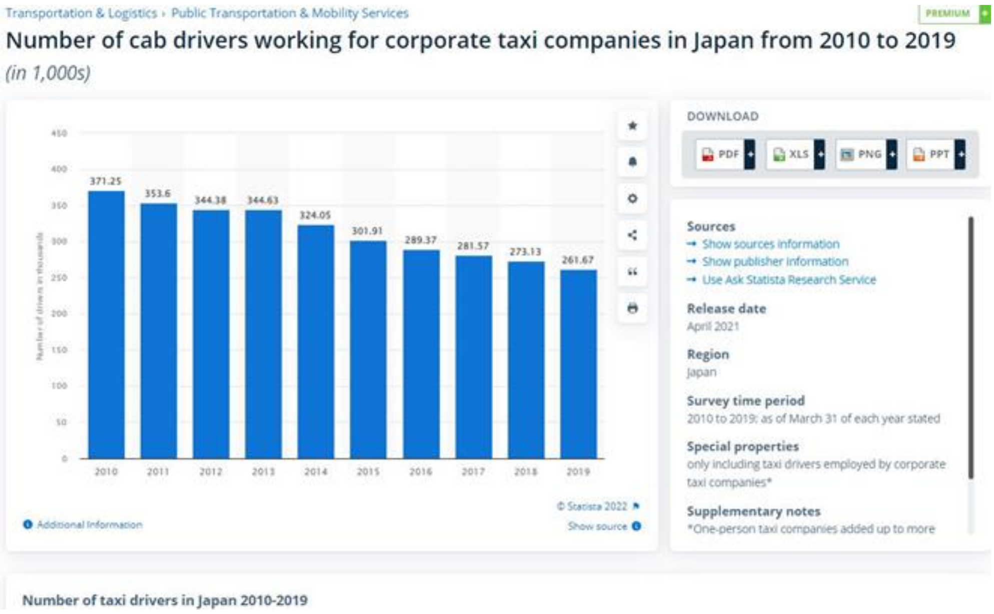

Figure 1.1. Number of taxi drivers decrease.
[Source: Statista (2019)[1]]

My research analyzes current operational conditions, finds existing problems, and ascertains potential for productivity gains. Also, we are trying to reduce costs by reducing operating costs and increasing demand by creating a new efficient taxi system based on real-world understanding. During peak hours, taxis are undoubtedly the most popular mode of transport. That is, taxis are often in short supply. During off-peak hours, however, the difference between high-profit and low-profit drivers becomes apparent. The driver should be aware of the locations where they can pick up people at any time of the day. Another factor is the length of the typical journey that results from a pick-up location. As we know, demand for taxis is generated by transit hubs, stations, retail malls, restaurants, and hotels. Instead of random cruising, an experienced local driver usually has a place to go to pick up a new passenger after dropping off a passenger. Instead of wasting fuel on random trips, experienced drivers will wait for passengers in a parking lot, Yuan *et al.* (2011). Optimizing taxi idle time costs for

---

[1] Source: *https://www.statista.com/outlook/mmo/shared-mobility/shared-rides/ride-hailing-taxi/japan*

minimizing CO2 emissions refers to the process of reducing the amount of time that taxi cabs spend idling in order to minimize their greenhouse gas emissions. This approach is important for reducing the carbon footprint of the transportation industry, as taxi cabs can spend a significant amount of time idling while waiting for passengers or traffic.

Optimizing taxi idle time costs by using technologies such as GPS and real-time traffic data to minimize the time that taxis spend idling. For example, taxis can be directed to areas with high demand for rides, reducing the amount of time they spend waiting for passengers. Additionally, taxi companies can implement policies that encourage drivers to turn off their engines when idling for more than a certain amount of time. The necessity to minimize traffic in urban environments caused by automobiles looking for parking supplied the momentum for this thesis. As a result of traffic congestion, this wastes time and fuel for cars seeking parking and contributes to idle time and fuel for other drivers, Geng and Cassandras (2013).

Therefore, shared taxi services have gained significant popularity in recent years due to their cost-effectiveness and environmental benefits. These services involve multiple passengers sharing a single vehicle to reach their destinations, resulting in reduced traffic congestion and lower carbon emissions. However, the allocation and routing of shared taxis present complex optimization challenges that need to be addressed to ensure efficient and reliable service.

The allocation and routing problem of shared taxis involves determining the optimal assignment of passengers to available vehicles and finding the most efficient routes for these vehicles to minimize travel time and operational costs. This problem becomes even more challenging when considering real-world constraints such as varying demand patterns, dynamic passenger requests, limited vehicle capacity, and time windows for pick-up and drop-off. To tackle this problem, the development of mathematical models and heuristic algorithms has become crucial. Mathematical models provide a formal representation of the problem, capturing its key components, constraints, and objectives. Heuristic algorithms, on the other hand, offer efficient and practical solutions by employing approximate techniques to find near-optimal solutions within a reasonable time frame.

The goal of this research is to develop mathematical models and a heuristic algorithm for the allocation and routing problem of shared taxis. The proposed models will take into account various factors such as travel distances, passenger waiting times, vehicle capacities, and time windows to ensure efficient and reliable service. The heuristic algorithm will employ intelligent search strategies to efficiently explore the solution space and find near-optimal solutions.

1.3 Research Questions, Objectives, and Originality

The goal of this thesis is to determine taxi demand and the best taxi fleet allocations and costs for shared taxis in order to reduce the taxi service's system. The three factors of the taxi identified in the research and used in the models are passengers, drivers, and the environment.

1.3.1 Research Questions

In light of this context, the study's goal is to fill in the mentioned above. The primary research question is expressed as follows:

RQ1. What is the most accurate and effective Hybrid Machine Learning approach for predicting taxi demand in a suburban area?

RQ2. How can the shared taxi routing problem promote affordability and accessibility for passengers in Japan, particularly in underserved or rural areas where general transportation options may be limited?

RQ3. How to design the appropriate trips cost for urban, suburban, and rural areas in a mathematical model to meet the conditions of the shared taxi problems with pickup and drop-off in time windows?

RQ4. How to cost minimize through optimizing idle time in taxi operations and improving the utilization of taxi fleets to increase efficiency?

RQ5. How can taxi allocation be optimized while also improving the quality of service for passengers optimized in urban areas with high demand and rural areas with low demand, such as reducing wait times and improving safety?

1.3.2 Research Objective

The most important objectives of this thesis are:

- **Hybrid Machine Learning Algorithm**: To build a new Machine Learning model and predict how time varies for passenger demand and taxi drivers waiting time.
- **Minimizing the total trip costs of passengers:** To formulate a nonlinear shared taxi model to ensure that passengers receive a fair and reasonable cost for their shared taxi

trips, while also encouraging them to use the service during off-peak hours to reduce traffic congestion and increase profitability.

- **Improving the overall performance of the shared taxi service:** To optimize various aspects of the service, such as route planning, and scheduling, to ensure that the service provides a valuable transportation option for passengers, while also contributing to the broader goals of reducing traffic congestion, improving air quality, and promoting sustainable urban mobility.
- **Minimizing the number of taxis fleet utilized:** To optimize the allocation of taxis based on demand patterns, to ensure that the service remains cost-effective and efficient, while also minimizing the environmental impact of excessive taxi usage.
- **Minimizing the cost of taxi idle time:** To minimize the total operational cost when minimizing the idle time of the taxi driver and the number of taxis utilized of potential demand.
- **Minimize CO2 emissions**: To determine the idle time and reduce taxi numbers to optimize taxi operations thus reducing CO2 emissions.

1.3.3 Research Originality

The taxi model is a type of mathematical model used in transportation research to determine the fare charged to passengers in a shared taxi service and taxi utilization.

- The **proposed hybrid Machine Learning** models for taxi data represent a novel approach to the analysis and optimization of taxi services. These models integrate different Machine Learning techniques, use multiple data sources, develop new prediction models, and focus on real-time prediction, leading to more accurate, efficient, and profitable taxi services.
- The **development of mathematical models** for shared taxi routing problems represents a novel approach to the pricing of shared transportation services. These models integrate various factors, use optimization techniques, focus on sustainability, and develop new mechanisms, leading to more efficient, fair, and sustainable transportation systems.
- The **development of mathematical models for taxi allocations** represents a novel approach to the management of taxi fleets. These models integrate data sources, use real-time optimization, develop new algorithms, and focus on sustainability, leading to more efficient, sustainable, and profitable taxi fleets.

- The shared taxi fare model provides **a fair and efficient transportation service** by optimizing the allocation of taxis and adjusting fare rates based on demand patterns. This helps to ensure that passengers receive fair and reasonable costs for their trip, while also ensuring that the taxi service remains profitable.

Its originality lies in its ability to optimize the allocation of taxis, adjust fare rates based on demand patterns, and provide a fair and efficient transportation option for communities.

1.4 Publications Related to the Research Done in the Thesis

The following periodicals have published the contents of the current doctorate thesis:

- MONDAL M., SANO K., WATARI T., PUPPATERAVANIT C., PERERA F., and DE SILVA C.K., An Efficient Modelling Approach on Passenger Demand with Fare in Shared Taxis: A Case Study of Sanjo City, *Japan, Journal of the Eastern Asia Society for Transportation Studies*, 2022 Volume 14 p. 684-701, https://doi.org/10.11175/easts.14.684
- Mondal M., Sano K. Teppei K, and Chonnipa P., Optimization of Taxi Allocation for Minimizing CO2 Emissions Based on Heuristics Algorithms, *Smart Cities*, 2023, Volume 6, Issue 3, 1589-1611, https://doi.org/10.3390/smartcities6030075
- Mondal M., Sano K. Teppei K, Chonnipa P., and Watari T., Predicting Taxi Demand in Urban Areas by the Application of Hybrid Machine Learning Algorithm, *Journal of the Eastern Asia Society for Transportation Studies (Submitted)*

1.5 Scope and Limitations

The scope of the shared taxi industry in Japan includes the various companies and organizations that offer shared taxi services, as well as the passengers who use these services. This includes both general taxi companies that are beginning to offer shared taxi services, as well as newer startups that specialize in shared transportation options. This research is divided into two categories: academic and practical. The current study, which is related to the academic sector, tries to address the conceptual concerns of why this topic? What difference does it make whether it's been accomplished?

The shared taxi industry in Japan operates in a dynamic and evolving market, with a range of factors that can influence its growth and development. These include changing consumer preferences, technological advancements, regulatory requirements, and competitive pressures from other transportation providers. Research into the shared taxi industry in Japan may focus on various aspects of this market, such as pricing models, demand patterns, operational efficiency, and service quality.

However, there are also some limitations to the scope of research on the shared taxi industry in Japan. One limitation is the relatively emerging stage of the industry, which means that there may be limited data and information available for analysis. Additionally, the shared taxi industry in Japan is just one segment of the broader transportation market, and research may need to consider the impact of other factors, such as public transportation or ride-hailing services, on the demand and viability of shared taxi services. A questionnaire survey might be used to learn about people's attitudes toward taxi-sharing services with various operation methodologies.

Another limitation is the potential for cultural and linguistic barriers in conducting research on the shared taxi industry in Japan. Research may need to be conducted in Japanese, and researchers may need to have a deep understanding of Japanese culture and business practices in order to effectively analyze the industry.

Overall, while there are some limitations to the scope of research on the taxi industry in Japan, there is also significant potential for exploring the challenges and opportunities facing this emerging sector of the transportation market. By understanding the scope and limitations of research in this area, researchers can help inform the development of effective strategies for promoting sustainable and accessible transportation options in Japan.

1.6 Outline of the Dissertation

To begin, Chapter 1 provides an outline of this research. Following a brief overview of the topic's research needs and purpose, the research gaps are highlighted as the backdrop for defining research objectives. The research question is given, together with the response, as the study's objectives. This part is critical in determining which components of the study should be prioritized and why they should be taken into account.

Chapter 2, we review papers of taxi demand, taxi fleet utilization, taxi allocations, and shared taxi routing modelling, as well as papers on the impact of cost-efficiency of the shared taxi.

Chapter 3, the trajectory data is extracted by using Visual Basic macro-Excel. This chapter focuses on the process of data gathering, including data collection methods and extraction, with a particular emphasis on two types of data: taxi data obtained from Nagaoka City and Sanjo City Taxi companies in Japan, and weather data sourced from the Japan Meteorological Agency's website.

Chapter 4, which provides a concise overview of methods and techniques used to estimate taxi demand, offering valuable insights for analysing and predicting passenger requests in the transportation industry.

Chapter 5 presents a mathematical model for the shared taxi routing problem, focusing on minimizing idle time cost. The chapter discusses taxi allocation strategies and proposes a mathematical framework to optimize routing decisions, aiming to reduce the overall idle time cost of shared taxi services.

In Chapter 6, explores the application of two models in the taxi sector of Sanjo and Nagaoka cities, yielding significant results such as optimal taxi routes and reduced CO2 emissions based on idle time.

In Chapter 7, we present the cost-efficiency analysis of shared taxi operations, the cost-efficiency analysis is to compare the operational performance of shared taxis and general taxis in suburban areas, in Japan, with a focus on factors such as vehicle utilization, driver expenses, and operational costs.

Chapter 8, we discuss the validity of the calculation results, the limitations of this research, and future issues.

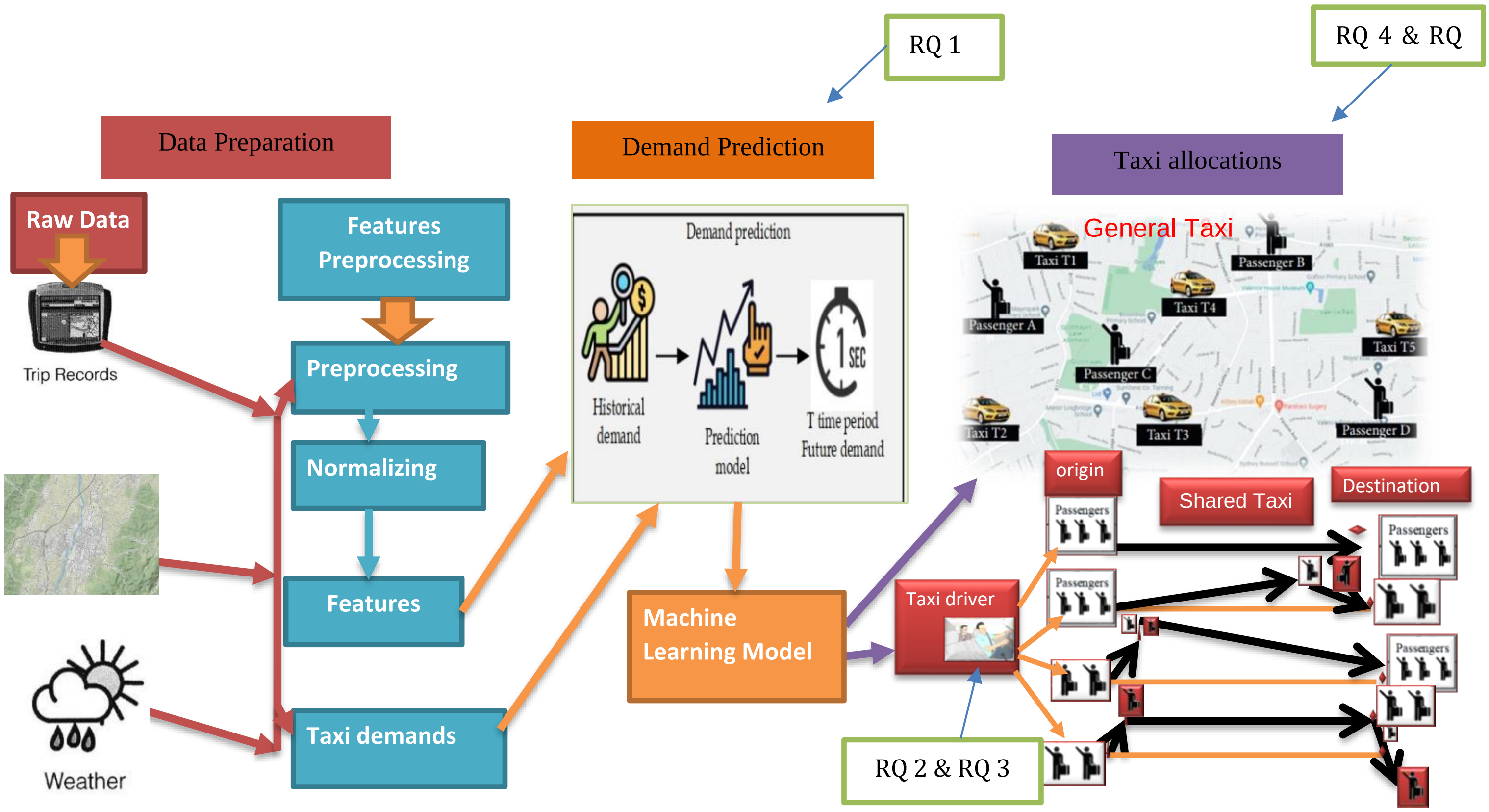

Figure 1.2. Research design – Conceptual framework

Figure 1.3. The framework by Chapter and organization of research

Chapter 2. LITERATURE REVIEW

This chapter begins by reviewing the concept of taxi demand, taxi fleet utilization, taxi allocations, and shared taxi routing modeling. The most important research in each area is detailed in greater depth, with a discussion of their significance to the subject addressed in this study. The chapter finishes with a review of the literature, based on which a solution to the posed problem is offered.

2.1 Review of the Taxi Demand

The demand for taxi services is a complex issue that has been studied extensively in the transportation literature. In the context of Japan, there are several key factors that have been identified as drivers of taxi demand, as well as challenges that the industry faces.

One of the key drivers of taxi demand in Japan is the country's aging population. As the population ages, there is a growing need for transportation options that are accessible and accommodating for older adults. Taxis, with their door-to-door service and ability to accommodate mobility devices, can be an attractive option for seniors who may have difficulty using public transportation. There has been a lot of research on taxi services around the world, and lots of data on the regulation and deregulation of the taxi industry. A lot of information, including position, time, number of customers, weather, traffic, and so on, may be captured in real-time by putting widely used mobile phones in network sensors and taxis. This information covers the way for the development of a smart transportation system capable of controlling and coordinating a large number of taxis. Both taxi drivers and operators may gain from it. Taxi drivers are permitted to operate in places where there is substantial demand for taxis and shared taxi services. They can reallocate their cars utilizing surge pricing in advance to satisfy passenger demand. According to Miao *et al.* (2019), passenger demand data may be employed in taxi dispatch systems to minimize passenger waiting time, taxi traveling time, and re-balancing expenses.

The taxi demand prediction problem has attracted more attention with ride companies' growth. Recently, some studies have been available concerning taxi demand forecasting (Mukai and Yoden (2012); Yao *et al.* (2018); Zhao *et al.* (2016); Zhao *et al.* (2019); Rajendran *et al.* (2021); Askari *et al.* (2020)). Askari *et al.* (2020) referred to points of interest, including the location, business type, and key timelines that emerge in demand as external features for the vanilla LSTM process. The study's findings reveal that external variables such as nearby sites and areas of interest might impact taxi demand forecasting in each region. Human movement patterns in an urban taxi transportation system were studied by Li *et al.* (2012). According to the variance of pickup quantity (PUQ) within an urban region, this can automatically estimate the boundaries of hotspots. In the related work on predicting taxi demand urban human mobility may help passengers have a more pleasant travel experience while also assisting the government in better managing the city's transportation system. Traditional time-series statistics like Auto-regressive Integrated Moving Average (ARIMA) (Li *et al.* (2012); Moreira-Matias *et al.*

(2013)) developed new machine learning time series like Recruit Neural Network (RNN) and Long Short-term Memory (LSTM). (Zhao *et al.* (2019); Vanichrujee *et al.* (2018)) were also improved taxi demand. Studies had also extended the analytical approach based on big data by merging geographical and temporal characteristics, Yao *et al.* (2018).

James Cooper (2010) performed a study on taxi services worldwide. Taxis played an important role in providing transportation to all parts of the world, providing instantly recognizes consistent services. The mode provided access to social activities, contributes to tourism, and significantly contributes to the economy of a location that provided socially desirable accessibility to people without vehicles and used in emergencies. Yang *et al.* (2005) used realistic distance-based and delay-based price structures to investigate the exclusive, socially optimum, and stable competitive solutions of cruising taxi services in the presence of traffic congestion. The authors generalized by taking into account the negative effects of traffic congestion on taxi businesses, customers, and society. They also noted that, when modeling taxi fleets in urban environments, the majority of fleet size and profit occur at the unit flexibility of customer demand, when the increase in revenue from a higher charge is offset by the decline in consumer demand. Yang *et al.* (2010) proposed a solution to solve the problem of bilateral search and meeting between taxis and consumers in a shared network. A generic meeting function has been developed that may capture the many natures of meetings in physically disparate places and contains the waiting time function, which has been widely employed in earlier research in a specific scenario. To provide a stable balancing solution with the least amount of time variable distribution between taxis and clients, a repeating numerical technique was designed. Salanova *et al.* (2014) provided a thorough examination of the evolution of the formulation for modeling taxi services in metropolitan regions. They also discovered and studied several formulations proposed in the literature for each of the three operating modes, offering distinct formulations for each (healing, stand, and dispatching).

Cordeau (2006) devised a branch-and-cut solution for the dial-a-ride problem (DARP) for transportation from a given origin to a specific destination using a mixed-integer programming model on a three-index with a polynomial number of constraints. The challenge entails a minimal cost constraint that satisfies the capacity, time length, time window, pairing, precedence, and ride-time restrictions of the vehicle routes. DARP specializes in providing door-to-door transportation for the elderly and disabled. Sub tour elimination constraints, precedence constraints, extended order constraints, and order-matching constraints are among the inequalities defined by the authors. Solving situations with up to four cars and 32 requests was made possible using this strategy. Ropke *et al.* (2007) created two new mathematical models for the pickup and delivery problem with time windows (PDPTW) of the dial-a-ride problem (DARP) in order to meet a collection of transportation requests between pickup and delivery. The PDPTW and the DARP did tests on numerous different instance sets to see if the branch-and-cut algorithms could solve these inequalities. The deregulation of the taxi sector in Japan has been the subject of several financial literature. Every day, a conventional taxi system transports thousands of customers. It is a wonderful illustration of a supply-demand system, with available taxis pointing in the supply direction and waiting for passengers pointing in the demand direction.

The finest personal judgments contribute to overall inefficiencies in the taxi system, with excess or deficit taxi service delivery and extended periods of idle time with no clients inside. Empty travel diminishes the system to use and adds to traffic congestion. As a result, lowering the efficiency of the taxi industry is a difficult but necessary task for city transportation authorities and governments. Over the previous 10 years, Yang *et al.* (2000) created a simultaneous equation for predicting passenger demand, taxi utilization, and service level-based regions in Hong Kong. The number of licensed taxis, taxi fares, disposable money, time of occupant taxi ride, daily taxi passenger demand, driver eating time, taxi accessibility, taxi use, and average taxi waiting time are all endogenous factors. Once provided external data or inferred from other socioeconomic factors, that model may be used to forecast future performance metrics of taxi services, including taxi demand, taxi utilization, and service quality.

Glaschenko *et al.* (2009) discussed the taxi system as a multi-agent system, offering a real-time model to reschedule taxi service before confirming order acceptance to finish recent trip calls. A very complex problem of operating London's largest taxi service was solved using a new multi-agent adapted scheduling method. This method is innovative, measurable, compatible with modern trends in optimization, and, effective in large-scale, practical, and commercial applications. Shen and Lopes (2015) proposed expanding and Target algorithms that can be readily combined with various scheduling techniques for dispatching autonomous cars to minimize passengers' normal waiting time and boost trip success rates. The authors implanted autonomous dynamics on the demand simulation platform and conduct an empirical study with New York City taxi data in 2013.

Qu *et al.* (2014) advocated that a low-cost referral system for taxi drivers be developed. The purpose of the design is to maximize revenues while adhering to the specified passenger-finding path. To estimate the prospective profit of the driving route, they devised the objective function of a net profit. Then, using previous taxi GPS data, they created a graph representation of the road networks and used a brutal-energy technique to determine the optimum driving path. However, a major roadblock in this direction is the graph-based approach's high processing cost. As a result, they built a creative repeat strategy based on a specific form of the net profit function to determine the best client route. The authors conducted extensive tests on a real-world data set acquired from the San Francisco Bay region, and the findings demonstrated that the suggested recommendation system is successful. To describe the behavior of empty taxis, Zheng *et al.* (2012) devised the non-homogeneous poisoning process, which can better forecast the real scenario and has a simple form for distributing the likelihood of waiting time. They calculated the time it would take for empty taxis to arrive at various points along the route and created a recommendation system for anyone who wanted to take a cab. They also conducted several tests to assess the real trajectory of simulated passengers and taxi drivers. They put the concept into reality by creating a model app that may assist users in finding an appropriate location to wait for a cab.

Qian *et al.* (2015) proposed a Sharing Considered Route Assignment Mechanism (SCRAM) to enhance egalitarian taxi services. SCRAM set out to deliver fair and efficient route recommendations to a community of competing taxi drivers. The fundamental purpose of SCRAM is to validate the accuracy of proposals offered by a group of competing taxi drivers. As a result of the proposed route's partitioning of present road segments, SCRAM's side effect is that taxi drivers' driving skills improve. Furthermore, rather than just directions, SCRAM may provide taxi drivers with an exact driving path. Yuan *et al.* (2011) demonstrated how to utilize the GPS trajectory of a large number of taxis to pinpoint parking spots. Taxi drivers often park their cars while waiting for passengers. The company calculated the chance of picking up a customer if the vehicle drove to a parking lot and parked in certain areas, enabling taxi drivers to use the advice. The authors proposed a feasible model for time-dependent taxi behavior (pickup/drop-off/parking), on both the highway and in parking lots, based on suggestions for taxi drivers and clients. They devised a partition-and-group architecture for acquiring city-wide statistical knowledge in order to provide timely suggestions based on historical data. The technique was tested utilizing a large number of historical GPS trajectories (12,000 cabs in 110 days) produced by taxicab companies. Zhan *et al.* (2014) investigated the effectiveness of the taxi service system using real-world large-scale taxi trip data. In the case of a hypothetical framework recommendation system, two methodologies have been proposed to identify the best theoretical strategies that reduce the cost of idle travel and so need the least number of taxis to fulfill all observed trips. Optimization issues that are converted to equivalent graph problems are tackled using polynomial time approaches. The performance discrepancy between the real system and the theoretically perfect system is assessed using data from cab journeys in New York City.

Ali *et al.* (2004) developed a mathematical model of a new metric to determine a resource allocation related to the functional characteristics of the desired system against multiple disturbances of different systems and environmental conditions. Also, the authors described a general robustness metric method called FePIA (Features, Perturbation, Impact, Analysis) to systematically find out the metric of consistency for different parallel and distributed asset allocation systems. The FePIA method is employed to find out the consistency metrics for three instance distribution systems. Geng and Cassandras (2013) developed a "smart parking" system that uses technology to identify drivers and estimate available parking spaces. It allocated parking spaces to automobiles rather than merely providing instructions. They focused on building an efficient and optimal allocation approach for both the system and the user by solving a sequence of mixed integer linear programming problems with a solution that fulfills specific fairness conditions. The taxi demand prediction problem has recently attracted more attention to ride companies' growth.

Mukai and Yoden, (2012), the authors discussed the taxi demand forecast from taxi historical probe data by a neural network. Since this is a study conducted in Tokyo, Japan, which may be similar in terms of business management and taxi passengers' behavior. This study used the previous demand in each region, day of the week, and rainfall to input the neural network. The results show that the day of the week is critical for forecasting demand because the demand for taxis repeats every week. Whether it was raining or not at the time the taxi picked up the passenger, precipitation was not a factor in the effectiveness of the learning model. The limitation of this study is that it uses rain as the only weather factor to represent the effect of weather. In addition, the second limitation of this study is the forecast timeframe. Instead of forecasting hourly or shorter periods, this study divided the 4-hour interval to forecast the next 4-hour interval. This forecast may be helpful in fleet allocation at the company level but not for drivers' customer acquisition strategy.

Many studies use time series models to build predictive models, mainly LSTM or ARIMA, with different approaches to enrich input features. Rodrigues et al., (2019) argued that a large part of the valuable information for the learning process is in the form of unstructured text. Therefore, this study attempts to build word embeddings and convolutional layers to collect rich text information before combining it with LSTM to forecast demand. Askari et al., (2020) referred to points of interest, including the location, business type, and key timelines that emerge in demand as external features for the vanilla LSTM process. The findings of the study reveal that external variables such as nearby sites and areas of interest might impact taxi demand in a certain region. Human movement patterns in an urban taxi transportation system were studied by Li et al., (2012). They concentrated on determining the pickup quantity (PUQ) for metropolitan hotspots with a high number of passengers boarding or exiting taxis. To cluster hotspots, they suggested an adaptive watershed algorithm. According to the variance of PUQ within an urban region, this algorithm can automatically estimate the boundaries of hotspots. Predicting urban human mobility can help passengers have a more pleasant travel experience and assist the government to enhance the development of a city's transportation infrastructure. They enhanced an ARIMA-based prediction system for predicting how many passengers would be at a specific hotspot in a specific period. These studies are based on the initial assumption that points of Interest (POIs) are related to differences in taxi demand at different times. However, they hypothesize that it is not only the POIs or neighboring locations that create the difference in demand. Therefore, the authors recommend cluster analysis and allowing historical data to voice for itself. The results of our cluster analysis will also be used as reinforcement input for demand forecasting models. Liao et al., (2018) developed the model of the Artificial Neural Network (ANN), Multi Regression, XGBoost, Decision Tree, Random Forest, LSBoost, STResNet, FCL-Net, and CNN-LSTM based on one million taxi records in New York City. Although considering a superior model, they concluded that a deep neural network is only effective when designed with a reasonable structure, where domain knowledge is key. Again, According to Liao et al., (2018), a transportation problem (namely, the forecasting of taxi demand) is well-suited for a convolutional neural network (CNN) based on a deep neural network, which was originally developed to handle picture data. The authors then over the details of two newly suggested deep neural network (DNN) designs, ST-ResNet and FCL-Net, which were both meant to directly address the three constraints, albeit in different ways. They implemented both networks and used them to estimate taxi rider demand using a common dataset, namely New York City taxi data. Zhao et al., (2022) analyzed the effectiveness of the conventional ARIMA prediction model and the current machine-learning model using 14 million sequence data. First, do some research to see which model has the most accuracy. After that, the team determined the maximum accuracy and examined which models exceeded it. The unexpected finding is that the machine learning algorithm is not the best prediction answer. ARIMA-based forecasting is not the optimum method, according to Moreira-Matias et al., (2013). They introduced a novel are included, demonstrating that their technique can achieve high-accuracy prediction. They concentrated on the challenge of real-time decision-making about where to go following a passenger drop-off (i.e., the frame where another passenger can be picked up more quickly). Another set of machine learning models is used that does not rely on time series data.

TaxiRec, a framework for analyzing and detecting the prospective passengers of road clusters, was proposed by Wang et al., (2018). In its neural network forecasting algorithm, TaxiRec considers three important factors: locations of interest, road length, and road type. To find difficult taxi supply-demand patterns, the author employed a deep neural network structure. They used multiple data sources, including car-hailing orders, weather, and traffic data, to forecast taxi demand for each construction block. They used time-series data to analyze taxi demand in a single area. The road type, road length, and point-of-interest kinds are all fixed parameters for one building block, therefore they can't be used to anticipate anything. Li et al., (2011) used the L_1-norm support vector machine as a feature selection method to find the most important feature patterns that affected taxi performance. They discovered that the suggested designs are capable of accurately interpreting empirical research results produced from raw data analysis, and they devised a novel way for describing passenger search techniques regarding the time triangle. A taxi performance prediction model was also built with high accuracy. Using the trajectories of past trips during the training phase, Li et al., (2018) suggested a multi-task representation learning method for taxi journey time estimate, which improved performance. To anticipate taxi demand in cities, recent studies have merged deep neural networks with different characteristics.

2.2 Review of the Modeling of Shared Taxi

Since the early 1970s, several studies on the taxi business have been published. While the early studies (1970–1990) employed aggregated models to examine the taxi sector's productivity and the necessity for government intervention, subsequent studies (1990–2010) used more realistic taxi industry models. New cab technologies like GPS, GIS, and GPRS were also simulated in different models to demonstrate their advantages and justify their use. Several of the models developed have been evaluated in various locations throughout the globe using data from different sources.

Agatz et al. (2012) focused on the optimization challenge of optimally matching drivers and passengers in dynamic ridesharing networks. There are two phases to this ride-matching puzzle. It initially calculates the most efficient vehicle routes before assigning passengers to them, balancing the competing goals of increasing the number of matched travelers while minimizing operating costs and passenger discomfort. They designed and built a taxi-sharing system that accepts real-time trip requests from taxi customers via smartphone and dispatches relevant taxis to pick them up through ridesharing, within time, capacity, and monetary constraints. Ma *et al.* (2015) devised and built a taxi-sharing system that takes real-time trip requests from taxi customers through smartphones and dispatches relevant taxis to pick them up via ridesharing, subject to time, capacity, and financial constraints. Hosni *et al.* (2014) stated that taxi sharing improves taxi service quality, reduces traffic congestion, and reduces environmental impact, and then utilized a Lagrangian decomposition approach to model the taxi sharing problem as a mixed integer program. They described the taxi-sharing dilemma and its solution using heuristics.

Jung *et al.* (2016) established a dynamic shared taxi problem to optimize. Three kinds of taxi dispatch algorithms are investigated: nearest vehicle dispatch, insertion heuristic, and hybrid-simulated annealing. To begin, the taxi demand in this study is derived from the KOTI regional transportation planning model, in which the survey and GPS data from taxis are based on traditional street taxi-hailing conditions, whereas the simulation study assumes an online e-hailing system, in which case actual demand patterns may differ. Second, we expected consumers to share their travels with others, and the simulation results show that ridesharing has a lot of potential for cutting taxi costs by improving system productivity. However, unless incentives are provided based on wait times and diversion hours, not all passengers are ready to share their travel.

Wang *et al.* (2017) formulated the dynamic shared taxi routing issue (D-STRP) and created a simulator that simulates the taxi-sharing procedure. The authors designed a generalized taxi allocation strategy to make a trade-off between passenger wait and detour time and taxi route length to optimize the suggested problem's goal. They also provided the insertion algorithm, which allocated the passenger request to the best option available at the lowest cost. The simulation findings indicated that the modeled taxi allocation strategy had a significant influence on D-STRP. They could determine the best taxi distribution strategy in each travel situation using simulations, which is critical for taxi service providers. Chen *et al.* (2019) used expected optimization approaches and a time-space network strategy to construct a stochastic taxi pooling model. They were written as NP-hard integer multiple-commodity network flow issues. They provided a solution approach that combines the CPLEX with an issue decomposition strategy. First, the authors developed a unique model, when vehicle trip times are stochastic that can be used to solve the taxi pooling issue. Second, a solution approach is developed to address the model efficiently and effectively with large-scale challenges encountered in reality. Third, in a case study, the adoption of this unique model outperforms the numerical method, which typically uses average vehicle journey durations. The Capacitated Vehicle Routing Problem (CVRP) was created by Ben-smida *et al.* (2020) to reduce the total cost of all journeys for the taxi sharing issue, in which items are swapped between passengers and trucks exchanged between taxis. They proposed a new approach called the Iterated Granular Neighborhood Algorithm (IGNA). They investigated how the IGNA algorithm handled a variety of real-world situations involving 9 to 57 people. They compared the IGNA to the parallel micro evolutionary algorithm (pEA) and observed that the IGNA had a lower average than the pEA. Mahmoudi and Zhou (2016) created a mathematical model for the vehicle routing issue with pickup and drop-off with time windows (VRPPDTW), which can identify time-dependent link travel time owing to traffic congestion at different times of the day. Using the Lagrangian relaxation framework, they optimized for matching passenger demands to transportation service providers, synchronizing vehicle routing, and computing requests for balancing transportation demand fulfillment and resource requirements on urban networks. They analyzed a variety of search space reduction methodologies and tested sub-gradient-based algorithms on medium and large-scale transportation networks, including the Chicago sketch and Phoenix regional networks, using a dedicated C++ application. Fagúndez *et al.* (2014) presented the application of an evolutionary algorithm to solve the problem of distributing a group of passengers traveling in different taxis from the same source to reduce the total cost of travel. They examined and compared the quality of the solutions obtained using the proposed algorithm which is calculated using an intuitive greedy heuristic.

Furuhata *et al.* (2013) presented a categorization of current ridesharing systems as well as the creation of appealing processes, such as a concierge-like ride arrangement and the establishment of confidence among unknown riders in online platforms, all of which point to new research prospects. The ridesharing taxonomy identified ridesharing matching patterns that can be satisfied by existing industry methods and some that are still challenging to meet. Both the degree of automation in matching formations and the intended demand segment were considered. For automated ride-matching, dynamic real-time ridesharing matching agencies used modern technologies such as GPS, web, and mobile technologies for real-time communication and developed rideshare matching systems based on routing and scheduling functions.

Based on two integer programming of the issue, Baldacci *et al.* (2004) offered both exact and heuristic approaches to carpooling. The exact system is built on a bounding approach that combines three lower limits derived from several issue reductions. The heuristic approach, which converts the solution of a Lagrangian lower limit into a feasible solution, yields a valid upper bound. The suggested precise technique may solve real-world Car-Pooling Problem (CPP) instances in acceptable computing time, whereas the heuristic takes limited computing time to produce effective solutions in all cases. Lokhandwala and Cai (2018) used agent-based modeling to investigate the advantages and downsides of taxi sharing. The authors developed a model that took into account individual preferences, compared conventional taxis to autonomous taxis, and investigated the geographical variation in service coverage in New York City taxis as a result of ride-sharing. They compared the advantages and disadvantages of ride-sharing with general taxis and shared autonomous taxis. They also examined how ride-sharing has affected the geographical distribution of service coverage. They discovered that moving from regular taxis to shared autonomous taxis may cut fleet size by 59 percent while maintaining service levels and avoiding major increases in passenger wait times. Watel and Faye (2018) presented a taxi-sharing dilemma in which the expense of the journey is split evenly among the passengers. The cabs were able to meet capacity and time constraints. They defined three versions of the problem: the maximum Dial-a-Ride Money problem with money, in which the goal is to drive the maximum number of clients with a large number of taxis arbitrarily; the max-1-Dail-a-Ride Money problem, in which the goal is to drive the maximum number of clients with one taxi; and the 1-Dail-a-Ride Money problem, in which the goal is to drive the maximum number of clients with one taxi. They demonstrated that the 1-Dail-a-Ride Money issue is NP-Complete and that the maximum Dial-a-Ride Money and max-1-Dail-a-Ride Money can't be approximated in polynomial time to within any variable ratio. Al-Ayyash *et al.* (2016) evaluated the potential for market demand for shared taxi services in an organization-based setting. It provided an integrated choice and latent variable modeling technique for calculating how often a shared taxi would be used in a company if it were implemented. They also looked at how cost incentives and a variety of shared taxi features influenced passenger behavior in real-world policy contexts. Jung *et al.* (2013) investigated how shared-taxi algorithms might improve taxi services, as well as what sorts of functions and limits can be employed to reduce excessive passenger flow. Hybrid simulated annealing is dynamically conducted to effectively assign passenger needs, and a series of simulations are undertaken to utilize two different taxi operating methodologies. The authors assumed that all passengers want to share their rides with others, and the simulation results show that increasing ride-sharing system productivity has a large potential to reduce taxi prices. Paraboschi *et al.* (2015) proposed a framework for measuring the impact of ride-sharing at the city level, which they applied to the present taxicab service. The authors included ride-sharing in a typical economic model of the regulated taxi industry. Based on many sample factors and data, the model was able to anticipate the interactions between the demand and supply of a shared taxi service. They tested the approach using a New York taxi market case study. The study emphasized the significant effect of pricing policy and taxi fleet management on the systemic results of a shared taxi system at the city level. Santos and Xavier (2013) proposed a system that can help people find shared rides, reducing greenhouse gas emissions and saving money. The authors provided the best way to get to the destination by modeling a moving scene so that people can share the ride even if the car has left its source. They proposed a greedy randomized adaptive search procedure heuristic algorithm.

Nguyen *et al.* (2015) developed a feasible hybrid transportation solution for Tokyo, in which a passenger and freight share a taxi. To make constraint formulation simpler, they proposed a time-dependent model. They used real-world trial data from the Tokyo-Mussen taxi firm. Over 20,000 daily requests, 4,500 daily serviced taxis, and over 130,000 crossing points on the Tokyo map are all part of the data gathering. They looked at the overall advantages, as well as the distance stored in days, the number of cabs used, and the number of shared requests.

Hosni *et al.* (2014) formulated the mathematical model of the shared taxi problem. The authors optimized the profit of the shared taxi problem under the multi-taxi dynamic dial-a-ride problem (DARP). They have approached a Lagrangian decomposition of the taxi-sharing problem and solved it with the new heuristics method. The inefficiency in transportation causes some economic and environmental problems, such as traffic congestion and severe traffic pollution. So, ride-sharing has some benefits from these problems. Co-ride has some advantages of (i) minimizing the vacancy rate of taxis; (ii) reducing the operating costs of taxi companies; (iii) fares are more favorable for passengers; (iv) reducing road congestion; (v) minimizing the effect of traffic on the environment. First, the passengers that the taxi has already served are called onboard passengers, and passengers who have issued requests to prepare for boarding are called seekers. Taxi-seekers issue a ride-hailing request, including passengers' boarding place and destination, and set the appropriate time, the earliest taxi arrival time expected, the latest alighting time, and the maximum boarding time. Neoh *et al.* (2017) examined how to improve the carpooling market. The authors identified twenty-four carpooling factors, categorized into four dimensions such as demographic, intervention, judgmental, and situational factors. It provided information on policymaking in the demand of the transport management field of the application of meta-analysis.

Yazhe Wang *et al.* (2018) developed a stable matching algorithm to minimize all potential participants' total travel distance, either in a successfully paired ride or an unsuccessfully paired ride. The method can significantly increase the stability of ridesharing at the cost of only slightly reducing system-wide performance. The authors found that during peak hours, the shared taxi demand is very high. Zhang *et al.* (2019) developed the multi-objective optimization model and solved the detour problem of taxi carpooling using a genetic algorithm. They optimized the passenger cost of the detour and non-detour time and maximized the income of the drivers. They also obtained the detour and non-detour payment ratio and travel distance. Ye *et al.* (2015) designed a method to optimize ridesharing routes. The authors optimized the total sharing fare, the traveling time of passengers, and the fuel cost of taxis under the constraints of the driver's income, the traveling fare of passengers, and the fuel charge of taxis. They solved the model by a genetic algorithm. Taking part load network of Lanzhou Chengguan district as an example shows that, when implementing the same travel demands, ridesharing can shorten the taxis' total distance, thereby saving fuel and energy and alleviating urban traffic congestion to a certain extent. Santos and Xavier (2015) examined an optimization model of both dynamic taxi sharing and ridesharing and suggested taking travel distance as the critical factor when calculating the taxi fare allocation. This kind of allocation method is fairer to the sharing of passengers. The optimization model was solved by a Greedy Randomized Adaptive Search Procedure (GRASP) and compared the heuristic and the proposed methods.

Due to the similarities between the ride-sourcing and general taxi markets, supply-demand assets in equilibrium have their origins in research on street-hailing taxi services (Yang and Yang (2011)), e-hailing taxi services (He *et al.* (2018)), and dynamically adjusted trip fares (Sun *et al.* (2019)). Other specific research involves the coordination of demand and supply using price and wage (Bai *et al.* (2019)); pricing and surge-pricing strategies (Chen *et al.* (2020); Yang *et al.* (2020b)); government regulations and policies (Yu *et al.* (2020)); impacts on traditional taxi markets (Nie (2017)); driver labor supply (Zha *et al.* (2018)); supply and demand predictions (Ke *et al.* (2019)); and multi-modal transportation with ride-sourcing and public transit services (Zhu *et al.* (2020)). Based on a queueing model with endogenous supply and demand, (Bai *et al.* (2019)) found that if potential customer demand is significant, the platform should charge passengers a high price, pay drivers a high wage, and implement a high payout ratio (the ratio of wage over price). Taylor (2018) examined the impacts of two essential features of a ride-sourcing market, i) delay sensitivity and agent independence, ii) a platform's optimal strategies in terms of price and wage. The platform should respond by decreasing the price and increasing the wage. Yang *et al.* (2020a) optimized the matching time interval and matching radius of the ride-sourcing demand. They investigated the impact of the new passenger waiting and idle taxi and improved the real-time matching of the ride-sourcing platform. Rong *et al.* (2019) identified the termination of the sharing economy. They developed a theoretical model influenced by the market structure, economic strategy, multi-homing, externality, and public cost. Wang and Li (2020) developed an Agent model based on time-sharing pricing and studied travel decisions for involved individuals and their changes in the share of different modes of travel in transportation systems. The authors analyzed the sharing rate of peak and off-peak periods under the shared economic environment on the Beijing real data and improved the rate of incomes and equilibrium of the sharing rate of different types of travel. Wallsten (2015) discovered the competitive effect of the taxi industry. They analyzed a detailed dataset of New York and Chicago taxi data. Pepić (2018) discussed the concept of the sharing economy and optimized the operational cost of the taxi company. The authors examined Uber's economic model, fare systems, driver income, and working conditions. Finally, they estimated the impact of Uber taxi companies in the present and future.

2.3 Reviews of Taxi Allocation Problem

Taxi transportation has developed as a main steam public transportation in several areas of the world (James Cooper (2010); Schaller (2007); OECD and European Conference of Ministers of Transport (2007)). The taxi allocation problem is placed in the context of taxi problems, models, and algorithms difficulties in the literature study.

Taxis are an essential part of the public transportation system in urban and rural areas Wong *et al.* (2001); Yang and Wong (1998); Wong *et al.* (2008); Wong *et al.* (2014b); Zhan *et* *al.* (2016). They offer a flexible and on-demand service that can accommodate multiple passengers, thereby reducing the need for individuals to rely on their private vehicles. For example, in 2019, New York City had 116,854 prearranged service vehicles and 16,678 street hail service vehicles, but more than one million rider trip requests was received every day. Hua et al., described in 2020, the 18,163 taxis in Hong Kong carried nearly one million passengers daily Hua *et al.* (2022). Although these studies analyzed the taxi demand in urban areas, feeder services of taxis are also important in areas where public transport services are not comprehensive enough to serve all locations, especially in more rural or suburban areas. Taxis can provide an efficient and reliable link between public transport hubs and other parts of the community, allowing people to access public transport more easily and quickly. Furthermore, taxis are accessible to people who may have difficulty using other public modes of transportation, such as those with health, mobility, or vision impairments. The taxi demand again increases when the older population increases as a more convenient mode of transportation. Most developed countries experience this in many sub-urban areas, Vazifeh *et al.* (2018). Further, taxis are more convenient than public transportation because they can be hailed at any time and can take passengers directly to their destinations. By using taxis and public transportation instead of owning a car, individuals can save money, reduce traffic congestion, and lower their carbon footprint.

According to Abbas Abbas (2016), future taxi fleet size is determined by using three methods: "generic algorithm for estimation of taxi fleet size," "taxi fleet requirements based on taxi demand model," and "taxi fleet requirements based on taxi availability index". Yao et al. established a bilevel programming model for taxi fleet size demand according to capacity configuration and ticket cost, Yao *et al.* (2016). Li et al. introduced a strategy for logistics companies to optimize their public charging infrastructure localization and route planning. The proposed approach utilized a bilevel program and a two-phase heuristic approach, combining a two-layer genetic algorithm (TLGA) and simulated annealing (SA). The author focused on determining the optimal locations for public charging stations in Chengdu, a major city in southwest China. The study highlights the advantages of employing a bilevel optimizing approach in addressing the challenges of citywide charging station location selection and logistics routing problems, Li *et al.* (2022). Abbas (2016) discussed taxi fleet size allocation and demand forecasting. The study explained three models, future taxi fleet size, taxi fleet requirements a genetic algorithm based on taxi demand, and evaluated taxi allocation and demand prediction. Yao *et al.* (2016) established the bilevel programming model for the fleet size demand according to the capacity configuration and ticket price.

Mingolla and Lu associated vehicle technology to reduce the carbon emissions of the taxi fleet Mingolla and Lu (2021). The authors discussed a taxi cost-benefit analysis of the CO2 equivalent. Li et al. aimed to examine how various optimizations of taxiing paths can contribute to the reduction of CO2 emissions. The authors analyzed to determine the influence of different parameters, such as aircraft speed, taxiway layout, and environmental conditions, on the reduction of CO2 emissions Li *et al.* (2019). Eslamipoor (2023b) proposed a comprehensive model for inventory transportation planning that incorporates multiple products, multiple periods, and CO2 emission considerations within a two-echelon framework. The problem focused on the deterministic and dynamic demands of retailers, where the supplier must adhere to delivery schedules and product quantities based on a vendor management inventory policy. The environmental pollution factor is recognized as a significant influence in assessing and choosing a supplier. The suggested model aimed to achieve several objectives, including reducing supplier costs (covering total ordering and shortage costs), minimizing the receipt of low-quality goods from suppliers, minimizing delivery time, and mitigating the emission of environmental pollutants caused by vehicles, Eslamipoor (2022). Nilrit and Sampanpanish conducted tests on a chassis dynamometer in an emission lab to determine automobiles' greenhouse gas (GHG) emissions. They obtained results for CO2 and methane emission rates from taxis and passenger cars that run on alternative fuels at different driving speeds. These findings can serve as a database for decision-making in developing transportation projects and controlling GHG emissions in Thailand's mitigation plans Nilrit *et al.* (2018). Zhang et al. conducted a case study in Shanghai to explore the potential for reducing carbon emissions from urban traffic based on CO2 emission satisfaction. The study utilized data collected through travel surveys and transportation models to assess the impacts of various measures on CO2 emissions reduction and travel satisfaction. The research team examined the effectiveness of measures such as optimizing traffic signal timings, promoting public transportation, and encouraging non-motorized transport modes, Zhang *et al.* (2020). Ghahramani and Pilla (2021a) put forward an unsupervised learning method to explore how taxi trips affect CO2 emissions. They utilized a hierarchical clustering algorithm optimized to identify emissions-related clusters, enabling the identification of the most polluting trips. By doing so, they could pinpoint the vehicles associated with these trips, thereby prioritizing CO2 emissions and enabling informed decision-making in the future. Their results could give decision-makers a clearer knowledge of fuel use and policy recommendations. The accessibility of data on taxi operations opens new avenues for mitigating the CO2 emissions brought on by taxis. Most previous studies on taxi emissions and greening the supply chain are based on surveys and statistical data Wang *et al.* (2015); An *et al.* (2011); Zhang *et al.* (2020); Eslamipoor (2023a). These studies adopted a new approach where the authors assessed and examined the CO2 emissions from different forms of urban passenger transportation, such as cars, buses, taxis, and rail transit. Based on such information, several models may be created, and it is possible to determine the average amount of air pollution caused by vehicles.

In addition, there are also several studies trying to improve the heuristic algorithms of transport systems. Abdel-Basset et al. presented a new nature-inspired meta-heuristic algorithm named the spider wasp optimization (SWO) algorithm, which was based on replicating the hunting, nesting, and mating behaviors of the female spider wasps in nature, Abdel-Basset *et al.* (2023). Kaya et al. aimed comparison of the performance of 7 meta-heuristic training algorithms in the neuro-fuzzy training for maximum power point tracking (MPPT), including particle swarm optimization (PSO), harmony search (HS), cuckoo search (CS), artificial bee colony (ABC) algorithm, bee algorithm (BA), differential evolution (DE) and flower pollination algorithm (FPA) Kaya *et al.* (2023). Ouyang et al. proposed a new heuristic solver based on the parallel genetic algorithm and an innovative crew scheduling algorithm, which improved traditional crew scheduling by integrating the crew pairing problem (CPP) and the crew rostering problem (CRP) into a single problem Ouyang and Zhu (2023).

The Spatial-Temporal Diffusion Convolutional Network (ST-DCN) was also shown to outperform seven existing state-of-the-art baseline techniques to anticipate city taxi problems for waiting time (Rahimi *et al.* (2021), Sun and Fan (2021)). Currently, Yao *et al.* (2022) analyzed changes in the context of COVID-19 from both the demand and supply of taxi travel under situations such as closures and travel limits, using system dynamics to mimic taxi travel systems. A typical Chinese middle-sized city, Ningbo in Zhejiang province, discovered elements that contributed to a considerable decrease in taxi travel due to the outbreak. Liu *et al.* (2015) wanted to emphasize the necessity of communication problems for taxi supply among taxi drivers to estimate where and when a customer may require a vehicle. The routing data can be used in a machine learning process, which can then be integrated with an optimization process to reduce taxi idle time. You *et al.* (2021) found that the neural network model can dynamically predict taxi passenger waiting time in real-time given specific spatiotemporal restrictions. Chen *et al.* (2022) presented a generative adversarial network-based technique that can greatly minimize computation. The generative adversarial network technique has two significant drawbacks: selecting the generative adversary network generator's input and finding many optimum solutions to aid in the generative adversary network's training. In order to identify the next passenger, it is also a hub for study on how to travel. Zhao *et al.* (2015) developed a method that uses a decision graph and a data field to cluster trajectory data. By comparing it with common clustering methods, it can determine parameters automatically rather than relying on experience. In addition, the method was applied to the discovery of urban hotspots in Wuhan City, China, based on trajectory data. Using taxi trajectory data, the distribution and dynamics of the hotspots were examined in relation to holidays, weekdays, and weekends. Kong *et al.* (2017) proposed a Time-Location-Relationship model for a data-driven taxi service suggestion to the taxi hotspot region. The model goes a step further to reveal the relationship between a passenger's ups and downs, which is then used to forecast the spatiotemporal distribution of passengers across various social functional zones and raise taxi drivers' earnings. This model assisted taxi drivers in quickly locating the next possible passenger. The authors suggested rewarding regions close to the driver's present location, which would be more practical in practice because drivers aren't ready to take a certain route or travel further only to get the optimal pick-up position.

The model uses a weighted confusion matrix and a modified Viterbi algorithm to anticipate the availability of empty taxis and provide pick-up locations within arbitrarily defined zones and time intervals, considerations such as time of day, and traffic conditions. Hu and Thill (2019) offered a successful Hidden Markov Model-based method for predicting the arrival of unoccupied taxis at predetermined places and times. The model was applied to a large-scale dataset of taxi trajectories in Beijing. Cruz *et al.* (2020) proposed a mathematical model for the vehicle allocation problem in logistic systems, focusing on freight road transportation. The proposed model is solved by a new technique using a branch-and-price algorithm and a column-generating approach with internal point stabilization based on Dontzig-Wolfe decomposition to tackle the problem. Machine learning and fuzzy methods have been around longer than optimization and simulation of transportation (de la Torre *et al.* (2021)). Ramezani and Nourinejad proposed a network-scale taxi dispatch model that optimizes an efficient dispatching system while considering the inter-related influence of typical traffic flows and taxi dynamics problems. To manage the taxi dispatch system, a model predictive control strategy is developed by Ramezani and Nourinejad (2018).

A mixed-integer program's (MIP) purpose is to balance the routing and client allocation costs, and it's done through heuristics (Labbé and Laporte (1986)). The dynamic vehicle allocation problem (DVAP) is raised in this context, which entails assigning empty cars to terminals where they are needed, given a set of required demand services between terminals, to maximize the profits generated by empty vehicle movements (B.Powell (1988)). Reihaneh and Ghoniem (2019) investigated vehicle routing with demand location and proposed an effective branch and price algorithm. Lin Lin (2014) established a framework for analyzing facility location, demand allocation, and resource capacity. Tanizaki (2013) improved the points to adjust the number of standby taxi cars at taxi stands. Crama and Pironet (2019) explored a multi-period stochastic vehicle allocation issue that occurs in transportation planning and proposed a variety of heuristics for dealing with it in a rolling-horizon scenario, in which information is gradually delivered to the decision-maker.

A queuing model and an automated planning system were applied to taxis to an airport terminal when demand was strong and compared the existing system's taxi and passenger waiting times vs. one that employs ChangiNOW (Anwar *et al.* (2013)). Yao *et al.* (2020) proposed a multi-objective multivariate evolutionary algorithm based on decomposition and a dynamic resource allocation strategy (MFEA/D-DRA), and a dynamic resource allocation strategy with a multi-factorial environment. Based on data analytics, the suggested approach is used to address a multi-objective multi-factorial operation optimization issue of a continuous annealing process. De Oliveira *et al.* (2015) devised a method for serving all current customers in a fair period while lowering the distance traveled by existing free taxi cars. The authors investigated the effects of the Greedy Algorithm based on Euclidean Distance (GAED), the Greedy Algorithm based on the Shortest Path (GASP), the Hungarian Algorithm based on Euclidean Distance (HED), and the Hungarian Algorithm based on the Shortest Path on taxi selection (HASP).

2.4 Reviews of Machine Learning

This sub-section reviews the time series as well as multistep forecasting techniques during the aspects of machine learning methods that are practically useful for creating a forecasting system. In a time series, the order of the data points is usually monitored consecutively. Cochrane (1997) arranged dimensions captured during an occurrence in exact chronological order in a time series The two types of time series are discrete and continuous time series.

Observations are recorded at different points in time in a continuous time series. A continuous time series counts observations at each point in time, while a discrete time series counts observations at different times. Then, at equal intervals, a disconnected series of different periods, such as hourly, daily, monthly, or annual time divisions, were frequently recorded, KEITH W. HIPEL (1994). The process of fitting a time series into a suitable model is known as time series analysis. Model parameters are often derived from known data values in models that seek to analyze and grasp the nature of the series. A time series is assumed to be influenced by four major variables: trendy, cyclical, seasonal, and irregular features. These four components may be distinguished using the data collected. For one-time series decomposition, addiction, and qualitative models are often used, taking into consideration the effects of these four components. The addition model presupposes that a time series observation's four components are independent of one another. On the other hand, the qualitative method permits four components to interact with one another, meaning that they are not necessarily independent.

2.4.1 Machine Learning in Data Networks

Over the last decade, machine learning algorithms have become serious challengers to traditional statistical models in predicting. Following that, the notion expanded to additional models, including support vector machines, decision trees, random forests, gradient boost, and others, which are together known as machine learning models, according to Alpaydin (2004). Preliminary statistical models are used in some of these models. There has been tremendous progress in this area over the years, in terms of the quantity and variety of models as well as the theoretical knowledge of models, Trevor Hastie, Robert Tibshirani (2008).

Many machine learning algorithms are designed to solve classification issues, but they may also be used to solve regression problems, which are used to forecast taxi demand, according to Hou (2010). Luo *et al.* (2021) proposed multi-task deep learning (MTDL), a multi-task deep learning model capable of forecasting the intensity of real-time demand in many traffic zones at the same time, to anticipate short-term taxi demand at a multi-zone level. A nonlinear Granger causal teat is used to discover the spatial temporal causal link between distinct traffic locations to filter the most relevant characteristics for the MTDL model. As a result, the suggested prediction model might automatically learn spatiotemporal characteristics to better capture spatiotemporal patterns and increase predictive performance. Several hyperparameter optimization approaches are also used to find the optimum hyperparameters and tweak the

MTDL model for improved prediction accuracy. The suggested MTDL model with feature selection is compared to the original MTDL model, the equivalent STDL model, and many benchmark algorithms such as LSTM, SVM, and K-NN. The models are validated using real-world data from New York City, with the findings demonstrating that the MTDL model outperforms conventional prediction approaches for calculating MAPE, RMSE, MAE, and R-square in large samples, including feature selection. The scientists used a clever multi-task deep learning model that took into account spatiotemporal dependencies to look at short-term taxi demand estimates.

Xu *et al.* (2021) proposed a smart taxi dispatch system that cut passenger wait times while also reducing the vehicle's idle driving distance. Two distinct models were used to forecast the distribution of taxi demand and prospective destinations, respectively. Models are created using a combination of long-term memory cells and a mixed-density network and are learned from past data. Taxi dispatching is stated as a mixed integer programming problem using these assumptions. In a real-world dataset of taxi travel in New York City, we test the performance and overall system of predictors. Between 2014 and 2019, Xie *et al.* (2020) did a thorough review of current development approaches for urban spatial flow forecasting, which is becoming increasingly significant in urban computing research and is connected to traffic management, land use, and public safety. They categorized the complex and dynamic influencing aspects of urban spatiotemporal flow, as well as the overall data preparation method for prediction. They also looked at current research on well-known and advanced urban flow forecasting approaches, such as data-based, classical ML-based, DL-based, and reinforcement learning-based methods for urban space-time-flow forecasting functions.

Zhang *et al.* (2016) used an integrated off-line data analysis and a real-time roaming sensor network to assess passenger demands, with taxicabs identifying the number of passengers and the time of arrival. They provided a fancy parameter called the pickup pattern that can be used to compare how a taxicab operates daily in a huge dataset, such as 900GB per year in Shenzhen. Omrani (2015) used four machine learning algorithms to predict the travel node of persons in Luxembourg, including Artificial Neural Net-MLP, Artificial Neural Net-RBF, Support Vector Machine, and Multi-nominal Logistic Regression. Data on an individual's attributes, means of transportation, place of employment, and lodging are used in the technique described. The data used in this study came from a nationwide poll. It provides data about persons living or working in Luxembourg's everyday movements (e.g., working from home). They devised unique characteristics to link the characteristics of working individuals to everyday movement (travel between home and work). They analyzed data from public transportation as well as certain home and office areas. The authors used successful prediction rates and cross-validation produced by neural networks to assess other alternative strategies for predicting travel mode preferences.

Zhang and Xie (2008) investigated the use of outmoded models for travel mode choice modeling that rely on different preferred models, such as multinomial logit models. For modeling trip mode choice, a novel support vector machine, and an artificial intelligence model were used. Based on data collected in the San Francisco Bay Area of California, this support vector machine model was evaluated and compared to a multinomial logit model and a multilayer feed-forward neural network model. Because of its promising performance and ease of implementation, their findings suggest that the support vector machine model be employed as an alternate technique for simulating travel mode choice. Xie *et al.* (2003) looked at the capabilities and performance of two developing pattern-recognized data mining approaches for travel mode choice modeling: Decision Tree (DT) and Neural Network (NN). The models are rated as particular, approximate, and comparable to a typical multinomial logit (MNL) model based on these two methodologies. The authors have offered a unique three-tier structure of the MNL model for comparison and have highlighted similarities and differences in the mechanism and structure of the model, as well as discrepancies in the model specification and assumptions. To assess and compare model performance, two performance metrics are used: individual prediction rates and overall prediction rates, which quantify predictive accuracy at the individual and mode overall levels, respectively. Models estimate and assessment are based on diary records from the San Francisco Bay Area Travel Survey (BATS) 2000. In terms of modeling results, the two data mining models are comparable but slightly better than the MNL model, with the DT model demonstrating maximum predictive efficiency and the most obvious explanations, and the NN model giving a higher predictive efficiency and the most obvious explanations, and the NN model giving a higher predictive performance in most cases.

Without any explicit prior programming, machine learning algorithms may learn directly from data. These are the kind of issues that might develop in a networking environment. Machine learning models may be used to solve a wide range of data network challenges, including Logistic Regression (Dreiseitl and Ohno-Machado (2002)), Random Forest (Bentéjac *et al.* (2021)), Gradient Boosting, K-Nearest Neighbors (KNN) (Yu *et al.* (2002)), Decision Tree, Principal Component Analysis (Fitria (2013)), and many more. There are five sub-group of machine learning such as:

1. Supervised: We start with some train data that includes the ground truth values (responses) that we want to predict later (Zhao *et al.* (2019)).
2. Regression: Predicting value is a constant process.
3. Classification: Predictive value is categorical. The task at hand is to apply the forecast to test data for which the expected right value is unknown.
4. Unsupervised: We begin with some training data that is devoid of values to predict. Without any prior knowledge, the purpose is to uncover the structure of this data. Clustering, dimensionality reduction (latent variables identification), outliers' detection, and cumulative distribution function learning are some of the subsets. The task is to predict fresh test data while making use of the extra information offered by the unlabeled data.
5. Semi-supervised: The goal is to figure out the appropriate sequence of actions to maximize the predicted total of future rewards while engaging with an often-unknown environment that provides the rewards and defines the interaction's dynamics. Without understanding the dynamics of the environment, the problem is to train a policy function that yields the optimal current action from an uncorrelated and perhaps sparse series of rewards.

This thesis is mostly concerned with supervised classification techniques, specifically:

- Machine Learning: Linear Regression, K-Nearest Neighbors, Decision Trees, Random Forest, Gradient Boosting.
- Proposed model: Hybrid Machine Learning

2.5 Review of Heuristic and Metaheuristic Methods

Heuristic and metaheuristic methods are widely used in optimization problems across various domains, including transportation, engineering, and economics. These methods can be used to find solutions to problems where it is difficult or impossible to find an exact solution using traditional optimization techniques. Heuristic methods are problem-solving techniques that use practical rules or heuristics to find approximate solutions to optimization problems. They are often simple and easy to understand but may not always produce the best solution. Examples of heuristic methods include greedy algorithms, hill-climbing algorithms, and simulated annealing.

Metaheuristic methods, on the other hand, are more advanced optimization techniques that use heuristic methods to guide the search for a solution. Metaheuristic methods typically work by exploring the search space of a problem systematically and then using a combination of randomness and heuristic knowledge to determine the next steps in the search process. Examples of metaheuristic methods include genetic algorithms, particle swarm optimization, and ant colony optimization.

In recent years, there has been significant research on the use of heuristic and metaheuristic methods in transportation problems. One key area of focus has been on the use of these methods in optimizing transportation network design and traffic management. Researchers have also explored the use of heuristic and metaheuristic methods in vehicle routing problems, such as the traveling salesman problem and the vehicle routing problem with time windows. In the early stage, heuristics and metaheuristics were usually built around available local search operators that have worked well on vehicle routing problems. A heuristic algorithm that sets a route from scratch is commonly called a root-building heuristic. Root-improvement, on the other hand, is an algorithm that seeks to generate a better solution based on an existing one. Metaheuristics are powerful methods that may be applied to a variety of problems. A metaheuristic is an adaptive expert strategy that intelligently integrates multiple conceptions for exploring and exploiting the search space to guide and change the operations of sub-heuristics. Heuristic and metaheuristic algorithms were developed in the late 70s for linear programming, Nonlinear programming, and dynamic programming optimization. When the traditional algorithm is slow then a heuristic algorithm is faster to solve a problem. A metaheuristic is a higher technique for an optimization problem. Because it is theoretically similar to the natural system of annealing, in which an element is heated to a liquid state and then returned to a crystallized solid form, it is called "simulated annealing". Vasco and Morabito (2016) presented a mathematical model for the dynamic vehicle allocation problem. They developed to solve the allocation problem involving creating a greedy and local search heuristic and simulated annealing metaheuristic methods. Daniel Fuentes and Sano (2020) presented a mathematical model of the vehicle routing problem with time windows developed by the greedy and local search algorithm.

The application of heuristic and metaheuristic methods in solving taxi demand problems has been a topic of interest in recent years. Several studies have explored the effectiveness of these methods in optimizing taxi demand solutions, including route planning, dispatching, and pricing. One area of focus has been on the use of heuristic and metaheuristic methods in taxi dispatching problems. Another area of focus has been on the use of heuristic and metaheuristic methods in taxi pricing problems. Despite the effectiveness of heuristic and metaheuristic methods in solving taxi demand problems, there are also challenges to their use. One challenge is the need for accurate and timely data, which is often difficult to obtain in real-world settings. Another challenge is the complexity of the taxi demand problem, which may require the use of advanced optimization techniques and computational resources.

In summary, heuristic and metaheuristic methods have shown promise in solving taxi demand problems, including dispatching and pricing. Ongoing research in this area will be important in further refining these methods and improving their effectiveness in addressing the challenges of taxi demand in different cities and regions.

2.6 Summary of the Chapter

Based on my review of the literature on taxi demand, shared taxi routing models, taxi fleet management, and machine learning, the following conclusions can be drawn:

- Taxi demand is influenced by a range of factors such as time of day, day of the week, weather conditions, and special events. Understanding these factors can help taxi companies optimize their services and increase efficiency.
- Shared taxi routing models have the potential to reduce costs for both passengers and taxi companies. However, the success of these models depends on factors such as customer demand, pricing strategies, and regulatory frameworks.
- Taxi location management is a complex task that involves balancing the supply of taxis with the demand in each area. Several studies have suggested that the use of dynamic pricing and dispatching algorithms can improve the efficiency of taxi fleets.
- Machine learning techniques have been increasingly applied to taxi demand forecasting, dispatching, and pricing. These techniques have shown promising results in improving the accuracy of demand forecasting and the efficiency of taxi dispatching.

Therefore, the literature suggests that the use of shared taxi routing models, dynamic pricing and dispatching algorithms, and machine learning techniques can help taxi companies increase efficiency, reduce costs, and improve customer satisfaction.

Chapter 3. DATA COLLECTION AND ANALYSIS

This chapter delves into the specifics of data gathering, data collection methods, and data extraction. We've spoken about two different types of data. Then we begin data processing for each data set. The Nagaoka City and Sanjo City Taxi companies have provided all the original taxi data in Japan. The Japan Meteorological Agency provides weather data for download from the website.

3.1 Taxi Demand Data

Data collection and analysis are essential components in understanding the taxi market and demand. In the context of the taxi industry, data collection and analysis refer to the process of gathering information on various aspects of the market, such as the number of trips, pick-up, and drop-off locations, fare pricing, and passenger demographics. The collection of taxi data can be done through various methods, including surveys, interviews, and technology-based solutions such as GPS and mobile applications. These data sources provide different types of information, which can be used to analyze the market from different perspectives. Data analysis involves the process of transforming the collected data into useful insights and knowledge. This can be achieved through the application of statistical methods and data mining techniques such as clustering and regression analysis. The results of data analysis can be used to identify patterns, trends, and relationships in the data, which can be useful for developing effective taxi demand solutions.

In recent years, the availability of big data and advanced data analytics tools has enabled more sophisticated analysis of taxi market data. For example, machine learning algorithms can be used to develop predictive models that can forecast taxi demand based on historical data. This can be useful for improving taxi dispatching and route planning.

3.1.1 Nagaoka City Taxi Data

The target area is the entire area of present-day Nagaoka City. Nagaoka City, Niigata Prefecture, is located in the central part of Niigata Prefecture, "Nagaoka Area", "Nakanoshima Area", "Koshiji Area", "Mishima Area", "Yamakoshi Area", "Oguni Area", "Washima Area", "Teramari area," "Tochio area," "Yoita area," and "Kawaguchi area." The current administrative area is 891.06 Km^2, of which about 50% is habitable. The Shinano River, which boasts the longest length and flowing water in Japan, runs through the central part of the city, and fertile alluvial plains spread on both banks. To the east and west, the Higashiyama Mountain range and the Nishiyama hills are connected. The number of households is about 109,263, the population is about 263,728, the elderly population is about 80,000, and about 30 % of the citizens are elderly[2].

[2] nagaoka city website

There are three main public transportation lines operating in Nagaoka City, the Joetsu Shinkansen, Shin-Etsu Line, and Echigo Line. Regarding taxis, seven taxi companies, including Mitsukoshi Taxi, which is the subject of this study, also carry out regular taxi business. Regarding buses, highway buses to Niigata City, the Chuo Circulation Line, the Miyauchi Circulation Line, the Kawasaki Circulation Line, and public transportation projects are widely introduced throughout the city. We will analyze the actual situation of taxi operation by using GPS data of all vehicles and 50 vehicles belonging to Mitsukoshi Taxi Co., Ltd., which is the largest taxi company with a business area in the entire Nagaoka city. The analysis period is one year from January 1 to December 31, 2019. The number of acquired data is 178,254. The data acquisition method is input by the driver and recorded by the date, time, direction, operating status, latitude, and longitude acquired from GPS, and the recording device installed in the vehicle (see Table 3.1). This makes it easy to know where the taxi has secured passengers.

Table 3.1. Selected data (Nagaoka city) for the study

Pickup Date Time	Dropoff Date Time	Pickup Lon	Pickup Lat	Dropoff Lon	Dropoff Lat	Fare (Yen)	Travel Time
1/1/2019 11:32	1/1/2019 11:46	138.8247	37.4634	138.8488	37.44448	1641.288	0:13:55
1/1/2019 12:28	1/1/2019 12:39	138.8233	37.44263	138.8528	37.44718	1260.81	0:10:47
1/1/2019 14:22	1/1/2019 14:38	138.8289	37.46078	138.8526	37.42944	2321.094	0:16:18
1/1/2019 15:09	1/1/2019 15:21	138.829	37.46065	138.8388	37.44359	1408.019	0:12:01
1/1/2019 16:01	1/1/2019 16:21	138.7792	37.44842	138.8517	37.43882	2987.91	0:19:57
1/1/2019 16:51	1/1/2019 17:04	138.8287	37.46076	138.8533	37.44719	1623.026	0:12:26
1/1/2019 17:38	1/1/2019 17:53	138.8273	37.4669	138.8624	37.43584	2548.618	0:14:41
1/1/2019 18:22	1/1/2019 18:29	138.8365	37.47388	138.851	37.45392	1758.715	0:06:39
1/1/2019 19:01	1/1/2019 19:11	138.829	37.46092	138.8529	37.44707	1622.905	0:10:00
1/1/2019 20:52	1/1/2019 21:01	138.8549	37.44654	138.8839	37.43415	1446.137	0:09:24
1/1/2019 21:21	1/1/2019 21:25	138.855	37.44648	138.8707	37.43641	1037.239	0:03:55
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
12/28/2019 14:11	12/30/2019 14:14	138.8547	37.44659	138.8566	37.4406	732.6815	0:02:47
12/29/2019 14:30	12/30/2019 14:34	138.8546	37.44661	138.8611	37.45165	630	0:03:46
12/30/2019 15:50	12/30/2019 15:55	138.855	37.44653	138.8532	37.4326	814.4494	0:04:44
12/30/2019 16:22	12/30/2019 16:31	138.8549	37.44655	138.8838	37.43792	1262.108	0:08:29

In this study, taxi operation data such as date, time, operation status and latitude, and longitude were used. The operating status of taxis recorded in the recording device installed in the car is classified into the following three types.

a. Actual vehicle: Running with passengers.

b. Empty car: A state in which passengers are picked up from the waiting area or are moving toward the taxi waiting area after unloading the passengers.

c. Waiting or resting (hereinafter referred to as waiting): Waiting for passengers at the taxi waiting area. In addition, the state of resting at the resting place.

The state of not performing business such as resting or moving between waiting places due to insufficient data cannot be separated from the waiting state, so it is treated as a waiting state.

Figure 3.1. Taxi operations in Nagaoka City (Google Maps)

Figure 3.1 shows a partial screen of the time when the taxi was replayed using the data recorded in the recording device installed in the car. In addition to the actual taxi operation status shown in Figure 3.1 Active vehicle (green), empty vehicle (blue), and standby (black), there are payments, meals, replenishment, etc., but the data is the actual vehicle, empty vehicle, and standby.

3.1.1.1 Average Number of Rides Hourly

The colors of bars in a visualization of taxi demand by the day of the month can vary depending on the specific context and design choices. However, typically in such visualizations, the colors are used to represent different levels or categories of taxi demand. The specific color scheme used can vary depending on the purpose of the visualization and the preferences of the designer.

Figure 3.2 shows the average number of trips hourly and time zone from January 1, 2019, to December 31, 2019. The unit on the vertical axis shows the average number of rides in each time zone for each hour.

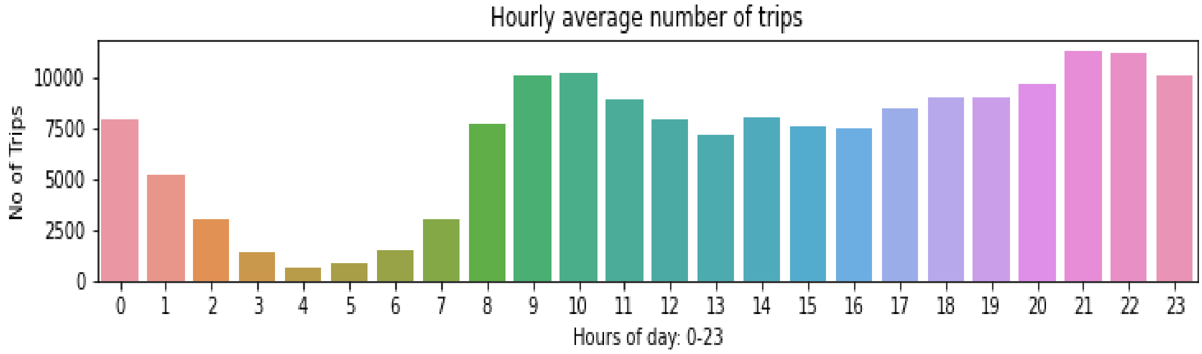

Figure 3.2. Hourly Number of trips

Looking at the time zone fluctuations from this figure, it was confirmed that there was a peak in demand from 9 am to 10 am. The number of operations after 10 o'clock.

It decreased and became the lowest in the afternoon. In addition, it was confirmed that the number of operations increased after 17:00 and there was a peak at 21:00. It is thought that there is a lot of demand because there is a rush hour for commuting and there are no buses after 9 pm.

3.1.1.2 Average Number of Trips Per Month

Figure 3.3 shows the average number of rides per month from January 1, 2019, to December 31, 2019. From Figure 3.3, the demand for January and August was the highest throughout the year, followed by March and December, and the lowest was May. The most common reason for January is that the snowfall has changed the use of bicycles and walking to taxis, and there are many year-end parties, so it is thought that taxi demand has increased more than usual. Since there are many farewell parties and welcome parties in March, taxi demand is higher than usual.

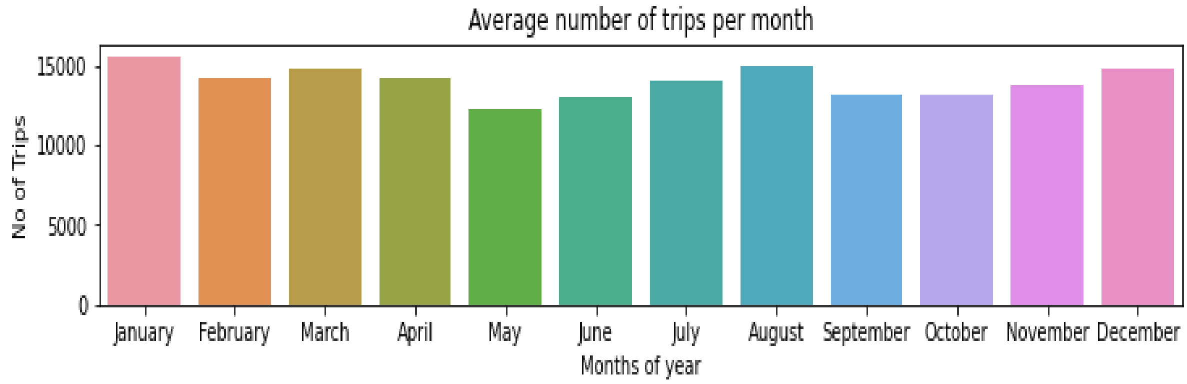

Figure 3.3. The average number of trips

Next, the month of May was the lowest, probably because there were many holidays during Golden Week, so the demand for taxis was the lowest. The reason why it seems that the average number of rides in August is higher than in July and September is that it is fireworks display in Nagaoka City on August 2nd and 3rd. It was confirmed that the demand for the day was high.

3.1.1.3 Average Number of Trips Weekly

Looking at the day of the week fluctuations, it was confirmed that the demand on Sunday was less than that on other days. It is thought that demand in the morning will decrease on holidays. The lowest demand in the morning is on Sunday, followed by Saturday see Figure 3.4.

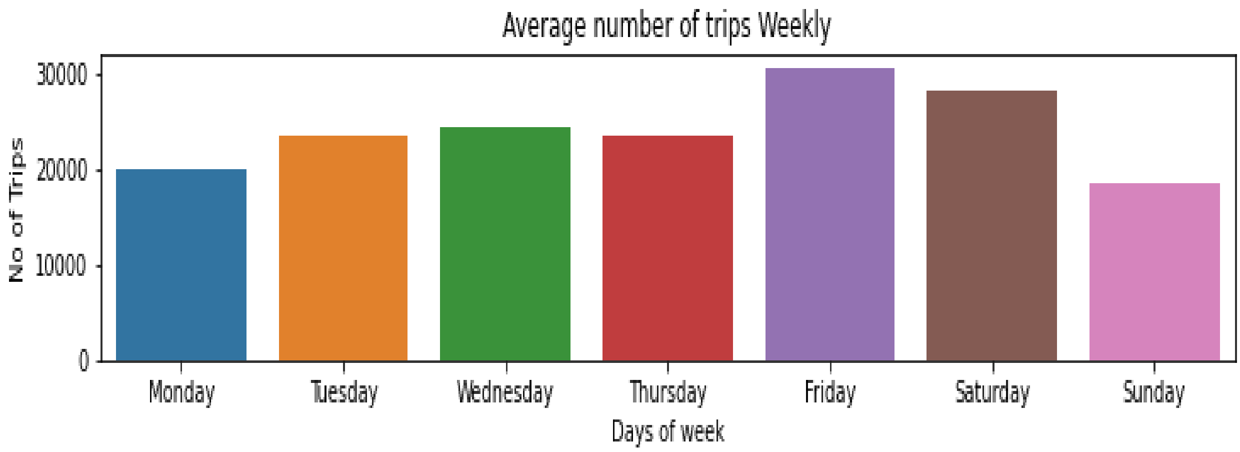

Figure 3.4. The average number of trips Weekly

It is thought that there will be fewer opportunities to go out on business trips and hospital visits. Looking at the time zone, it seems that the demand on Monday is less than the demand on other weekdays because there are many holidays on Monday.

3.1.1.4 Pickup Heatmap

Figure 3.5 shows that the demand was overwhelmingly higher than on other days of the week after 19:00 on Friday and Saturday and between 0:00 and 4:00 on Saturday and Sunday. Since there are many opportunities to eat and drink, it is thought that there is a lot of demand in the middle of the night on weekends. In addition, the demand after 19:00 between Monday and Saturday tends to increase depending on the day of the week.

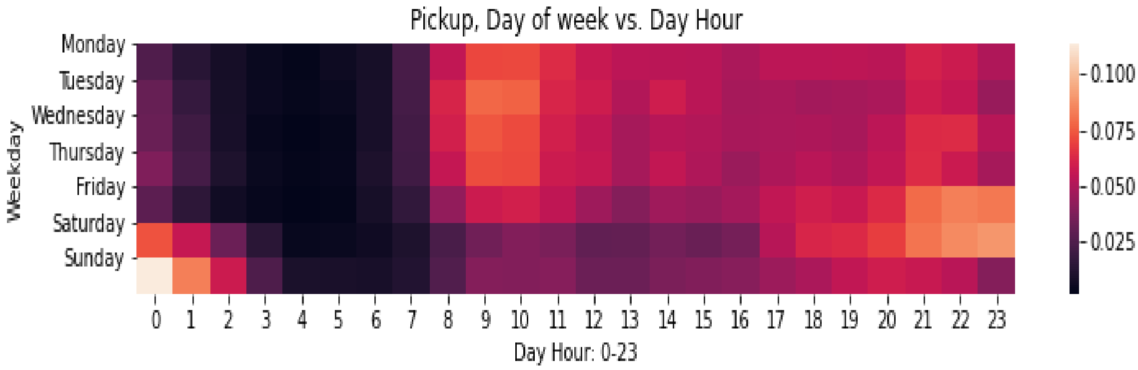

Figure 3.5. Pickup, Day of week vs. Day Hour

Figure 3.6 shows the average number of rides by month and day of the week from January 1, 2019, to December 31, 2019. Looking at the day of the week fluctuations, the most demanded are Friday and Saturday. The lowest demands are on Mondays and Sundays. The reason for the low demand on Sundays is thought to be that there are fewer opportunities to eat and drink in the middle of the night, and there are no opportunities to go out on things such as business trips.

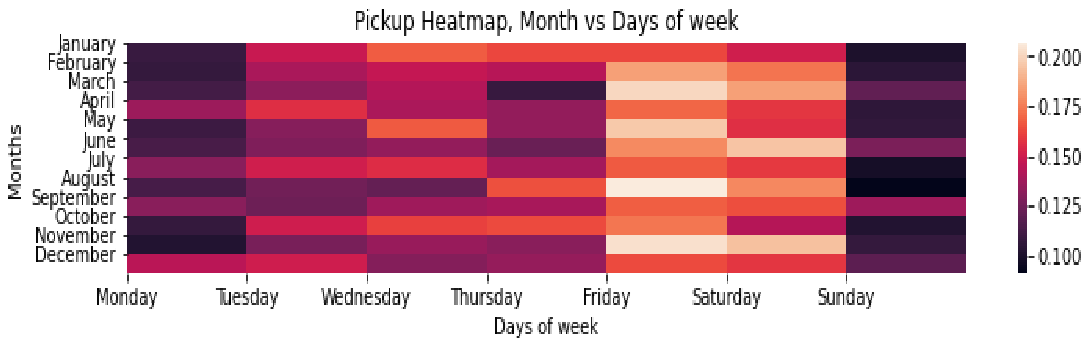

Figure 3.6. Pickup, Month vs Days of the week

Looking at the trends by looking at the monthly fluctuations, it was confirmed that there was a lot of demand in December. Time of day, it is thought that usage will increase on Fridays and Saturdays when there are many opportunities to eat and drink in the middle of the night. It is thought that there will be a lot of demand in December due to the cold December and the end of the year.

3.1.2 Sanjo City Taxi (Sharing) Data

Sanjo City is located in the central part of Niigata Prefecture and is composed of the plains formed by the Shinano River, Igarashi River, and Kariyata River and the hills and mountainous areas in the southeastern part. In addition, it is a town blessed with abundant nature and agricultural and forestry resources. The total area is about 432 km^2, the population is about 94,514, and the elderly population is about 30,000.[3])

The main public transportation that operates in Sanjo City is the widespread public transportation business throughout the city, including the shared taxi "Hime Sayuri" which is the subject of this study, the city circulation bus "Guruttosan", and the high school student school liner bus. The analysis period is one year from June 2015 (Table 3.2). The number of acquired data is 5039.

[3] sanjo city - demographic data / sanjo city

Table 3.2. Selected data (Sanjo city) for the study

Taxi Company	Pickup datetime	Pickup longitude (degree)	Pickup latitude (degree)	Dropoff longitude (degree)	Dropoff latitude (degree)	Dropoff datetime	Passenger count
Chuetsu Transportation	6/1/2015 7:20	138.8817	37.60497	138.9737	37.62851	6/1/2015 7:43	1
Chuetsu Transportation	6/1/2015 7:30	139.1317	37.5412	138.9717	37.63309	6/1/2015 8:18	1
Hinomaru sightseeing taxi	6/1/2015 7:50	139.0528	37.56554	138.9675	37.63287	6/1/2015 8:23	2
Hinomaru sightseeing taxi	6/1/2015 7:51	139.1178	37.50412	138.9717	37.63309	6/1/2015 8:29	2
Hinomaru sightseeing taxi	6/1/2015 7:55	139.0344	37.56621	138.9735	37.64175	6/1/2015 8:17	1
S taxi	6/1/2015 8:00	138.9613	37.6301	138.9546	37.64321	6/1/2015 8:08	1
Sanjo taxi	6/1/2015 8:00	138.9768	37.63565	138.9533	37.61257	6/1/2015 8:10	1
Hinomaru sightseeing taxi	6/1/2015 8:00	138.9328	37.64898	138.9717	37.63309	6/1/2015 8:10	2
Hinomaru sightseeing taxi	6/1/2015 8:01	138.9393	37.64862	139.0262	37.6448	6/1/2015 8:24	1
Hinomaru sightseeing taxi	6/1/2015 8:02	139.0034	37.59727	138.9889	37.60494	6/1/2015 8:06	1
Sanjo taxi	6/1/2015 8:05	138.9496	37.64019	138.9827	37.62533	6/1/2015 8:15	1
Hinomaru sightseeing taxi	6/1/2015 8:09	139.0503	37.56646	139.0287	37.57637	6/1/2015 8:15	1
Sanjo taxi	6/1/2015 8:10	138.8844	37.60965	138.9735	37.64175	6/1/2015 8:30	1
Sanjo taxi	6/1/2015 8:13	138.9393	37.64862	138.9506	37.65328	6/1/2015 8:17	2
Sanjo taxi	6/1/2015 8:15	138.9737	37.62851	138.9889	37.60494	6/1/2015 8:24	1
Chuetsu Transportation	6/1/2015 8:15	138.9512	37.6304	138.9717	37.63309	6/1/2015 8:22	1
Chuetsu Transportation	6/1/2015 8:20	138.9848	37.6233	138.955	37.63783	6/1/2015 8:29	1
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
Sanjo taxi	6/1/2015 8:30	138.9751	37.60675	138.9686	37.63008	6/1/2015 8:35	1
Sanjo taxi	6/1/2015 8:30	138.9737	37.62851	138.9619	37.63639	6/1/2015 8:35	1

3.1.2.1 Average Number of Trips Hourly

Figure 3.7 shows the average number of trips hourly and time zone in June 2015. The unit on the vertical axis shows the average number of rides in each time zone for each hour.

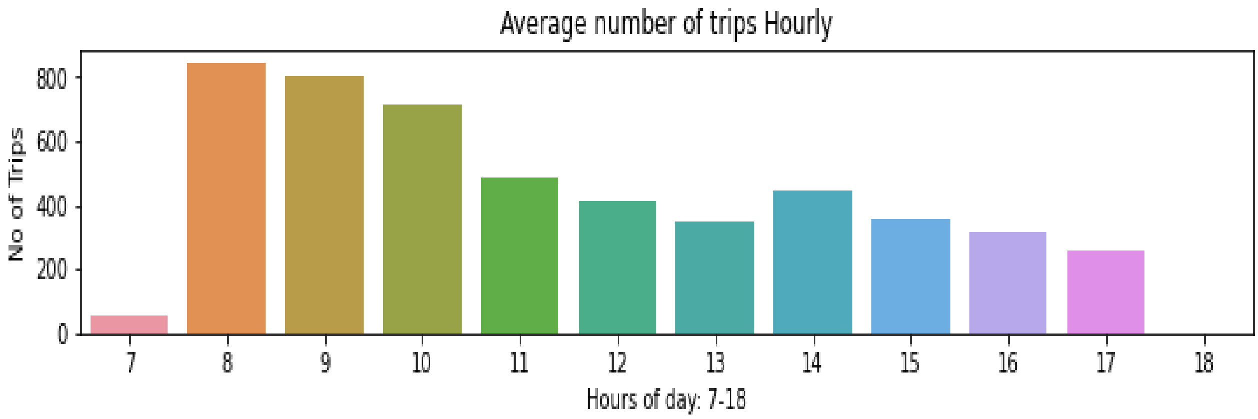

Figure 3.7. The average number of trips Hourly

Looking at the time zone fluctuations from this Figure 3.7, it was confirmed that there was a peak in demand from 8 am to 10 am.

3.1.2.2 Average Number of Trips Weekly

The day of the week variations (Figure 3.8), it was confirmed that the demand on Monday and Tuesday was greater than that on other days.

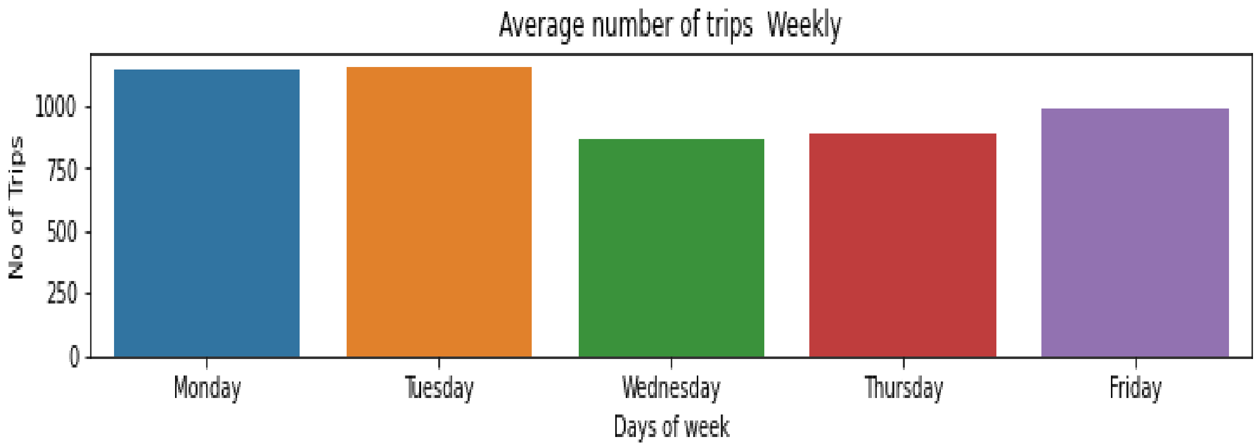

Figure 3.8. The average number of trips Weekly

3.1.2.3 Pickup Heatmap

In Figure 3.9, and Figure 3.10, at the day of the week variations, the most demanded are Monday and Tuesday. The lowest demand is on Wednesday.

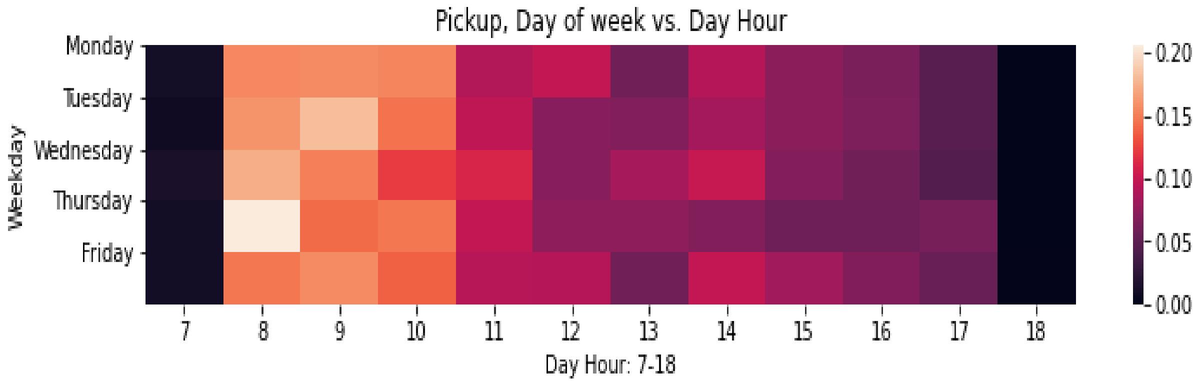

Figure 3.9. Day of week vs. Day Hour

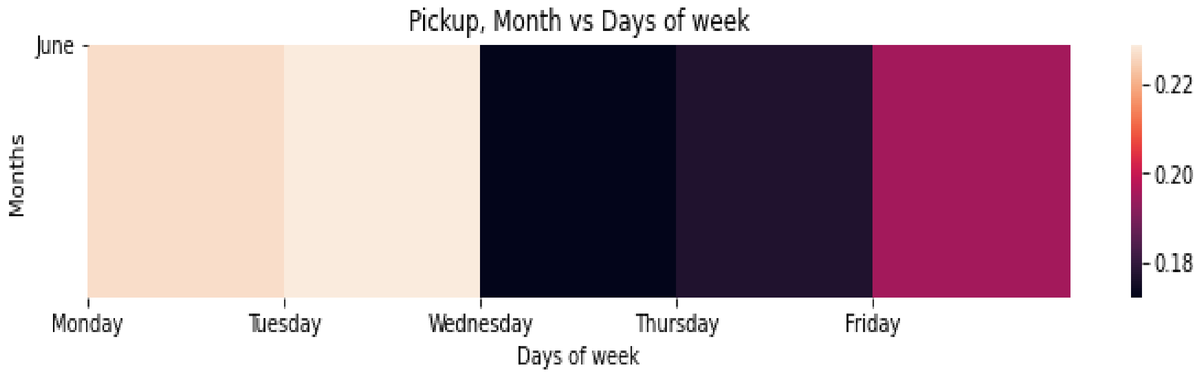

Figure 3.10. Month vs Days of the week

3.2 Weather Data Analysis

The urban traffic system is exceedingly complicated, time-dependent, and random since it incorporates people, cars, roads, the environment, and other complex components. Various external factors, like policy, the economy, and the weather, have a direct or indirect impact on the traffic system's operational performance. The link between meteorological conditions and passenger movement, or ridership, is gaining scholarly attention. Most of their research emphasizes the importance of weather conditions in varying the strength of various climatic measures such as temperature, wind speed, snowfall, and precipitation. This section will look at how the weather affects the taxi ride.

Table 3.3. Selected data for the study

Date/Time	Precipitation (mm)	Temperature (°C)	Wind speed / direction (m / s)	Snowfall (cm)	Snow cover (cm)
1/1/2019 0:00	0	0.7	0.5	0	6
1/1/2019 1:00	0.5	1.3	2.2	0	7
1/1/2019 2:00	1.5	1	2.3	0	7
1/1/2019 3:00	0.5	0.8	2.3	0	7
1/1/2019 4:00	1.5	0.5	1.9	0	7
1/1/2019 5:00	1	0.7	0	0	7
1/1/2019 6:00	0	1	1.3	0	7
1/1/2019 7:00	1.5	0.9	1.7	1	8
1/1/2019 8:00	0	1.2	3.4	0	7
1/1/2019 9:00	3	0.6	2.5	1	8
1/1/2019 10:00	0.5	1	2	0	7
1/1/2019 11:00	0	1.8	3.5	0	7
1/1/2019 12:00	0	2.4	3	0	6
1/1/2019 13:00	0	2.5	2.4	0	6
1/1/2019 14:00	3	1.7	1.2	0	6
1/1/2019 15:00	1.5	1	1.3	1	7
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
12/29/2019 21:00	1	0.7	1.1	0	6
12/30/2019 22:00	0.5	0.7	1	0	6
12/31/2019 23:00	0	0.8	0.1	0	6

In general, the impact of different weather conditions on road conditions is mostly based on two effects: one when the weather directly affects drivers' driving conditions, and the other traffic volume because the weather conditions affect a person's travel choice. Given the link between weather and road conditions, this study chose four typical meteorological variables to properly represent weather conditions: temperature, snow cover, wind speed, precipitation amount, and snowfall. Hence, the weather measurements are used as the principal data sources (Japan Meteorological Agency[4]) in this study to understand the weather condition. The data was collected for one year (from January 1, 2019, to December 31, 2019) (see Table 3.3).

[4] https://www.jma.go.jp/jma/indexe.html

Given the diversity of climate features among cities, this study focuses solely on general climatic changes and their impact on taxi movements, exposing the basic laws of transportation under present climate conditions. Therefore, some rare weather events in a particular city will not be considered in this study. Daily weather data used in this study were obtained from the Japan Meteorological Agency based on the 1-hour interval of ground observation with four major climate changes: temperature (F), wind speed (m / s), precipitation (mm), rainfall(mm) and snowfall (cm).

3.2.1 Impact of Weather

Figure 3.11 shows the frequency distribution of boarding time for each trip by weather. It can be confirmed that the number of operations, when it snows from 11 minutes, exceeds that of fine weather and rain. There are relatively few trips with a short boarding time and relatively many trips with a long boarding time when it snows. When the distribution of riding distance was confirmed, there was no significant difference due to the weather. It is presumed that the snowfall did not induce short-distance demand but that the vehicle speed decreased.

It can be seen that the curve when it rains and the curve when it is fine are almost the same, and it is considered that it does not affect the decrease in vehicle speed due to rain.

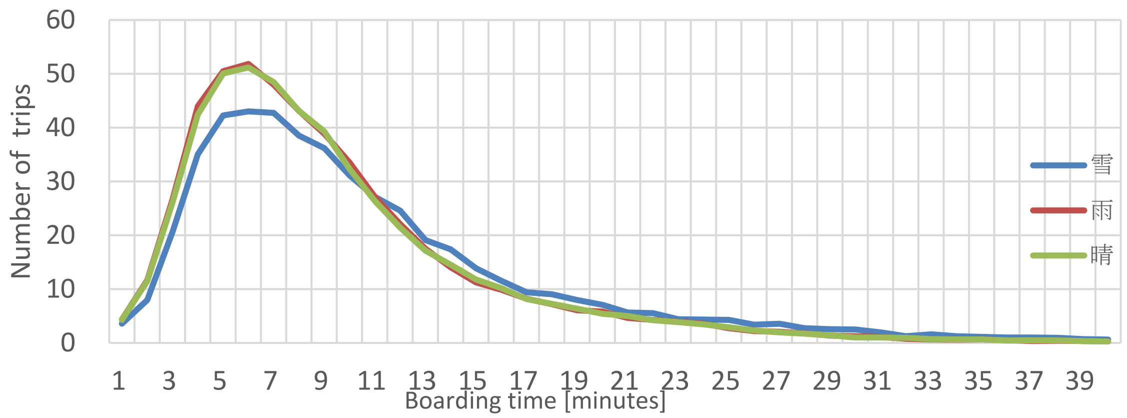

Figure 3.11. Number of trips by weather

3.3 Summary of the Chapter

In this chapter, we analyzed the actual operation by analyzing the GPS data of all vehicles and 42 cars belonging to Mitsukoshi Taxi Co., Ltd., which is the largest taxi company with a business area in the entire Nagaoka city. As a result, the following points were clarified.

- ✓ The peak demand in the morning is from 9:00 to 10:00.
- ✓ Weekday nighttime demand peaks from 21:00 to 22:00.
- ✓ Since there are many eating and drinking opportunities on Fridays, Saturdays, and Sundays, there is a lot of demand in the middle of the night.
- ✓ Throughout the year, demand was highest in January, followed by March and November, and the lowest was in May.
- ✓ There is a decrease in vehicle speed due to snowfall.
- ✓ Because there are many waiting times and breaks, the operation efficiency is poor.

However, the nature and extent of this relationship may vary depending on several factors, including the type of weather, the location, and the time of day. In general, it seems that extreme weather conditions, such as heavy rain, snow, or extreme heat, can lead to an increase in taxi demand as people are less likely to walk or use other modes of transportation. On the other hand, mild weather conditions may not have as significant an impact on taxi demand.

Regarding shared taxis specifically shared taxi demand may be more sensitive to weather conditions than general taxi demand. This is because shared taxis often follow fixed routes and schedules, which means that they may be more affected by disruptions caused by weather conditions.

Therefore, while weather conditions do seem to play a role in determining taxi demand, other factors such as traffic patterns, events, and the availability of alternative transportation options may also have an impact.

Chapter 4. DEMAND ESTIMATION

This chapter introduces machine learning methods, which are particularly useful for time series modeling and forecasting. Machine learning approaches are considered to be a superior method for predictive problems, including linear regression, random forest, gradient boost, decision tree, and hybrid machine learning, concepts, structures, and formulations to final. We've discovered that combining these mathematical forecasting approaches applied in the real-world taxi sector may drastically cut fuel costs, customer waiting times, and empty taxis if we can properly anticipate taxi demand. However, several problems remain beyond this encouraging study. The knowledge of driver-passenger demand is always changing. To forecast demand on a real-time basis, which model or technique should we use? What if the model you choose isn't suitable? How can you build a strong and dependable system that can locate or even always produce the best appropriate models? We'll deal with these issues in this chapter.

4.1 Introduction

To guarantee effective and secure means of transportation, modern transportation systems worldwide, such as taxis and trains, have been employing a varied spectrum of information technology-based applications. On the other hand, typical taxi systems often experience inefficiencies due to a lack of coordination when consumer demand shifts. Drivers are more likely to focus on probable destinations based on their job experience, Zheng *et al.* (2018) which is referred to as transfer to such a location based on a reward function. This situation leads to underserved areas where passengers have to wait long to be served and oversupply where vehicles exceed the demand (Huang and Powell (2012)). The disparity in taxi supply and demand causes uncomfortable waiting for passengers and creates useless waiting time for organizational resources such as drivers. Due to the lack of customer information, they are unaware of nearby areas with high taxi demand, and sometimes, drivers cannot pick up any customers. The forecast for taxi demand can improve the focus on the taxi-service ineffectiveness problem. Under such circumstances, the 'Taxi Demand Prediction Problem', an elementary requirement, yet not fully apprehended to achieve efficiency has attracted more attention recently.

Furthermore, traffic is the pulse of a city that impacts the daily lives of millions of people. One of the most fundamental questions for future smart cities is how to build an efficient transportation system. A critical component to addressing this question is to create an accurate taxi demand prediction model. A better way to predict demand for travel and pre-allocated resources is to adapt to the taxi demand in the city while avoiding excessive energy consumption. Moreover, with the increasing popularity of taxi requesting services, it is essential to collect massive demand data unprecedentedly. In our study, we chose 15 minutes. The drop-off points are generally within 15 minutes, and the boarding and alighting facilities were identified assuming that the target facility was used within 15m from the outer road of the target facility in Nagaoka City, Japan (Sano *et al.* (2020)).

So, there are still specific research gaps in previous works. The proposed machine learning (ML) model is to develop a Hybrid Machine Learning (HML) model for taxi prediction. Others ignore the usefulness of external meteorological features, as in the work of (Askari *et al.* (2020); Zhao *et al.* (2016)). The approach to predicting is to maximize the impact of prospective pickup sites, also known as points of attraction, such as pubs, restaurants, shopping malls, and so on, on the idea that taxi demand would vary from that of other regions (Askari *et al.* (2020); Vanichrujee (2017)).

A novel HML model combining LR, DT, RF, and GB algorithms was developed in this work to predict and evaluate the significant performance of taxi demand. Compared to the ML model using a single algorithm, the hybrid model combining four different algorithms can enhance the robustness and generalization ability of the ML model for fast and effective predictions. In addition, an HML algorithm has been used to determine the hyper-parameters of each algorithm (LR, DT, RF, and GB) to enhance the adaptive ability and increase the accuracy of the prediction.

Finally, we have conducted prediction and comparison tests based on a real-world travel demand dataset from Nagaoka City taxis. Numerical results show that HML outperforms benchmark models such as LR, DT, RF, and GB.

To summarize, the main address in this chapter is as follows:

1. We have proposed a new HML model to predict and evaluate the critical performance of taxi demand. The HML methods could effectively the features to reduce errors.

2. We have provided a 15-minute demand prediction period that is acceptable and sufficient for drivers and contact centers to determine what to do when passengers are dropped off.

3. We have examined several ML models to decide which one is the best based on prediction accuracy and data processing time. This data may be used to create an intelligent system that can govern and coordinate taxis on a broad scale, benefiting taxi drivers and businesses. Taxi drivers may travel to high-demand regions to accommodate customer demand, and ride-sharing companies can re-allocate their vehicles utilizing surge pricing in advance.

Other than machine learning approaches, according to Liao *et al.* (2018), the deep neural network is the most powerful. As a result, in this study, we use Nagaoka City raw data to quantify taxi demand by incorporating weather-related factors.

4.2 Demand Forecasting Using ML Algorithm

The focus of the chapter is to determine the future demand for taxi systems as a percentage of the present regular taxi demand. The methods of the study are created model testing under five ML techniques, namely, linear regression, decision tree, random forests, gradient boosting, and

Hybrid Machine Learning method, to ensure robust model creation and assessment.

One of the most often used classification algorithms in machine learning is the K-nearest neighbor (KNN) method. It is also one of the easiest ML algorithms. It can be used for classifications and regression tasks. KNN learning involves three main hyper-parameters: the number of neighbors k, the distance function, and the weighting function (for details, see Kang (2021); Yu *et al.* (2016)). There are two things to know about KNN. Firstly, a non-parametric algorithm means that no statements about the dataset are produced when the model is used. Secondly, when the KNN is used, the dataset is separated into train and test stage sets. So, all the training data is also utilized when the model is required to make predictions.

k -means the cluster is the process of grouping data points into numerous groups such that data points with similar qualities are grouped. In the literature, several clustering techniques have been published and widely utilized, including the partition algorithm, hierarchy algorithm, mass-based, diagram-based, and model-based approaches. The partitioning method generates clusters, but most algorithms automatically estimate the number of k cluster centroids depending on specified characteristics, where k is a user-specific number of clusters. Because we have specified limitations such as a minimum degree of customer service, the partitioning technique is a better fit for the suggested research. It would be advantageous to change the number of clusters to satisfy these constraints. Our study suggests a different of the k-means algorithm, for example, which is a partitioning algorithm (see details Frigui (2008)). The cluster is divided by the taxi demand (see Figure 4.1)

The centroid-based clustering algorithms can be implemented by following steps:

- Select k points at random as centroids of the created cluster dividers,
- Built on these centroids, allocate each point to a particular cluster,
- Assign data points to the closest cluster based on calculating the Euclidean distance from the assigned point to the centroids and specific points to the clusters where the distance is minimum.
- After the assigned points have been fixed to the clusters, find the new centroid of the cluster.

So, repeat k means multiple times with different initializations. The k means cluster selection was done based on the shortest cumulative Euclidean distance.

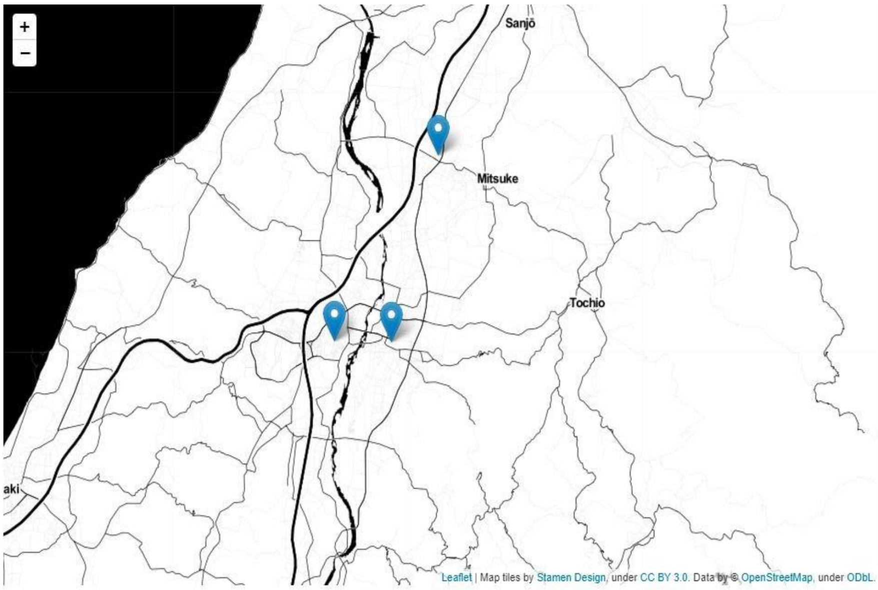

Figure 4.1. Nagaoka city can be divided into 3-means clustering.

Given a set of observations $(x_1, x_2, x_3 \dots\dots\dots\dots\dots. x_n)$ where each observation is a d-dimensional vector, k means clustering algorithm aims to divide the n observations into a set of k groups $k \leq n$ such as $G = G_1, G_2, G_3 \dots\dots\dots\dots\dots. G_k$, to reduce the within-groups sum of distance squares, which is defined as the sum of distance functions from each point in the groups to the appropriate center. The following is the objective function of k means: G_i Li *et al.* (2021).

$$\arg\min \sum_{i=1}^{k} \sum_{x \in G_i} \left\| x_i - c_i \right\|^2 \qquad (4.1)$$

4.2.1 Linear Regression (LR)

LR is applied to predict a nominal dependent variable, given several independent variables. We determine the possibility that the demand of the output variable corresponds to the relevant group using LR. Furthermore, LR may be utilized to determine the domain's solution. In ML, LR is appropriate for prediction problems ML (Zantalis *et al.* (2019)). A simple LR formula:

$$y = \beta_0 + \beta_1 x \qquad (4.2)$$

Where y is the dependent variable, x is the independent variable, β_0 is the model's intercept and β_1 is the slope.

A vector Autoregression Sujath *et al.* (2020) is a prediction of two-time series. The following formula

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + \varepsilon_t \qquad (4.3)$$

Where β_0 is the model's intercept, $\beta_1, \beta_2, \beta_2, ... \beta_p$ are the coefficient of the legs of Y till over p. Order 'p' means, up to p-lags of Y is applied and they are the predictors in the equation. The ε_t is error considered white noise.

The goal of LR is to find the line of best fit or the equation that best describes the relationship between the dependent and independent variables. LR is a simple yet powerful ML model that can be used to analyze a wide range of data sets. It is particularly useful for understanding the relationships between variables in continuous data sets, where the dependent and independent variables can be quantified and modeled using mathematical equations.

4.2.2 Decision Tree (DT)

The DT approach is a supervised ML method that constantly divides data according to a parameter. Two things describe the tree: decision nodes and leaves. The choices or outcomes where the data is divided are the leaves. By maximizing the separation of the data, this technique repeatedly breaks the data set into a tree-like structure (for details, see Dreiseitl and Ohno-Machado (2002)). The greedy construction process gives a major disadvantage to DT: the combination of the single best variable and optimal split-point is selected at each step. However, a multi-step looks ahead that considers variables' combinations may obtain more outstanding results. Entropy is an amount of confusion or ambiguity and the goal of DT.

The mathematical formula for entropy is:

$$E(s) = \sum_{i=1}^{c} -p_i \log_2 p_i \text{ , Where } p_i \text{ is simply probability of } i \qquad (4.4)$$

So, decision trees can be disposed to overfitting, which means that they can be too complex and capture the noise in the data, resulting in poor generalization performance on unseen data. To address this, techniques such as pruning, ensemble methods, and limiting the maximum depth of the tree can be used. Decision trees are a powerful tool for the prediction of various aspects of taxi operations and have been widely used in literature. It is important to carefully consider the limitations and potential for overfitting when using decision trees for prediction.

4.2.3 Random Forests (RF)

RF is an ensemble algorithm to estimate the outcome class by parallelly training processes on numerous DTs. RF is an ensemble of classifiers composed of DTs generated using two different sources of randomization. First, each decision tree is trained on a casual sample with replacement from the real data with the equal size to the given training set. The created bootstrap samples are likely to have approximately some duplicated instances. The second source of randomization applied in the random forest is attribute sampling. A subgroup of the input variables is randomly chosen at each node to search for the best split. For classification, the final estimate of the ensemble is given by margin voting (see details in Bentéjac *et al.* (2021); Breiman (2001); Sarica *et al.* (2017); Fitria (2013)).

RF algorithm:

Step 1: Pick the random K traffic data from the training data,

Step 2: From the data, construct a decision tree,

Step 3: Before repeating steps 1 and 2, choosing the number of trees to build,

Step 4: Find the predictions of each decision tree,

The general mathematical formula for an RF classifier Shah *et al.* (2020) is as:

$$n_{ij} = w_{left}c_j - w_{left(j)}c_{left(j)} - w_{right(j)}c_{right(j)} \tag{4.5}$$

Where, n_j = the significance of node j

w_j = weighted number of samples reaching node j

c_j = the impurity value of node j

$left(j)$ = child node from left split on node j

$right(j)$ = child node from right split on node j

The random forest algorithm trains multiple decision trees driven by a different subset of data. Using the random forest classification approach, a large number of classifiers are created from smaller subsets of the input data, and then the individual findings of each classifier are combined based on a voting process to produce the desired output of the input data set. The conceptual diagram for the random forest approach is shown in Figure 4.2. From a particular total training set, n sub-training sets are first randomly chosen. In this context, a decision tree is referred to as a sub-training set. Although the processing of the sub-training set is identical to that of a typical decision tree, the application of available variables takes unpredictable into account.

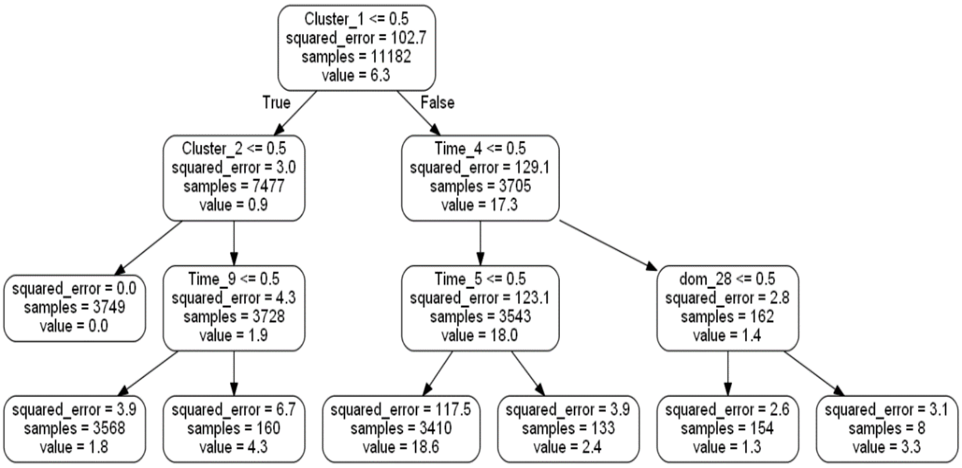

Figure 4.2. Classification of the random forest

4.2.4 Gradient Boosting (GB)

GB develops the tree model and updates it by lowering the expected value of the specified loss function. It can be used with other models that minimize the average of a loss function on training data, such as a square error or a complete error. The loss function calculates the disparity between the expected and actual values. The boosting approach is an "effective gradient descent" in regression issues. It is an optimization approach that reduces the size of a loss function by adding a base model to each step that reduces it. The final prediction of the demand class is the weighted voting of individual DTs; (for details, see Fitria (2013); Biau *et al.* (2019)).

The GB is to build a base model to predict the observations in the training dataset. Mathematically, the average of the first target column is

$$F_0 = \underset{\gamma}{\arg\min} \quad \sum_{i}^{n} L(y_i, \gamma)$$

Where L is the loss function, γ is the predicted value, y_i is the observed value.

The loss function is

$$L(y_i,\gamma)=\frac{1}{n}\sum_{i}^{n}(y_i,\gamma)^2$$

The following step of gradient boosting is to build a model:

Step1: To build a model to predict the training dataset;

Step 2: To calculate residual = (observed value – predicted value);

Step 3: To minimize the residuals and improve the model accuracy;

Step 4: To find the output values for each leaf of the decision tree;

Step 5: Finally, update the predictions of the previous model.

The gradient boosting can be computationally expensive, especially for large datasets, as multiple weak learners need to be built and combined. Additionally, gradient boosting can be less interpretable compared to decision trees, as the predictions are based on the combination of multiple weak learners rather than a single tree.

4.2.5 The Proposed Hybrid Machine Learning (HML)

HML is a combination of two or more ML algorithms that are used together to make predictions. In the context of taxi prediction, HML could involve combining multiple algorithms such as linear regression, decision trees, random forests, and gradient boosting to make more accurate predictions about taxi demand, supply, and trip duration. In HML models, firstly the LR model is used for estimating the association between the dependent and independent variables statistically. And then the RF is an ensemble learning method that constructs several decision trees at randomly selected features and predicts the class of a test instance by voting for the individual trees. The proposed HML defines a general framework for trees as a baseline model, such as a linear regression model. In order to tree baseline model, nonlinear models such as decision tree, random forest, and gradient boosting or combined models can be applied.

Figure 4.3 shows the conceptual diagram of an HML framework that combines the based dataset clustering technique and the based classification technique. First, clustering creates groups (i.e., clusters) and provides them as dependent variables for classification.

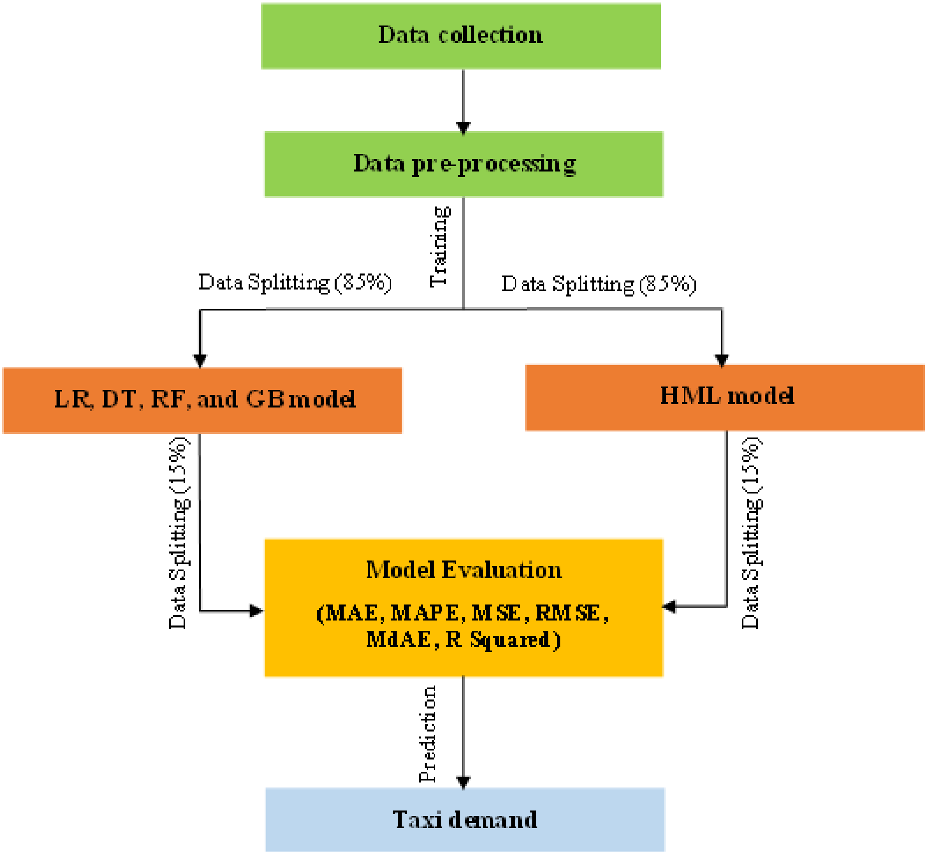

Figure 4.3. A logical framework for the prediction and evaluation using HML.

The HML framework generates six stages conducted in sequence. To use hybrid machine learning for taxi prediction, the following steps are as follows:

1. **Data collection and preparation:** The first step is to collect and prepare the relevant data for the prediction. This data includes information such as time of day, weather, location, and previous demand patterns. The data has been cleaned, normalized, and transformed as necessary.

2. **Feature engineering:** Next, the relevant features that will be used for prediction should be identified and extracted from the data. This involves creating new features based on the existing data or selecting the most relevant features through feature selection techniques.

3. **Model selection:** The next step is to select the machine learning (LR, DT, RF, and GB) algorithms that will be used in the hybrid model. This involves considering the strengths and

weaknesses of different algorithms, as well as the nature of the data and the problem being solved.

4. **Model training:** Once the algorithms have been selected, they should be trained on the prepared data. This involves splitting the data into training and testing sets, training each algorithm on the training data, and evaluating their performance on the testing data.

5. **Model combination:** After each algorithm has been trained, they should be combined into a single hybrid model. This can be done through a variety of techniques, such as weighing the outputs of each algorithm, combining the outputs through a meta-model, or using a stacking approach where the outputs of each algorithm are used as inputs to another algorithm. A hybrid model involves using the predictions from one or more models as input to another model. There are several ways to combine models, including stacking, blending, and boosting.

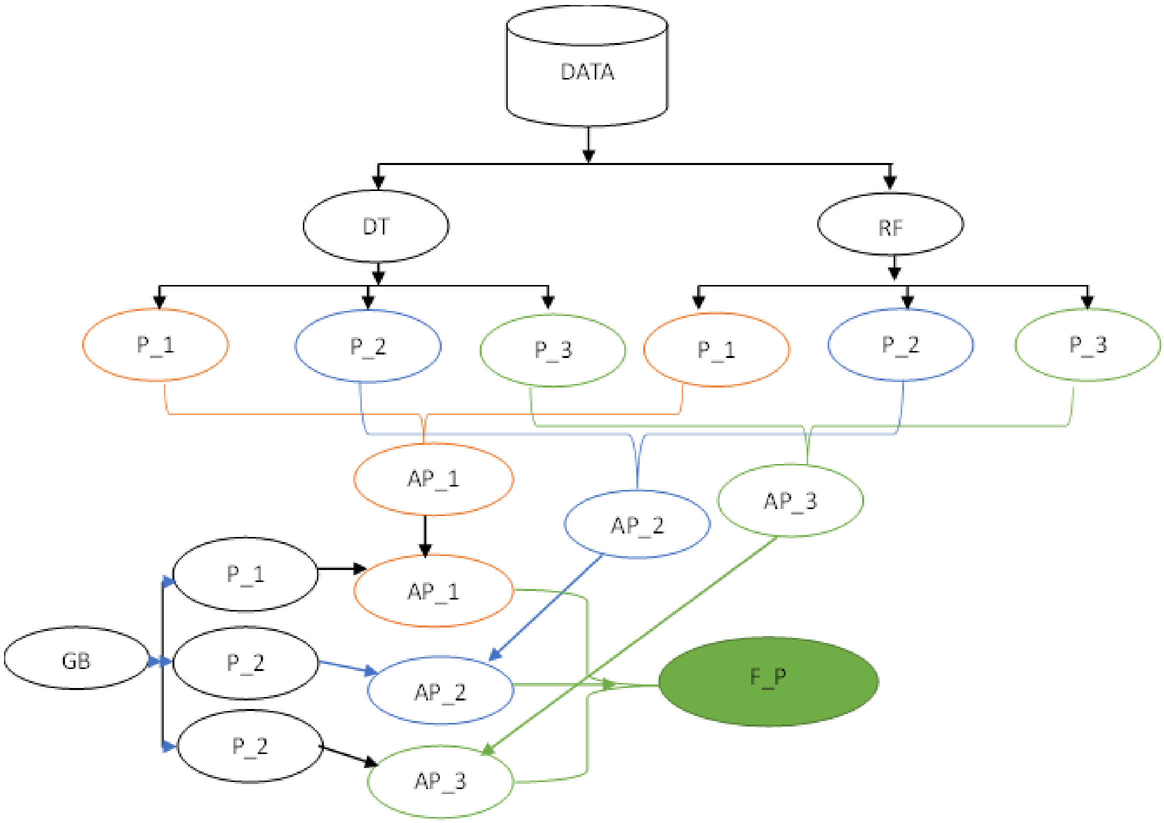

Figure 4.4. HML architecture

Stacking: In stacking, the predictions of the individual models are combined as input to a meta-model, which makes the final prediction. The meta-model is trained on the predictions of the component models, instead of the original features. This technique can be used to overcome the weaknesses of individual models, by allowing the meta-model to learn from the predictions of multiple models. We use the output of the decision tree model as input to the random forest model and then use the output of the random forest model as input to the gradient boosting model. The gradient boosting model makes the final prediction seen in Figure 4.4.

For example:

```
from sklearn.ensemble import StackingRegressor
baseEstimators = [
    ('DTR', DecisionTreeRegressor()),
    ('RFR', RandomForestRegressor())
]
hybridModel = StackingRegressor(
    estimators=baseEstimators,
    final_estimator=GradientBoostingRegressor()
)
```

6. **Model evaluation:** Finally, the hybrid model should be evaluated to determine its performance. This involves comparing the prediction accuracy of the hybrid model to the accuracy of each algorithm and other prediction models.

The use of hybrid machine learning for taxi prediction is an iterative process, and the steps may need to be repeated multiple times to fine-tune the model and achieve the best possible results. In this study, the demand for taxis has been modeled using a different hybrid method. The models combine multiple trees such as a LR, DT, RF, and GB to make predictions.

4.2.6 Hyperparameters Selection

Hyperparameter tuning is a very important step of ML to improve the performance of algorithms. And hyperparameter is also a component in increasing the model's accuracy. In linear regression, common hyperparameters include the regularization parameter, which determines the amount of reduction applied to the coefficients, and the choice of the loss function, which defines the objective to be optimized during training.

In decision trees, hyperparameters include the maximum depth of the tree, the minimum number of samples required to split a node, and the minimum number of samples required to be at a leaf node with 5-fold cross-validation (CV) in Table 4.1.

Table 4.1. Hyperparameter of decision tree

Hyperparameter	Assigned Value
max_depth	5
max_features	auto
max_leaf_nodes	50
min_samples_leaf	2
min_weight_fraction_leaf	0.1
splitter	random

For random forests, hyperparameters include the number of trees in the forest, the maximum depth of the trees, and the number of features to consider when splitting a node. A randomized search method with 5-fold cross-validation was used herein for exploring possible values of hyperparameters using the Scikit-learn package in Python in Table 4.2.

Table 4.2. Hyperparameter of Randomized search

Hyperparameter	Assigned Value
random state	1
no of iteration	10
cross-validation	5
no of jobs	-1

In gradient boosting, hyperparameters include the number of trees, the learning rate, and the maximum depth of the trees along with the 5-fold CV (Table *4.3*).

Table 4.3. Hyperparameter of GB

Hyperparameter	Assigned Value
learning_rate	0.05
max_depth	2
min_samples_leaf	5
min_samples_split	10
n_estimators	1000

For hybrid machine learning, the hyperparameters would depend on the specific models being combined in Table 4.4. For example, if a linear regression model and a decision tree model are being combined, the hyperparameters for each model would need to be set and optimized separately.

Table 4.4. Hyperparameters for HML

Hyperparameter	Assigned Value
n_estimators	1000
learning_rate	0.05
max_depth	2
min_samples_leaf	5
min_samples_split	10
min_weight_fraction_leaf	0.1
random_state	1
cross-validation	5
Max features	7

Additionally, the choice of how to combine the models and any hyperparameters related to this combination would also need to be set.

4.3 Baseline

We compare the HML model:

- HML is the model that combines multiple ML algorithms to generate more accurate predictions. In contrast, LR, DT, RF, and GB are all individual ML algorithms that are used on their own or combined with other algorithms to create hybrid models.
- LR is a statistical approach that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. It is a simple, straightforward approach that is well-suited to continuous data sets where the relationships between variables can be quantified and modeled using mathematical equations.
- DTs are the ML models that use a tree-like structure to model the relationships between variables. They work by dividing the input data into smaller and smaller subsets, making predictions based on the data in each subset, and building up a tree-like structure that models the relationships between variables. DTs have been used for both continuous and categorical data sets and are particularly well-suited for use in complex data sets with many variables.
- RFs are machine learning models that build multiple decision trees and combine their predictions to generate more accurate results. RFs work by randomly selecting subsets of the input data to use for building each DT and combining the predictions from all of

the trees to generate a final prediction. This approach helps to reduce the risk of overfitting, where a model becomes too closely tied to the input data, and can lead to more accurate predictions.

- GB is a machine learning algorithm that uses a combination of decision trees and gradient descent optimization to generate more accurate predictions. GB works by building a series of decision trees, where each tree is designed to correct the errors made by the previous tree. This approach helps to produce more accurate predictions by combining the strengths of multiple decision trees into a single, more powerful model.

The HML, LR, DT, RF, and GB models are all ML algorithms with their own strengths and weaknesses. The choice of which algorithm to use has depended on the particular data set being analyzed, the goal of the analysis, and the available resources and expertise. By combining multiple algorithms, HML models have offered a more powerful and flexible approach to data analysis and prediction and helped to overcome the limitations of individual algorithms.

4.4 Experiments

This section will analyze the actual situation of taxi operations. Firstly, the dataset used in the experiments. Then, analysis is done on the variables affecting how well taxi demand forecasting performs. The actual situation of taxi operations varies widely depending on the location, market conditions, and local regulations. However, there are some general trends and challenges that are common to many taxi operations around the world.

Taxi companies have needed to invest in technology and data systems to stay competitive, but this can also be a significant expense that can impact on their bottom line. Despite these challenges, there are also many opportunities for taxi operations to improve their performance and increase their competitiveness. There is also growing demand for services such as shared taxis and dynamic pricing, which can help taxi companies to better match supply and demand and improve their overall cost efficiency.

The actual situation of taxi operations is complex and dynamic and is impacted by a range of factors including competition, technology, and regulation. Taxi companies need to stay informed about the latest developments in the industry and be prepared to adapt and innovate in order to remain competitive. The experiments utilized the GPS data of 42 vehicles belonging to Mitsukoshi Taxi Co., Ltd., Nagaoka, the largest taxi company with a business area in the entire Nagaoka city, Japan (see Figure 4.5). Ride-related data (e.g., pickup and drop-off locations, time of the day) are from the taxi company, and environment-related factors (e.g., temperature and presence of snow) are derived from the commercial weather service provider dataset. Nagaoka is in the center of Niigata and the nearby Chūetsu region of Japan (see Figure 4.5). As of August 4, 2021, the city had an estimated population of 264,611 in 109,283 households and a population density of 300 inhabitants per square kilometer (780 sq/mi). The city's total area was 891.06 square kilometers. Well-established machine learning (ML) algorithms are developed to predict demand.

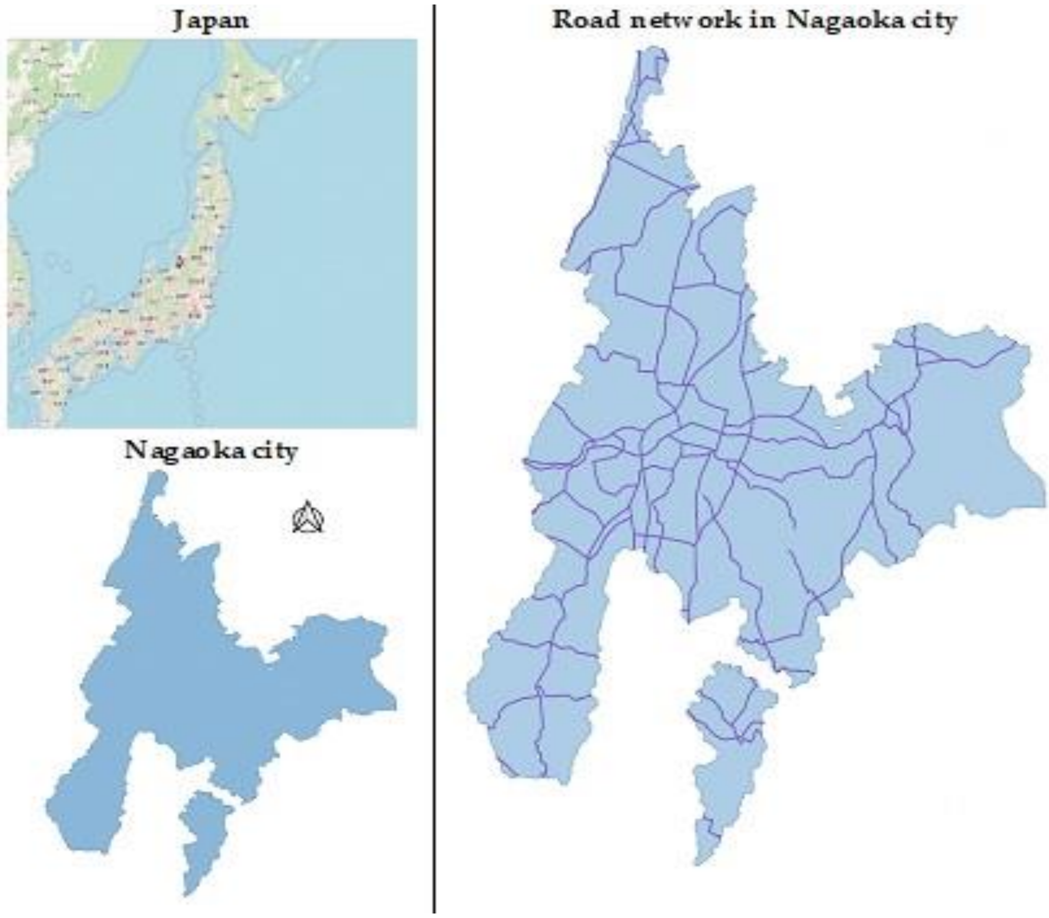

Figure 4.5. Location of the case study area: Nagaoka City, Japan.

The Nagaoka City Taxi collects information on individual taxi journeys. To estimate the demand, we used about 168156 individual taxi ride data for whole days in 2019. Each record includes the following field of core features (the meter was engaged or disengaged):

- Pick up date & time;
- Pick up longitude (unit: degree);

- Pick up latitude (unit: degree);
- Drop off date & time;
- Drop off longitude (unit: degree);
- Drop off latitude (unit: degree);

In order to expand the forecast accuracy, taxi demands are also dependent on weather conditions as the number of taxi users increases in poor weather conditions. Weather data is obtained from the Japan Meteorological Agency provided by the past amount of precipitation in Japan, (Japan Meteorological Agenc, 2019). We chose the following characteristics from the raw data, which comprise more than 19 fields:

- Date & time: e.g., 2019-01-01 11:00:00;
- Every 15 minutes temperature (unit: degree centigrade);
- Every 15 minutes wind speed (unit: m/s);
- Every 15 minutes precipitation (unit: mm);
- Every 15 minutes snowfall (unit: cm);
- Every 15 minutes snow cover (unit: cm);

The predictors and outcome variables are required for developing ML models which are prepared by gathering the fields available in the raw trip data. Firstly, to enhance the usability and generalizability of the prediction models, the pickup coordinates (latitude and longitude) are partitioned into different regions, also referred to as location ID, using a clustering algorithm. Specifically, the means of clustering process was selected for this research to control the number of clusters (i.e., the number of regions). Moreover, the pickup date is used to derive ride-related temporal factors - month, every 15 minutes of day indicator.

Table 4.5. Description of features.

Input Features	**Type and Description**
Month	Categorical: Jan, Feb,⋯, Dec.
Day of month	Categorical: #1, #2,⋯, #31.
Day of the week	Categorical: Monday, Tuesday,⋯Sunday.
Time of day (a day range (0, 23) with step 1 where each step is 15 minutes)	Categorical: #1, #2,⋯, (in mints)
Traffic Volume	Continuous
Region ID	Categorical: #1, #2, #3
Temperature	Continuous
Wind speed	Continuous
Precipitation	Continuous
Snowfall	Continuous
Snow Cover	Continuous

Table 4.6. Outcome variable.

Output Feature	**Type (Categorical)**
Demand	Continuous

In addition to these predictors, weather-related (or environmental) information, such as temperature, weather condition (snow, normal, etc.), and wind speed, are extracted for every 15 minutes of the day and each date in the set of data using the Japan Meteorological Agency provider. The output variable is the demand for the taxi and is obtained by aggregating and binning the number of passengers transported for each location, date, and day of the week. We have calculated the taxi demand every 15 minutes in each region. A summary of the input features is presented in Table 4.5.

The unified dataset contains inconsistencies, errors, and missing values and is typical for any real-world case. In particular, we observe several discrepancies in the data with zero, negative or irrational values in trip duration and travel distance. Likewise, in some cases, the longitude and latitude were recorded as 2,035. Remove 23 samples that are the outlier, i.e., has (longitude, latitude): (2,035, 2,035). As a result, the set of data is cleaned before being input into the ML model. After cleaning data, we were able to estimate the taxi demand depending on the weather.

Each of the levels associated with a categorical variable is converted using a one-hot encoding technique to enable the training and testing process of the ML model. For instance, the variable “Month” has 12 levels, and each level is encoded as an independent variable (e.g., encoded feature “Month” Jan = 1 if the trip is made in January and 0 otherwise). On the other hand, the continuous output variable is demand.).

4.5 Performance Metrics

LR, GB, DT, RF, and HML are compared using various performance parameters of taxi prediction such as mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), Median Absolute Error (MdAE), and R squared (R^2) in identifying the best ML model. The formulas for performance parameters are given below:

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i \in n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \tag{4.6}$$

MAE is a commonly used evaluation metric in machine learning. It measures the average magnitude of errors in a set of predictions, without considering the direction of the errors. The reason for using MAE as an evaluation metric is that it is relatively easy to understand and interpret, and it provides a simple way to compare the performance of different models. MAE is particularly useful in situations where the target variable has a linear relationship with the predictors, and where the magnitude of errors is more important than the direction. Another advantage of using MAE is that it is easy to optimize using gradient descent algorithms, as it has a continuous and differentiable objective function. This makes it a good choice for many optimization problems in machine learning, such as linear regression and neural network training.

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100}{n}\sum_{i \in n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \tag{4.7}$$

MAPE is another commonly used evaluation metric in machine learning. It measures the average percentage difference between the predicted and actual values of a target variable.

MAPE is particularly useful when we want to measure the performance of a model in terms of its ability to make accurate predictions on a relative scale, rather than an absolute scale. MAPE is commonly used in forecasting models, where we want to predict the future values of a variable based on its historical values. In such cases, MAPE can provide a useful measure of the model's accuracy in terms of percentage error. One advantage of using MAPE is that it is a relative measure of error that is scale independent. MAPE is also easy to understand and interpret, as it represents the average percentage error of the predictions. So, one potential drawback of using MAPE is that it can be sensitive to extreme values or outliers in the data. This is because the percentage error can become very large when the actual value is close to zero. As a result, some practitioners prefer to use other metrics like MAE or MSE instead of MAPE.

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i\in n}(y_i - \hat{y}_i)^2 \tag{4.8}$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i\in n}(y_i - \hat{y}_i)^2} \tag{4.9}$$

MSE and RMSE are two commonly used evaluation metrics in machine learning. Both metrics are used to measure the average magnitude of errors in a set of predictions, and they are particularly useful when the direction of errors is important. MSE is calculated by taking the average of the squared differences between the predicted and actual values of a target variable. RMSE is simply the square root of MSE. Both metrics provide a measure of the average squared error of the predictions, with RMSE providing a more interpretable measure of the error in the original units of the target variable. MSE and RMSE are commonly used in regression problems, where the goal is to predict a continuous target variable. They are useful for comparing the performance of different models and selecting the best one for a given problem.

Median Absolute Error (MdAE):

$$MdAE(y_i, \hat{y}_i) = median(|y_1 - \hat{y}_1|, \quad |y_2 - \hat{y}_2|, \ldots \ldots ., |y_n - \hat{y}_n|) \tag{4.10}$$

MedAE is an alternative evaluation metric in machine learning that is similar to MAE, but instead of measuring the average magnitude of errors, it measures the median magnitude of errors. MedAE is calculated by taking the median of the absolute differences between the predicted and actual values of a target variable. This makes it a robust measure of error, as it is less sensitive to outliers in the data compared to mean-based metrics like MAE or MSE. MedAE is particularly useful in situations where there are a few extreme values in the data that can heavily influence the mean-based metrics. By using the median, we can reduce the impact of outliers and obtain a more accurate estimate of the typical error. Another advantage of using MedAE is that it is easy to interpret, as it represents the median magnitude of the errors in the original units of the target variable. It is also less influenced by extreme values than mean-based metrics, making it a good choice for evaluating the performance of models in situations where extreme values are expected.

R-squared:

$$R^2 = 1 - \frac{\sum_{i \in n}(y_i - \hat{y}_i)^2}{\sum_{i \in n}(y_i - \bar{y})^2} \tag{4.11}$$

Where $\hat{y}_i$ are predicted values, y_i are observed values, $\overline{y_i}$ is the mean of all y_i, and n is the number of observations.

R-squared, also known as the coefficient of determination, is a widely used evaluation metric in machine learning. It measures the proportion of variance in the target variable that can be explained by the model. R-squared is calculated as the ratio of the explained variance to the total variance. The explained variance is the variance of the predictions of the model, while the total variance is the variance of the actual values of the target variable. R-squared ranges from 0 to 1, with a value of 1 indicating a perfect fit of the model to the data. R-squared is particularly useful in regression problems, where the goal is to predict a continuous target variable. It provides a measure of the goodness of fit of the model and can be used to compare the performance of different models. One advantage of using R-squared is that it is easy to interpret, as it represents the proportion of variance in the target variable that can be explained by the model. It can also be used to assess the predictive power of the model, as a high value of R-squared indicates that the model is able to explain a large proportion of the variance in the target variable. However, one potential drawback of using R-squared is that it can be misleading when used inappropriately. For example, R-squared can increase even when the model is not improving the predictions, as long as the variance of the predictions is increasing. This means that R-squared should be used in conjunction with other metrics, such as MAE or MSE, to provide a more complete picture of the model's performance.

Therefore, machine learning models need to be evaluated using appropriate evaluation metrics to ensure their accuracy and reliability. There are a variety of evaluation metrics that are commonly used in machine learning, each with its advantages and disadvantages. MAE is a simple and interpretable evaluation metric that measures the average magnitude of errors between the predicted and actual values of the target variable. It is particularly useful when the direction of errors is not important. MAPE is a variation of MAE that measures the average percentage error between the predicted and actual values of the target variable. It is useful when we want to measure the relative magnitude of errors. MSE and RMSE are evaluation metrics that measure the average squared error between the predicted and actual values of the target variable. They are particularly useful in regression problems and can be easily optimized using gradient descent algorithms. However, they can be sensitive to outliers in the data. MedAE is a robust evaluation metric that measures the median magnitude of errors between the predicted and actual values of the target variable. It is less sensitive to outliers in the data compared to mean-based metrics and is particularly useful when extreme values are present. R-squared is a widely used evaluation metric in regression problems that measures the proportion of variance in the target variable that can be explained by the model. It provides a measure of the goodness of fit of the model and can be used to compare the performance of different models. However, it should be used in conjunction with other metrics to provide a more complete picture of the model's performance.

4.6 Results

The chapter is implemented using the sci-kit-learn in Python statistical computing packages. Besides, a computer configured with Intel i7 quad-core processor, 16 GB RAM, and Windows 10 operating system is used to execute the study procedure. In this chapter, the ML models are applied to find the demand for our problem. We present some results and discuss different ML algorithms in our problem setting. The two most important parameters to consider when using ML techniques are how computationally concentrated and how quickly a given technique is. Depending on the application type, the most suitable ML algorithm is selected with the parameters (Zantalis et al. (2019); Mohammed et al. (1985)).

Three sorts of listings were checked to identify the clustering of taxi demand in Nagaoka, Niigata: density area, railway station, and supermarket. A list of popular periods represents the number of people that visit this location on specific days and weeks of the week. Clusters may not reveal the exact number of visitors because they do not record all, but what is intriguing is that the number of clusters with popular time data and the number of places without this information in urban areas have the same spatial dimension tendency. From January 1, 2019, to December 31, 2019, its period depicts the spatial distribution of each zone's total number of pickups. The supermarket area has fewer visits during the year than the station area. In 2019, the central railway station received 129,244 pickups during the year.

4.6.1 Performance Prediction

For the LR model, the input is the demand intensity sequence in a single region after sequence stabilization. For DT, RF, and GB, the input that the model accepts each time is a feature vector of a single region in a time interval. For HML, the input consists of an adjacency matrix describing the spatial relationship between regions and the feature matrix composed of feature vectors of all regions in the same time interval. Multiple algorithms work together to complement and augment each other. Table 4.7 reveals the performance of the proposed prediction network compared with the modes. The proposed HML model achieves the best prediction results in the minimum values of MAE (1.49), MAPE (149.2), MSE (8.26), RMSE (2.87), MdAE (0.43), R^2 (0.92).

Table 4.7. Performance of different ML algorithms.

No	Model	MAE	MAPE	MSE	RMSE	MdAE	R^2
1	LR	4.04	404.24	34.02	5.83	2.86	0.67
2	DT	1.97	197.15	14.06	3.75	0.51	0.86
3	RF	1.74	174.17	11.78	3.43	0.44	0.89
4	GB	2.01	200.61	15.16	3.89	0.73	0.85
5	DT+RF+GB	1.5	150.24	8.44	2.9	0.41	**0.92**

The first three clusters were defined based on the demand, and then each ML algorithm was run separately in Table 4.6. Each of the used algorithms is described in more detail in the previous section. For splitting the data frame, train (85%) and test (15%) data sets are considered. Table 4.7 shows the comparison of various ML algorithms. The model is fitted with a 15-minute timestamp in this table. As per Table 4.7, the HML model, with the lowest errors and highest R^2-score (0.92%), was identified as the best model.

4.7 Discussion

This research considers machine learning methods for predicting taxi demand across different months, days, weeks, periods, and locations. Based on the presented results, the performance of the taxi demand forecasting models is as follows:

- The taxi demand in each region might be impacted by the surrounding areas and the clusters located in that area.
- Predicated demand for taxis is affected by weather-related features, such as temperature and the presence of snow. It can be confirmed that the number of operations when it snows, exceeds that of fine weather and rain. There are relatively few trips with a short boarding time and relatively many trips with a long boarding time when it snows. When the distribution of riding distance was confirmed, there is no significant difference due to the weather. It is presumed that the snowfall did not induce short-distance demand but that the vehicle speed decreased. As a result, basic HML techniques may be applied to avoid prejudice and achieve justice in the taxi demand problem.
- Modifying the values of hyperparameters can result in performance improvement.
- The HML model can provide more accurate results than the others (LR, DT, RF, and GB) for dynamic purposes.

4.8 Prediction and Closeness of Predict Data

The prediction and closeness of predicted data are key measures of the accuracy of an HML model. Predictions refer to the output generated by an ML model, which is based on input data and the model's underlying algorithms. The closeness of predicted data refers to the degree to which the predicted output values match the actual values of the target variables.

HML models combined multiple ML algorithms to generate more accurate predictions. These models used a variety of techniques, such as LR, DT, RF, and GB, to analyze different aspects of the input data and generate more accurate predictions. The prediction and closeness of predicted data are critical measures of the effectiveness of these models, as they can help to determine the degree to which the models can make accurate predictions based on the input data. There are various techniques for measuring the accuracy of predictions in HML models. One common approach is to use a metric such as mean squared error MAE, MAPE, MSE, RMSE, MdAE, and R^2, which can help to quantify the difference between the predicted values and the actual values of the target variables. Another approach is to use cross-validation, which involves splitting the input data into multiple parts, using different parts for training and testing the model, and evaluating the performance of the model using a variety of metrics.

Ultimately, the prediction and closeness of predicted data are key indicators of the performance of an HML model. By carefully analyzing these measures and identifying areas for improvement, it is possible to further optimize the performance of these models and improve their ability to make accurate predictions.

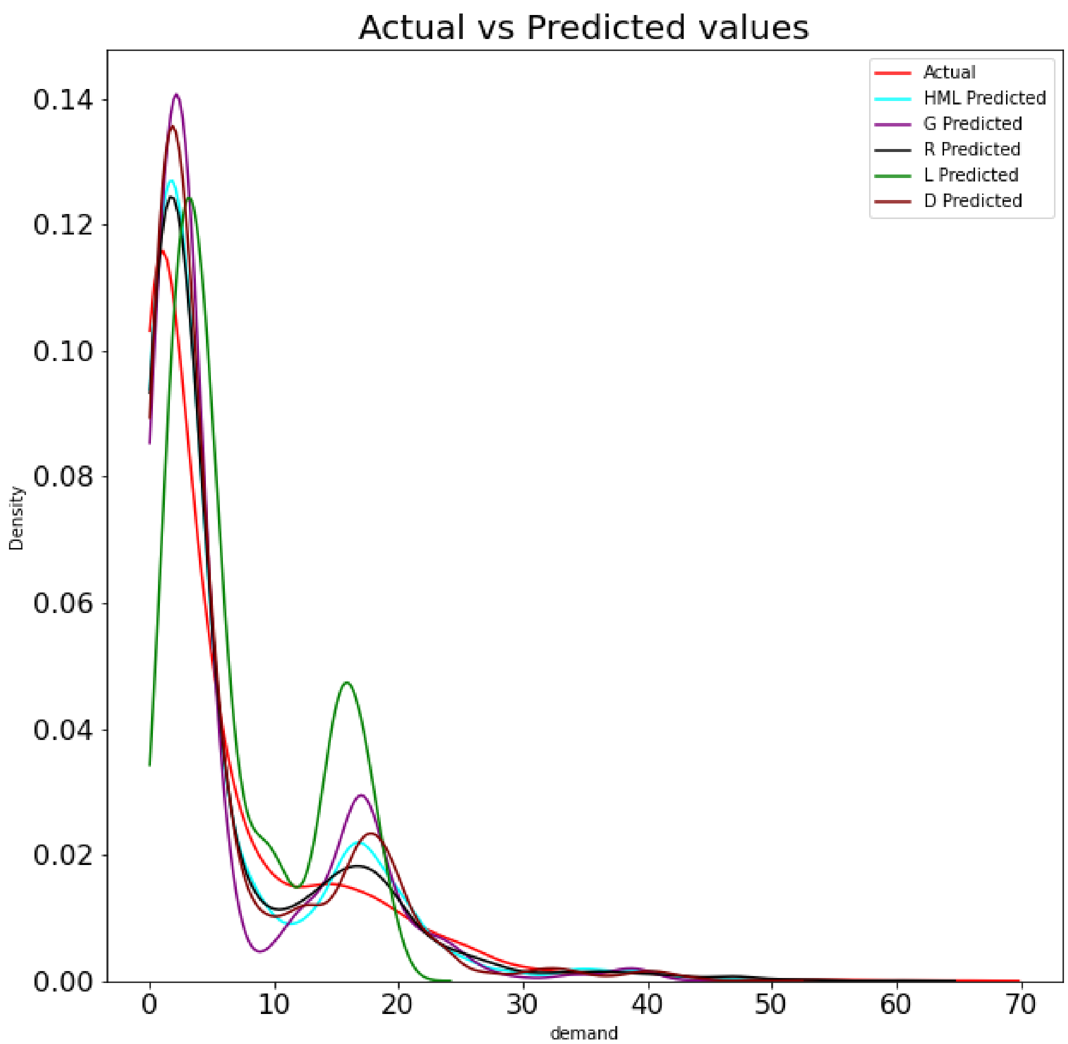

Figure 4.6. The prediction results of 5 models.

The actual vs predicted values of a hybrid machine learning model refers to the comparison between the observed real-world values (actual values) and the values that are estimated or predicted by the model. The accuracy of a hybrid ML model is evaluated by comparing the actual vs predicted values. Plot the actual and predicted values in a line plot to visually compare the performance of the hybrid model. We added a line of best fit to the plot to see how closely the predicted values match the actual values. The result shows that taxi demand fluctuates due to its dynamic nature, which is influenced by various factors such as the time of day, day of the week, and season. Also, unexpected events like major public gatherings, festivals, or inclement weather conditions can significantly impact taxi demand. The fitted curves between the actual values and the different results obtained from various models are given in Figure 4.6, from which it can be concluded that the proposed HML model presents the best performance.

4.9 Summary of the Chapter

This chapter discussed the prediction demand in 15-minute timestamp periods. So, we prepared the data and used four popular ML models and the HML model to build a demand prediction model. We accepted five types of input data: demands of each area, months, days of weeks, 15-minute timestamps, and weather for ML and evaluated their impact on taxi demand. The relationship between the predictors and demand with different accuracy tests has been discussed. Therefore, considering the input features and weather conditions for enhancing the model's performance, a taxi operator can estimate taxi demand. In this chapter, the implementation of these models has been evaluated using MAE, MAPE, MSE, RMSE, MdAE, and R^2. The results demonstrated the capability of ML models to predict the demand level. The HML algorithm consistently achieved the best classification performance.

Therefore, predicting taxi demand in urban areas is a crucial task for improving the efficiency and sustainability of the transportation system. HML algorithms, combining different ML methods, are indicated to be effective in accurately predicting taxi demand in urban areas. This can provide valuable insights for taxi companies and city planners to optimize their resources and improve urban transportation services. Additionally, the study revealed several critical predictors for estimating demand levels. The insights obtained from the study have numerous practical implications for aiding decision-makers at strategic, tactical, and operational levels.

Chapter 5. MATHEMATICAL MODEL

In this chapter, we will delve deeper into the specific mathematical models used for shared taxi routing and taxi allocation, exploring the key variables, constraints, and objectives that are involved. The mathematical model of shared taxi routing involves the use of algorithms and optimization techniques to determine the most efficient routes for taxis to take and the optimal allocation of passengers to those taxis. By using a mathematical model, transportation planners, and operators can make informed decisions about how to allocate resources and maximize the benefits of shared taxi routing.

5.1 Introduction

A mathematical model of shared taxi routing and taxi allocation in transportation is a system of equations that describes how an allocation of taxis can be optimized to serve the transportation needs of a population efficiently and cost-effectively. This model typically involves several variables, including the number of taxis in the fleet, the number of passengers requiring transportation, the location and distance of each passenger from their destination, the expected travel time for each trip, and the cost to be charged for each trip.

The shared taxi routing problem is calculated based on the distance traveled, the number of passengers in the taxi, and other factors such as time of day or traffic congestion. The cost is typically split among the passengers, with each passenger paying a proportionate share of the total trip cost based on the distance they traveled. The goal of the model is to maximize the utilization of the taxi fleet while minimizing the cost to passengers. This is achieved by grouping passengers based on their proximity and destination and then assigning them to a single taxi for the trip. This reduces the number of taxis required to serve the population, which in turn reduces the cost to passengers. The model typically uses a set of algorithms to calculate the optimal route for each taxi based on the location and destination of its passengers. This may involve complex calculations to determine the shortest and most efficient route, taking into account traffic conditions, road closures, and other factors that may affect travel time.

Therefore, the mathematical model of shared taxi routing problem and taxi allocation in transportation is a sophisticated system that can help improve the efficiency and cost-effectiveness of taxi services, while also reducing traffic congestion and improving overall transportation outcomes for the population.

5.2 Shared Taxi Routing Model

Recently, shared taxis are the most popular for public transportation. The sharing charge influences the taxi-sharing enthusiasm of passengers. Due to the difference in the service order, travel distance, number of shared passengers, and waiting time, the cost-sharing ratio should be changeable. Again, establishing a generalized pricing method is necessary based on those factors. It can lower fare costs for each customer and reduce travel costs. Meanwhile, it increases the taxi company’s revenue as it accepts drivers to pick up multiple passengers.

Dial-A-Ride Problem (DARP) requests trips from indicating the pickup locations to drop-off locations, while individual requirements on the ride (such as the initial pickup time and final drop-off time) (Cordeau and Laporte (2007)). Several variants of DARP exist, including single-vehicle and multi-vehicle DARP and static and dynamic DARP. The dynamic version accommodates changes that include users leaving and entering the system. It is essential to handle new customer requests during the whole operation of the taxi rider fleets spread around and while traveling. Ke et al. (2020) examined the empirical probability density functions of trip time for ride-pooling service and non-pooling service for hire-vehicle ride-sourcing services in New York City and developed a mathematical model of a ride-pooling market. The authors found the optimum solution for passengers choosing ride-pooling and non-pooling services. The ride-pooling service has a slightly longer average trip time.

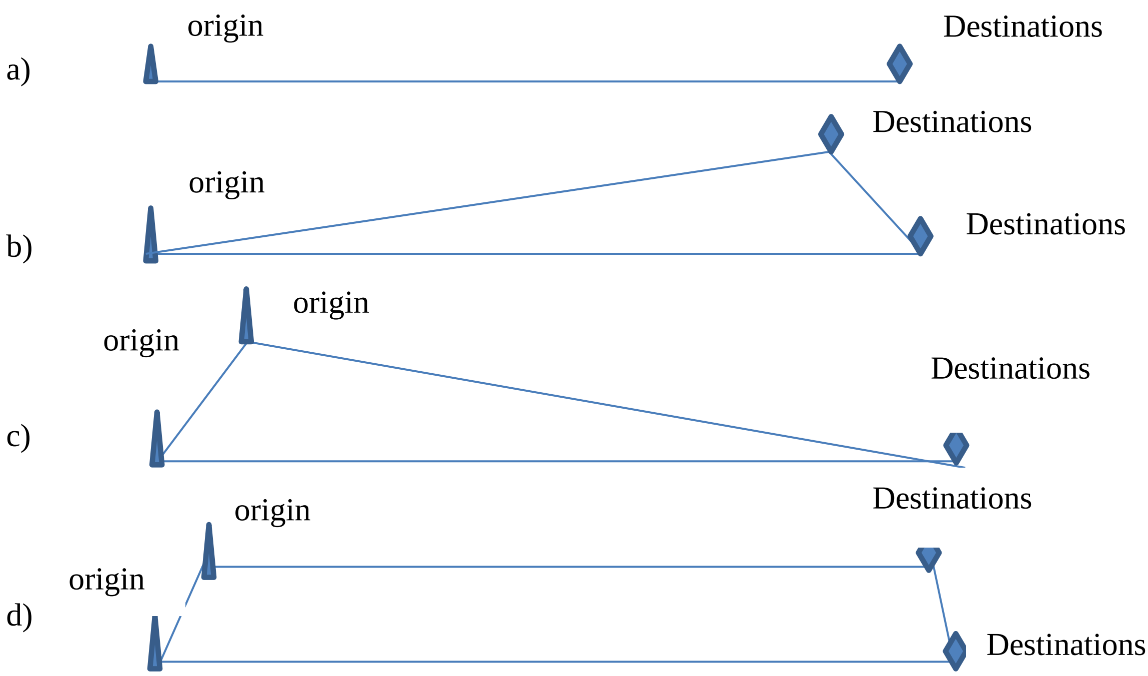

Figure 5.1. Four prominent taxi-sharing cases.

The taxi trip cost is calculated on the pickup and drop-off location of each passenger. The shared taxi problem is obtained to determine the optimal passenger position to taxi, the optimal route of each taxi, the optimal cost of each passenger, etc. The taxi can pick up different origins and drop off customers who go to various destinations in this model.

The four characteristics of the routing problem (for details, see Figure 5.1) of shared taxis are given below: (a) passengers have the same origin and the same destination, (b) passengers have the same origin and different destinations, (c) passengers have the different origins and the same destination, (d) passengers have the different origins and different destinations. Two main classes of taxi sharing are studied: static taxi sharing and dynamic taxi sharing. Static taxi sharing requires all taxi trips to be known beforehand. Therefore, a globally optimal sharing plan could be derived to maximize the collective benefits of sharing, e.g., cumulative trip length reduction. Dynamic taxi sharing matches real-time trip requests with running taxis. The trip requests are unknown beforehand, and the taxis can reroute and respond to new trip requests on the fly.

The taxi allocation method based on the number of passengers has some benefits. It can calculate the cost per passenger. The payment limitation can be solved by turning where the early fare is equal to the next fare.

5.2.1 Formulations

For the mathematical model, some assumptions are needed:

1. Each taxi driver is different from the others.
2. The taxi capacity of each taxi is four passengers.
3. Passengers are not allowed to transfer.
4. The driving speed of the taxi remains unchanged in terms of taxi sharing.
5. The pickup and drop-off must always happen within the specified time windows.
6. The discount parameter value is from 10% to 30%.
7. The first ride fare is 630 yen up to 1200 m, and additional fares are 90 yen every 249 m. (Source: https://www.taxisite.com/far/info/15.aspx)

Sets, indices, and parameters:

The shared taxi of the pickup and drop-off problem contains n requests and k taxi. Passenger i request is represented by nodes i and $i + n$. Given a directed graph $G = (V, A)$ where, the V, the set of all points, $i, j \in V$, and i, j, connect; and A the set of arcs; K, the set of all taxi, $k \in K$, let $\tau_k = 2n + k$, $k \in K$ and $\tau'_k = (2n - 1) + k$, $k \in K$ be the nodes representing the start and end terminals, respectively, of taxi k.

Symbols		Descriptions
q_k	:	Capacity at k^{th} taxi
C_0	:	Initial fare [yen]
N_i^k	:	The number of passengers at k^{th} taxi at point i
B_i^k	:	Departure time at k^{th} taxi at point i, [hrs.]
s_i	:	Service time at i point, [min]
e_i	:	The earliest departure time for each point $i \in V$, [min]
l_i	:	The latest arrival time for each point $i \in V$, [min]
d_{ij}	:	The distance of i to j, [km]
t_{ij}	:	Travel time from i to j, [min]
W_i^t	:	Passenger waiting time at i point, [min]
R_{ij}	:	The fare rate of going from i to j, [yen/km]
C_{ij}^k	:	The fare of going from i to j at k^{th} taxi, [yen] $$C_{ij}^k = \begin{cases} C_0\,, & if \sum_j \sum_i x_{ij}^k d_{ij} \le d_0 \quad \forall k \in K \\ C_0 + R_{ij}, & \text{Otherwise} \end{cases}$$
d_0	:	Initial distance in fixed fare, [km]
α	:	The discount parameters of the passenger waiting time
β	:	The fare of the passenger distance
P_i	:	Number of passengers at i point

The decision variable

$$x_{ij}^k = \begin{cases} 1, & \text{If and only if the arc (i, j) used by taxi k} \\ 0, & \text{Otherwise} \end{cases}$$

The shared taxi fare problem is then formulated as follows.

$$Min \sum_i \sum_j \sum_k \left(\frac{C_{ij}^k}{N_i^k}\right) x_{ij}^k \tag{5.1}$$

Here, the objective function minimizes the total trip fare of passengers and the number of taxis utilized for potential demand (maximizing passenger satisfaction).

Subject to constraints:

1. The conservation of taxi flows specifies that each taxi must leave the depot and finally arrive at the same depot. Again, when a taxi visits a customer node, it must leave that node and arrive at the next node, respectively.

$$\sum_{j\in V} x_{ji}^k - \sum_{j\in V} x_{ij}^k = 0 \qquad \forall k \in K, \forall i \in V$$

$$\sum_{j\in V} x_{ij}^k - \sum_{j\in V} x_{j,n+i}^k = 0 \qquad \forall k \in K, \forall i \in V$$

$$\sum_{i\in V}\sum_{k\in K} x_{ij}^k = 1, \qquad \forall j \in V$$

$$\sum_{i\in V}\sum_{k\in K} x_{ji}^k = 1, \qquad \forall j \in V$$

$$\sum_{k\in K} x_{jj}^k = 0, \qquad \forall j \in V$$

$$\sum_{k\in K} x_{j0}^k = 0, \qquad \forall j \in V$$

$$\sum_{k\in K} x_{\tau_k',j}^k = 0, \qquad \forall j \in V$$

$$\sum_{j\in V} x_{0j}^k = 1, \qquad \forall k \in K$$

$$\sum_{i\in V} x_{i,\tau_k'}^k = 1, \qquad \forall k \in K$$

$$x_{0,\tau_k'}^k + x_{\tau_k',0}^k = 0, \qquad \forall k \in K$$

2. Each taxi's service time is reached by each taxi, while customers can only be picked up after the specified earliest pick-up time and drop-off before the specified latest drop-off time.

$$\left(x_{ij}^k = 1\right) \Rightarrow B_j^k \geq B_i^k + (s_i + t_{ij})x_{ij}^k \qquad \forall k \in K, \forall i,j \in V$$

$$e_i \leq B_i^k \leq l_i \qquad \forall k \in K, \forall i \in V$$

3. Capacity constraint defines the maximum number of passengers per trip request. This constraint avoids forming shared trips that exceed taxi capacity.

$$\left(x_{ij}^k = 1\right) \Rightarrow N_j^k \geq N_i^k + P_i x_{ij}^k \qquad \forall k \in K, \forall i,j \in V$$

$$0 \leq N_i^k \leq q_k \qquad \forall k \in K, \forall i \in V$$

$$N_0^k = 0, \qquad \forall k \in K$$

$$N_{\tau'_k}^k = 0, \qquad \forall k \in K$$

4. The waiting time may be longer for those passengers who arrive first than for those latecomers. If her or his waiting time is too long, the passenger should be given a discount.

$$W_i^t = (B_i^k - e_i), \qquad \forall k \in K, \forall i \in V$$

$$W_i^t \geq 0, \qquad \forall i \in V$$

5. Waiting time defines the maximum time allowed when a traveler submits a request and when a taxi is assigned to serve the trip.

   Taxi fare rate constraint R_{ij} reflects the total shared taxi fare based on the shared trip's waiting time and travel distance.

$$R_{ij} = \alpha W_i^t + \beta \left(\sum_k x_{ij}^k d_{ij} - d_0 \right) \qquad \forall i, j \in V$$

Finally, constraints specify the domain of the variables.

5.2.2 Summary of Shared Taxi Model

The shared taxi routing model is a mathematical model that calculates the fare for shared taxi trips based on the distance traveled and the number of passengers in the taxi. The trip cost is typically split among the passengers, with each paying a proportionate share based on the distance they traveled. The model aims to reduce the cost to passengers by grouping them based on their proximity and destination and assigning them to a single taxi for the trip, reducing the number of taxis required to serve the population.

Therefore, the model can improve the cost-effectiveness of taxi services while providing a more affordable transportation option for passengers. So, the mathematical model for the shared taxi routing problem involves calculating the total cost for the ride based on the total distance traveled by all passengers and then dividing that cost among the passengers based on the distance each passenger travels. This ensures that each passenger pays a fare proportional to the distance they travel, and the total fare is fairly divided among all passengers.

5.3 Taxi Allocation Problem

The taxi allocation model is a mathematical model that aims to optimize the use of a fleet of taxis to serve the transportation needs of a population efficiently and cost-effectively. With the growth of urbanization and the increase in the number of people using taxi services, there is a need for a more efficient and effective way to utilize taxi fleets to ensure the provision of high-quality services at affordable prices. The taxi allocation model addresses this need by grouping passengers based on their proximity and destination and assigning them to a single taxi for the trip. This reduces the number of taxis required to serve the population, which in turn reduces the cost to passengers. The model also calculates the optimal route for each taxi based on the location and destination of its passengers, taking into account traffic conditions and other factors. Overall, the model can help improve the efficiency and cost-effectiveness of taxi services while reducing traffic congestion and improving transportation outcomes for the population.

Climate change and global warming are increasing in today's world, which leads to finding strategic solutions to reduce carbon dioxide (CO2) emissions. Urban areas significantly impact CO2 emissions in traffic congestion due to high transport demand and excessive street parking. Vehicles typically run on gasoline or diesel fuel, both of which emit CO2 and other harmful pollutants into the air when burned. Increased traffic congestion caused by private vehicles can lead to decreased travel speed and increased travel time for passengers, which emits higher CO2 in urban areas (Yang *et al.* (2021); Ghahramani *et al.* (2021); Ghahramani and Pilla (2021b); Ghahramani *et al.* (2020); Franckx and Mayeres (2016); Aggoune-mtalaa (2023)). Such air pollution is becoming a serious problem with the rapid increase in vehicle ownership in developing countries (Singh *et al.* (2017); Wang *et al.* (2015); Zhang and Nian (2013)). In order to address the issue of air pollution caused by traffic congestion, it is essential to promote the usage of public transportation and taxis as substitutes for personal vehicles, thereby reducing the number of vehicles on the road network.

On the other hand, inefficient or improper taxi service management can cause numerous issues such as increased idle time, high costs of operation, and increased CO_2 emissions (Mingolla and Lu (2021); Postorino *et al.* (2019); Ashkrof *et al.* (2022)). In the case of idle time, there are two types of idle time for the driver, moving idle time and stationary idle time. Moving idle time is the time when taxi drivers travel while searching for passengers to pick up or while returning to their original location. Stationary idle time is the period when the driver parks the car for various reasons, such as a driver's break time or waiting for a customer's request. The idle time occurs when the vehicle's engine is running, but the taxi is standing, while delay time occurs when the taxi is stuck in traffic or waiting for passengers. When a taxi is idling, it is still burning fuel without moving, which leads to unnecessary CO_2 emissions. As a result, the delay time can be identified as the key factor for CO_2 emission in urban areas (Lazaroiu and Roscia (2012)). Most scholars argued that reducing idle and delay time is essential to mitigate the emission of CO_2 in taxis (Song *et al.* (2019); Nicolaides *et al.* (2019); Pourakbari-Kasmaei *et al.* (2020); Hou *et al.* (2020); Wang *et al.* (2020)). The idle time cost is a measure of the time that taxi drivers spend idling without generating revenue and has a direct impact on profitability. Therefore, for effective taxi operation, taxi allocation is necessary to prioritize. Taxi allocation refers to the process of assigning taxis to pick up passengers Yang *et al.* (2000). Once the taxi allocation is optimized, idle time can be reduced, which results in reduced taxi operational costs. Therefore, reducing idle time by optimizing taxi allocation can lead to several benefits, including time, lower costs, and decreased CO_2 emissions.

However, the problem of taxi fleet allocation has been studied extensively in the literature, especially at the operational level of taxi companies. Most of the existing studies have focused on urban areas (Zhang *et al.* (2020); Wang *et al.* (2015); Zhang *et al.* (2023)), where the demand and supply of taxis are relatively high and dynamic. In contrast, suburban areas have different characteristics, such as lower demand, longer travel distances, and higher idle time. These factors affect not only taxi companies' service quality and profitability but also the environmental impact of taxi operations, as idle time is correlated with CO_2 emission. Therefore, this study aims to fill the gap in the literature by proposing an optimization model for taxi allocation in suburban areas, using authentic data from actual taxi firms. The model considers both the economic and environmental objectives of taxi companies and the constraints of demand, supply, and vehicle capacity. The model is applied to a case study of a suburban area, and the results show that the optimal fleet size and reallocation to meet future passenger demand can reduce both the idle time cost and CO_2 emission of taxi companies while improving the community's living standards, thereby creating a sustainable future. Furthermore, this study proposes a new dynamic greedy heuristic algorithm to obtain better solutions to the problem and evaluate the performance of this algorithm in comparison with greedy and simulated annealing heuristic algorithms.

We formulate a mathematical model of the taxi allocation problem to optimize the taxi drivers' idle time costs as a mixed-integer program, which will support decision-makers, urban transport planners, and governmental and taxi companies to make their decisions more strategic. The proposed model focuses on taxis in suburban areas. Most of the existing studies, (e.g., Zhang *et al.* (2020); Wang *et al.* (2015); Zhang *et al.* (2023)), consider only the taxi model for urban areas other than suburban areas. Since the behavior of the taxi model for suburban areas is different from that of urban areas, our model can contribute to the reduction of CO_2 emissions and the efficient operation of a taxi service in suburban areas. In order to accomplish this objective is as follows:

5.3.1 Establishing the Problem

The cost of taxi services is influenced by the idle time of the drivers, which is the time they spend without passengers. Idle time is not only unproductive but also increases fuel consumption and emissions. One way to reduce idle time is to match the supply and demand of taxis more efficiently by predicting where and when passengers will need a ride. For example, if a driver received a call from a nearby location, they would move towards it, but if the call were from a faraway or congested area, they would ignore it or decline it. This subsection presents the challenges of modeling the cost function that incorporates idle time and the benefits of using data-driven methods to optimize it. This way, we could capture the complexity and variability of the taxi service in our simulation.

By reducing the idle time, each taxi will have greater use, while the taxi driver spends less time finding passengers. Further, a smaller number of taxis can cover the same amount of demand. Accordingly, there are two objectives of this model:

- To minimize the total idle time of the taxi driver when using the optimum number of allocated taxis.
- To minimize the total operational cost when minimizing the idle time of the taxi driver and the number of taxis utilized of potential demand.

The above goals are directly tied to taxi system efficiency and driver profitability. In addition, idle time costs can be considered important factors in reducing CO2 emissions.

5.3.2 Problem Assumptions

The following assumptions were used in this research to better understand the taxi problem:

a) Assumption 1 - Passenger demand for taxis is assumed to be random:

This assumption highlights the importance of reducing idle time and optimizing routes to avoid unnecessary driving. Through the developed program, it is able to achieve significant reductions in both idle time and CO_2 emissions while improving efficiency and reducing costs for all stakeholders involved.

b) Assumption 2 - An empty car repeatedly moved over the planning horizon:

This assumption highlights the importance of reducing idle time and optimizing routes to avoid unnecessary driving. Through the developed program, is able to achieve significant reductions in both idle time and CO_2 emissions, while improving efficiency and reducing costs for all stakeholders involved.

c) Assumption 3 -The pick-up and drop-off must always happen within the specified time windows i.e., the drivers waiting time for the passenger's and the passenger's waiting time for the taxi.

Assumption (3) is the requirement for a successful taxi service to ensure that passengers' pickup and drop-off are done on time. It means that the drivers and the passengers must respect the agreed time windows for each trip, creating a more sustainable and efficient transportation system.

d) The first ride fare is 630 yen up to 1200 m, and additional fares are 90 yen every 249 m. (Source: https://www.taxisite.com/far/info/15.aspx)

5.3.3 Formulation of the Idle Time Cost

This section shows a mathematical formulation of the taxi allocation problem. We consider an undirected graph $G = (N, A)$ that consists of N locations and the A arcs ($A = N \times N$). Let S_i^k and E_j^k represent the start and end working times of the taxi driver k at the locations i and j. For arc $(i, j) \in A$, we assign a positive distance and travel time T_{ij}. Location $i \in A$ has a service time ser_i . The service time represents the time required for pick-up and drop-off, and the time indicates when the visit to the location must start. A taxi is allowed to arrive at a location before the start of the time, but it has to wait until the start of the time before the trip can be performed. The maximum capacity of a taxi is denoted by Q.

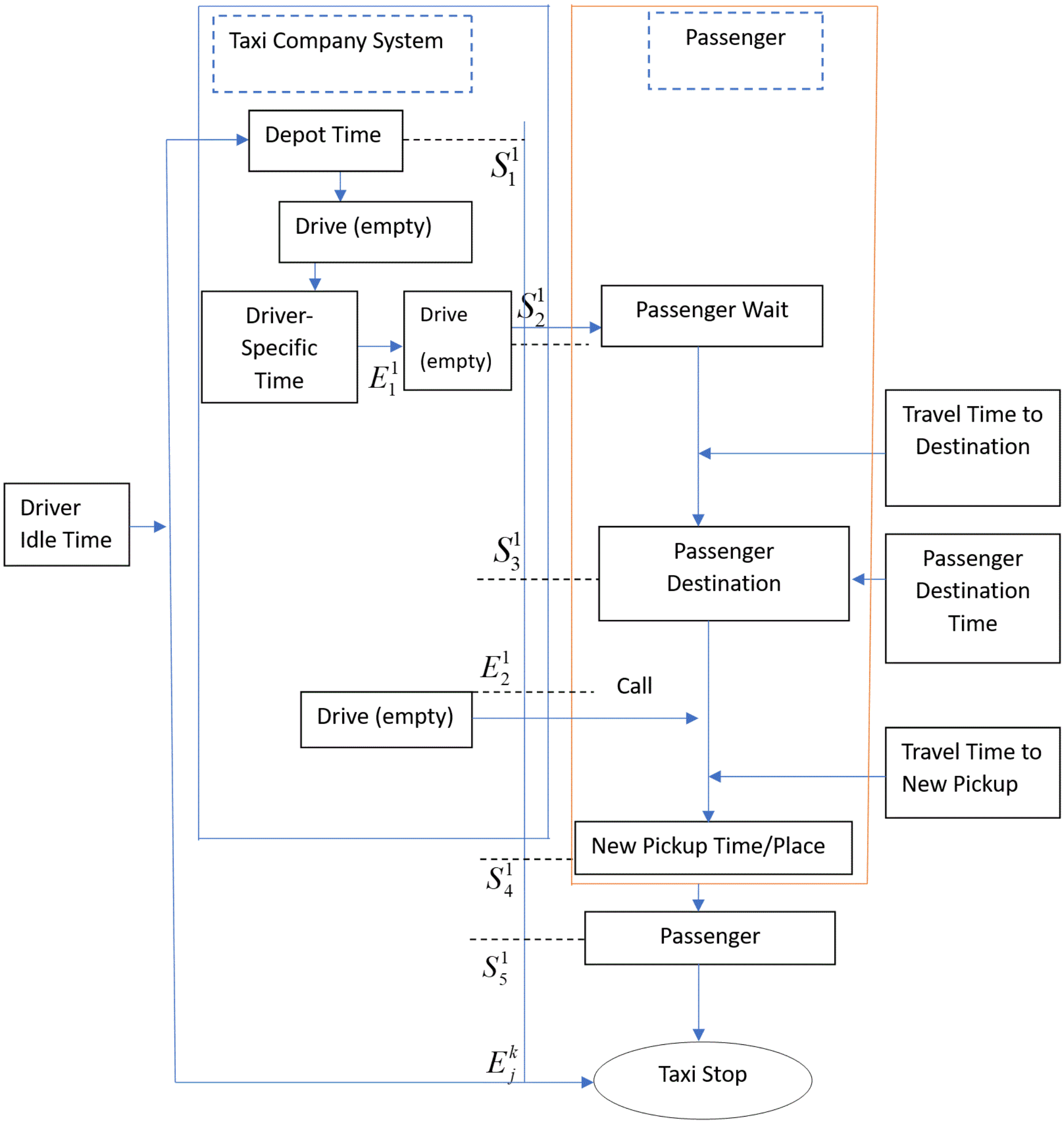

Figure 5.2. A sample of taxi allocation for explaining the idle time

Figure 5.2 shows a sample of taxi allocation. This taxi (numbered as 1 ∈K) starts working time at location 1 at the time S_1^1. Location 1 is where this taxi remains stationed for passengers who are waiting. Then this taxi is assigned a passenger at the time E_1^1 and moves to location 2 where the passenger is waiting. Then, this taxi pick-up the passenger at location 2 at the time S_2^1; after that, the onboard passenger drop-off at location 3 at the time S_3^1. After that, the driver waits at specific locations for the passenger, then the taxi is assigned a passenger at the time E_2^1 and moves to new passenger pickup location 4 at the time S_4^1 and the onboard passenger drop-off location 5 at the time S_5^1. We define idle time as the amount of time a driver spends waiting for a new passenger request. Then, the total idle time can be calculated as $(E_1^1 - S_1^1) + (E_1^1 - S_2^1) + (E_2^1 - S_3^1) + \cdots .. + (E_j^k - S_i^k)$ for this sample.

5.3.3.1 Notations

In this paper, the notations and parameters to be applied in the formulations are as follows:

Symbols		Descriptions
A	:	The set of arcs
N	:	The total number of taxi drivers
S_i^k	:	Start time for driver k his working day at a specific location $i \ \forall i \in A$ (Sec.)
E_j^k	:	Destination time for driver k his working day at a specific location $j \ \forall j \in A$ (Sec.)
c_0	:	Idle time unit cost (yen)
q_i	:	Demand at the location $i \ \forall i \in A$
ser_i	:	Service time at the location $i \ \forall i \in A$, each taxi maximum stay at the location (Sec.).
T_{ij}	:	Travel time from the location $i \, to \, j \ \forall i, j \in A$
Q	:	Maximum capacity of each taxi
M	:	A large number

The decision variables are as follows:

$$Z_{ij}^k = \begin{cases} 1 \text{ if traveling along (i,j) with taxi k} \\ 0 \quad \text{otherwise} \end{cases}$$

5.3.3.2 Formulations

The taxi driver idle time problem is formulated by several equations. Equation (5.2) develops from this study, and the flowing constraint functions are found in the previous literature surveys Ropke and Pisinger (2006).

$$Min\sum_{k=1}^{N}\sum_{i,j=1}^{A}[c_o(E_j^k - S_i^k)]Z_{ij}^k \tag{5.2}$$

$$\sum_{j=1}^{A} Z_{ij}^k - \sum_{j=1}^{A} Z_{ji}^k = 0 \quad \forall \boldsymbol{i} \in \boldsymbol{A}, \forall \boldsymbol{k} \in \boldsymbol{N} \tag{5.3}$$

$$\sum_{j}^{A} Z_{0j}^k = 1 \ \forall \boldsymbol{k} \in \boldsymbol{N} \tag{5.4}$$

$$\sum_{i=1}^{A} Z_{i0}^k = 1 \ \forall \boldsymbol{k} \in \boldsymbol{N} \tag{5.5}$$

$$\sum_{i=1, i\neq j}^{A} Z_{ij}^k = 1 \ \forall \boldsymbol{j} \in \boldsymbol{A}, \forall \boldsymbol{k} \in \boldsymbol{N} \tag{5.6}$$

$$S_i^k + ser_i + T_{ij} - M(1 - Z_{ij}^k) \le E_j^k \ \forall \boldsymbol{i}, \boldsymbol{j} \in \boldsymbol{A}, \ \forall \boldsymbol{k} \in \boldsymbol{N} \tag{5.7}$$

$$\sum_{i=1}^{A}\sum_{j=1, i\neq j}^{A} q_i Z_{ij}^k \le Q \ \forall \boldsymbol{k} \in \boldsymbol{N} \tag{5.8}$$

$$Z_{ij}^k \in \{0,1\}, \forall k \in N, and \, \forall i, j \in A \tag{5.9}$$

The objective function (5.2) minimizes the cost of taxi idle time, which is the objective function. C_0 is the idle time cost, which is dependent on the taxi company, but in this study, an idle time cost of JPY 300 has been used. Constraint (5.3) is that the number of taxis coming to a passenger's location and the number of taxis exiting are the same. (5.4) constraint that the taxi starts from the depot. Constraint (5.5) ensures that finally, the taxi returns to the depot. Constraint (5.6) means that each passenger is riding only once in one taxi. Constraint (5.7) means ensures that passengers do not consume more time at the taxi stand. Constraint (5.8) does not exceed the maximum capacity (Q) of the taxi. That means no more than one passenger per taxi. Constraint (5.9) is the binary condition.

5.3.4 Summary of the Taxi Allocation Problem

A formulation for the taxi allocation based on the taxi delay time has been studied, concluding in optimum cost for demand levels. The model involves grouping passengers based on their proximity and destination and assigning them to a single taxi for the trip to reduce the number of taxis required. The model also calculates the optimal route for each taxi based on the location and destination of its passengers, taking into account traffic conditions and other factors. A mathematical model for taxi allocation involves estimating the number of taxis required to meet the demand for taxi rides in a given area at different times of the day while minimizing the cost of maintaining and operating the taxi location.

Therefore, the taxi allocation model can improve the efficiency and cost-effectiveness of taxi services while reducing traffic congestion and improving transportation outcomes. Using this model, we can estimate the optimal number of taxis required to meet the demand for taxi rides in each zone at different times of the day, while minimizing the cost of operating the taxi fleet. This can help taxi companies to better manage their fleet, reduce operational costs, and improve customer satisfaction by providing reliable and timely taxi services.

5.4 Summary of the Chapter

The shared taxi routing problem model aims to reduce the cost of transportation for passengers by allowing multiple riders to share a single taxi. This not only benefits the passengers but also increases the revenue for taxi drivers by increasing the number of riders per trip. The shared taxi routing problem model can also reduce traffic congestion by reducing the number of taxis on the road.

The taxi allocation model aims to increase the efficiency of taxi fleets by optimizing the distribution of taxis to meet the demand for transportation services. This can be achieved through the use of data analytics and predictive algorithms that can analyze real-time data on passenger demand, traffic patterns, and weather conditions to make informed decisions on the allocation of taxis. By optimizing the allocation of the taxi, the model can reduce idle time costs for operators, increase the number of trips made by taxis, and reduce the number of empty taxis on the road and CO_2 emissions.

In conclusion, both the shared taxi routing model and the taxi allocation problem have the potential to improve the efficiency of the taxi industry and enhance the overall transportation system. The models can help to reduce costs for passengers, increase revenue for taxi drivers, and reduce traffic congestion. The success of these models depends on the implementation of effective policies and the use of advanced technologies to optimize the allocation of taxis.

Chapter 6. CASY STUDY

In this chapter, the two developed models are applied to the taxi sector of the city of Sanjo, providing significant results in terms of minimum and optimum taxi routes, and the city of Nagaoka, providing significant results in terms of optimum CO_2 emissions based on idle time.

6.1 A Case Study of Sanjo City, Japan

This section analyzes the actual taxi operation by using GPS data of all vehicles belonging to Sanjo Taxi Company, Ltd., the largest taxi company with a business area in the entire Niigata. The analysis period was in June 2015. Sanjo City, Niigata Prefecture, is in the center of Niigata Prefecture and is classified in the Chuetsu region. Three main public transportation lines are operating in Sanjo City: the Joetsu Shinkansen, Shin-Etsu Line, and Yahiko Line. Regarding taxis, the four taxi companies that operate the shared taxi "Hime Sayuri," which is the subject of this study, also carry out regular taxi business. Regarding buses, highway buses to Niigata City, city circulation bus "Gurutto-san," high school student commuting liner buses, and community buses in some areas are also running, and public transportation projects have been widely introduced throughout the city.

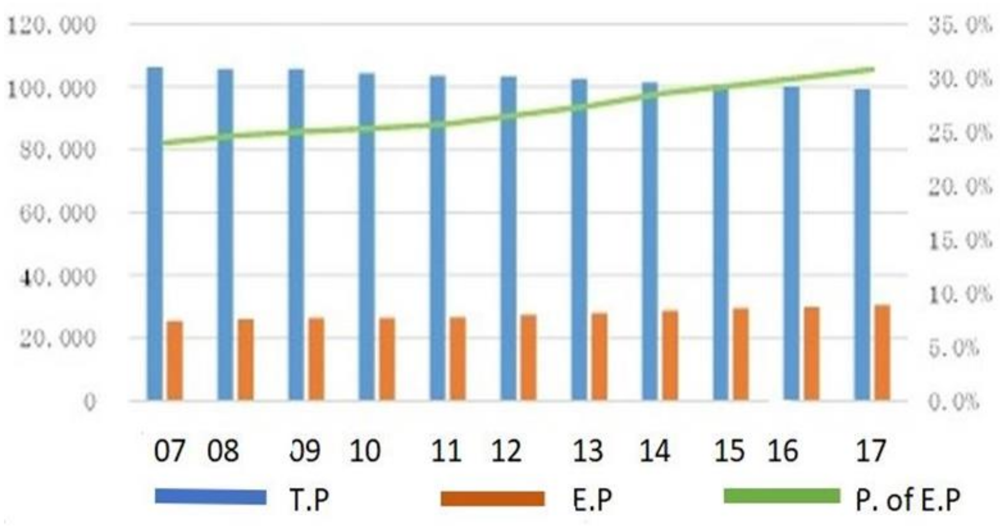

Figure 6.1. Changes in the total population(T.P), the number of elderly people(E.P), and the percentage of elderly people(P.E.P) in Sanjo City[5].

In low density, rural areas were focused based on two factors essential for shared taxi operations; Limited availability of general transport services and High spatial dispersion of travel demand (AINI *et al.* (2019)). This situation caused problems for older and disabled people with traveling difficulties in rural areas. Considering the requirements, a demand share taxi system is introduced, incorporating a traditional taxi system.

5 https://www.city.sanjo.niigata.jp/

6.1.1 Data Analysis

The taxi trips were categorized based on the day of the week (weekday) and time of day (“07:00-09:00”, “09:00–11:00”, “11:00-13:00”, “13:00-15:00”, “15:00-18:00). The peak time is from 8:00 to 10:00. Originally, the usage time is from 8:00 to 18:00, but it seems to operate around 7:00 according to the user's request. At 7 o'clock, the synergistic rate halved before and after the price revision. It is probable that while the number of users was initially small, the number of users decreased further due to the price increase, and carpooling decreased significantly. The number of trips is always decreasing as well. The synergistic rate increases most of the time slightly.

The dataset was cleaned for incorrect records. The taxi trips with 0 min travel time or kilometers traveled fare less than 630 yen (the initial fee for taxis in Niigata, Japan), trips originating and arriving at the same place, and trips with an average speed of less than 20 km/h. Figure 6.2 represents a partial screen of the operational status for pickup (red) and drop-off (yellow) locations. The taxis were deployed in June 2015, using the data recording devices installed in the car. The total number of recorded passenger rides is 5,041. The driver's data acquisition method is input and recorded by the date, time, vehicle ID, direction, operating status, latitude, & longitude readings acquired from GPS.

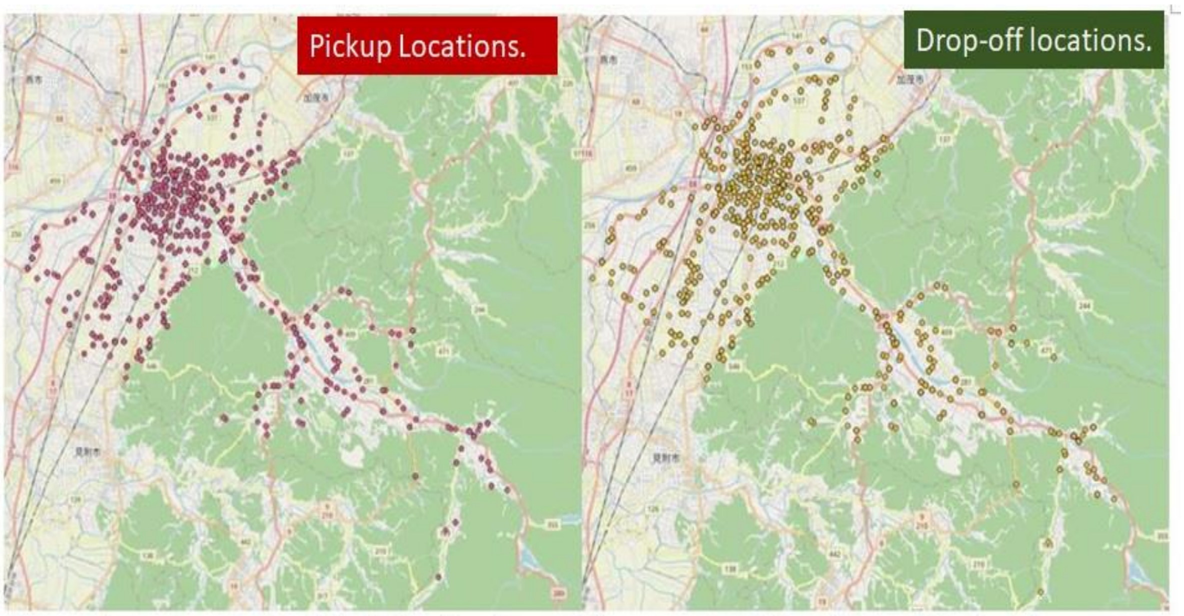

Figure 6.2. Geographical distributions of the customers' locations in Sanjo City

In this research, after pre-processing, we have used taxi operation data such as date, the number of users, pick up and drop off (longitude and latitude), and operation status (departure and arrival time) in

Table 6.1.

Table 6.1. Taxi GPS trajectory data

No	Operating month	No of Users	Pickup latitude (degree)	Pickup longitude (degree)	Departure time	Drop-off latitude (degree)	Drop-off longitude (degree)	Arrival time
1	6/1/2015	2	37.56554	139.0528	7:50: AM	37.63287	138.9675	8:23:AM
2	6/1/2015	2	37.50412	139.1178	7:51: AM	37.63309	138.9717	8:29:AM
3	6/1/2015	2	37.64898	138.9328	8:00: AM	37.63309	138.9717	8:10:AM
4	6/1/2015	2	37.64862	138.9393	8:13: AM	37.65328	138.9506	8:17:AM
5	6/1/2015	2	37.59501	138.9655	8:33: AM	37.6159	138.9477	8:43:AM
6	6/1/2015	3	37.52714	139.1182	9:15: AM	37.64862	138.9393	10:00:AM
7	6/1/2015	2	37.59258	139.0542	9:15: AM	37.64175	138.9735	9:35: AM
8	6/1/2015	2	37.61143	138.9941	9:27: AM	37.63639	138.9619	9:40: AM
9	6/1/2015	2	37.63685	138.9497	9:30: AM	37.60423	138.9875	9:46: AM
10	6/1/2015	2	37.6159	138.9477	9:47: AM	37.59501	138.9655	9:55: AM
11	6/1/2015	2	37.58343	138.9549	9:56: AM	37.6304	138.9512	10:10: AM
12	6/1/2015	2	37.6298	138.9879	10:00:AM	37.6448	139.0262	10:07: AM
13	6/1/2015	2	37.615	138.968	10:25:AM	37.62699	138.9397	10:38: AM
14	6/1/2015	2	37.5464	139.0775	10:30:AM	37.64175	138.9735	10:55: AM
15	6/1/2015	2	37.63639	138.9619	10:40:AM	37.61143	138.9941	10:55: AM
16	6/1/2015	2	37.64175	138.9735	10:46:AM	37.56554	139.0528	11:13: AM
17	6/1/2015	2	37.64175	138.9735	11:16:AM	37.59258	139.0542	11:45: AM
18	6/1/2015	2	37.62469	138.9558	11:30:AM	37.64175	138.9735	11:35: AM
19	6/1/2015	2	37.64175	138.9735	12:00: PM	37.62874	138.9568	11:55: AM
20	6/1/2015	2	37.63309	138.9717	12:10: PM	37.64898	138.9328	12:20: PM
21	6/1/2015	2	37.64175	138.9735	12:30: PM	37.5464	139.0775	12:55: PM
22	6/1/2015	3	37.62763	138.9781	1:30: PM	37.65328	138.9506	1:43: PM
23	6/1/2015	2	37.62699	138.9397	2:22: PM	37.615	138.968	2:34: PM
24	6/1/2015	2	37.61536	138.9733	3:50: PM	37.632	138.9737	3:57: PM

6.1.2 Taxi Trip Analysis: Travel Times, and Trip Frequency

Taxi trip analysis is the process of examining data related to taxi trips to gain insights into patterns and trends in travel times and trip frequency. This analysis can be useful for a variety of purposes, including improving transportation planning, optimizing routes, and understanding consumer behavior. Travel times refer to the duration of taxi trips, which can be analyzed to identify peak travel times and popular destinations. By examining travel times over time, it is possible to identify patterns in when and where people are traveling.

Trip frequency refers to the number of taxi trips taken in a given area over a certain period. By analyzing trip frequency data, it is possible to identify areas that are experiencing high demand for transportation services. This information can be used by transportation providers to optimize routes and schedules to better serve these areas.

6.1.2.1 Number of Trips in Monthly

The number of taxi trips in the monthly analysis is a process of examining the data related to the number of taxi rides taken during a particular month. This analysis is useful in understanding the travel behavior of individuals or groups over some time. To conduct a number of taxi trips in monthly analysis, data from taxi companies, ride-sharing services, or other transportation providers can be used. This data typically includes information on the number of rides taken, the pickup and drop-off locations, and the time and date of each trip. The number of taxi trips in monthly analysis can also be used to identify changes in travel behavior over time. For instance, it may be discovered that the number of taxi rides has increased or decreased significantly compared to the previous year or month. Figure 6.3 shows the average number of rides per month. The passenger demand varies every couple of days and is cyclical.

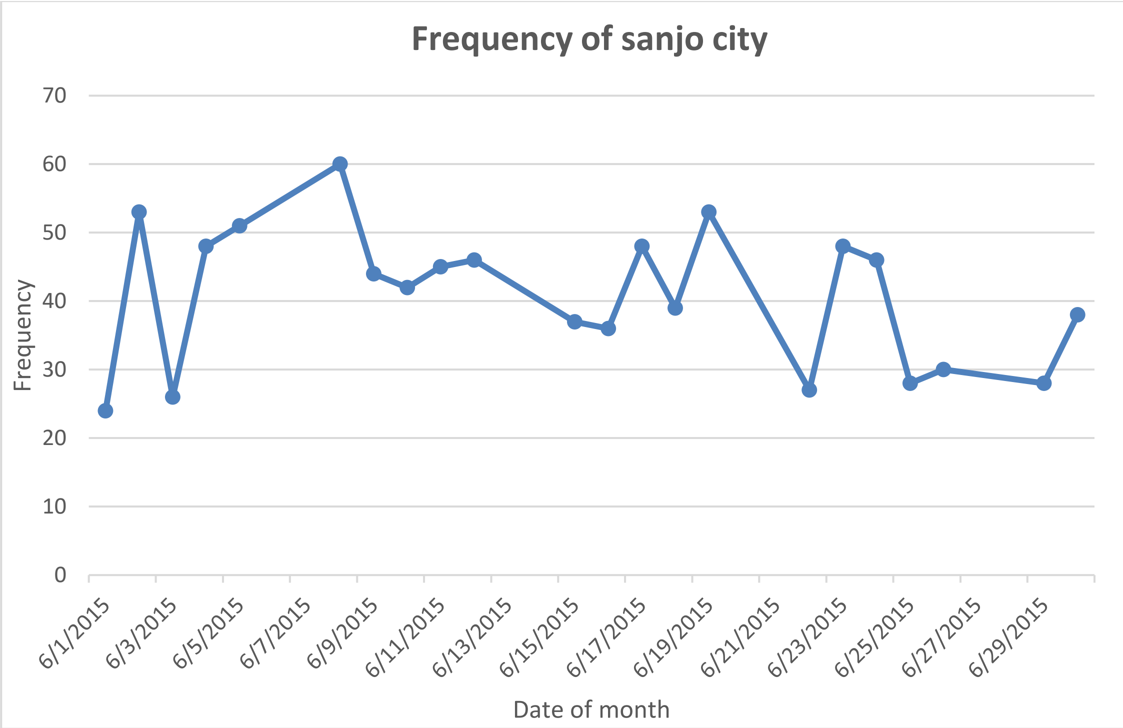

Figure 6.3. Taxi demand in the month (June 2015)

6.1.2.2 Number of Trips Per Hour of the Day

The number of taxi trips per hour of the day analysis is a process of examining the data related to the number of taxi rides taken during different hours of the day. This analysis is useful in understanding the travel behavior of individuals or groups during different times of the day. The number of taxi trips per hour of the day analysis is a valuable tool for transportation planners and providers to understand travel behavior and to optimize their services to better serve their customers during different times of the day.

By analyzing this data, it is possible to identify trends and patterns in the number of taxi trips taken during different hours of the day. For example, it may be found that the number of rides tends to increase during peak hours when people are commuting to and from work or school. Understanding these trends can help transportation providers to plan their services accordingly and allocate resources more efficiently.

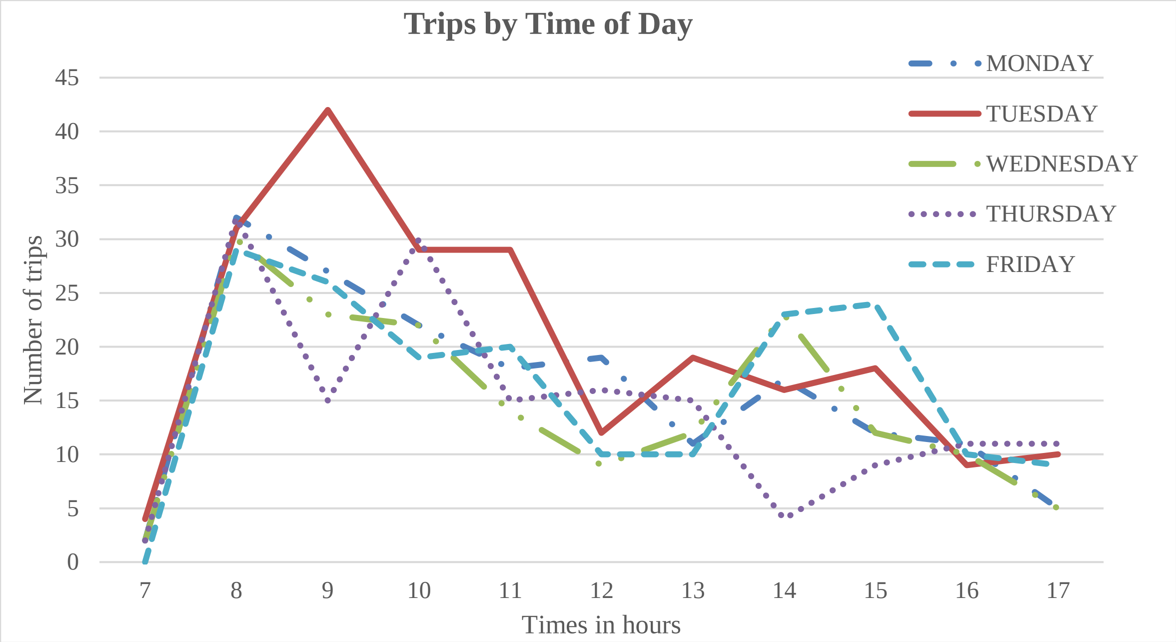

Figure 6.4. Number of taxi trips made at different times (06/2015)

Figure 6.4 shows the operating status by the time of day. The peaks are usually on ("8:00-9:00", and 9:00-10:00"). It can be confirmed that there is a peak demand after 8 am.

6.1.2.3 Weekday Trips Per Hours

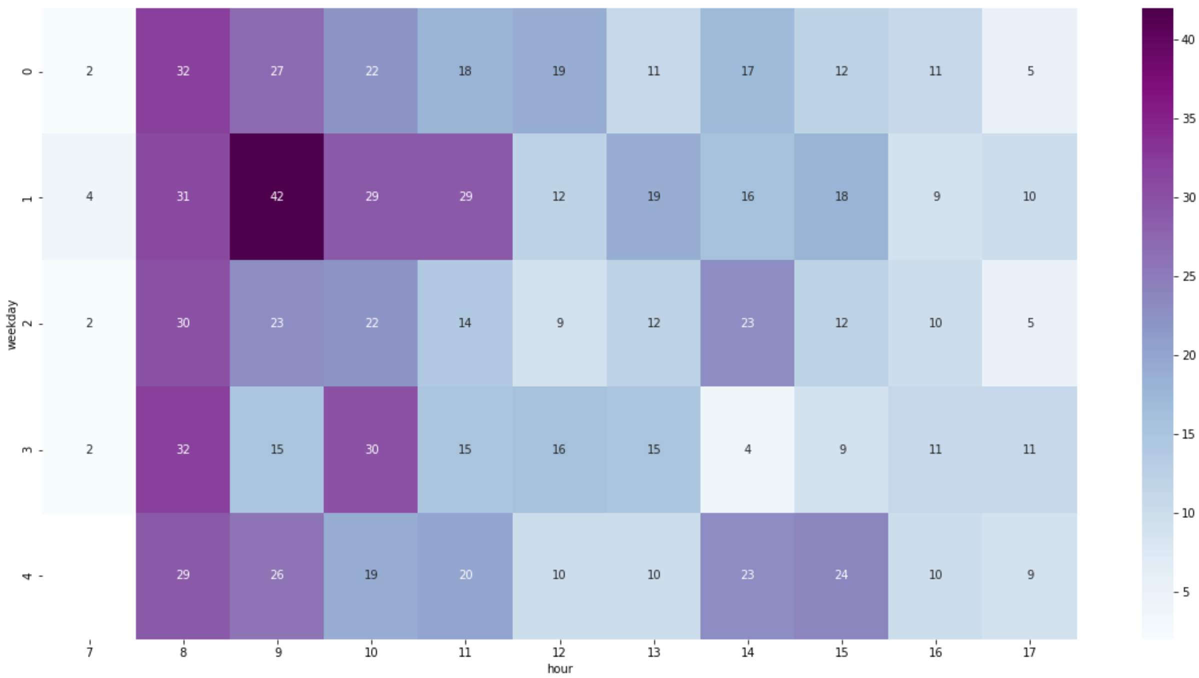

Figure 6.5. Cross-analysis of hours vs. weekday

Figure 6.5 shows the average number of rides by week and the time of the day. It can be confirmed that there is a lot of demand on the weekday. And shows the color-labeled value of suitability time. It is intuitive to expect the suitability time to increase with increasing travel time.

6.1.2.4 Data Description

Taxi data description typically refers to a collection of information related to the operation of taxi cabs in a given area. This data may include details about the pickup and drop-off locations, time and date of the trip, the fare charged, distance traveled, and other relevant information.

Taxi data is often collected by transportation authorities, taxi companies, or third-party providers using GPS tracking technology and other data collection methods. This data can be used for a variety of purposes, such as optimizing taxi routes and improving traffic flow, monitoring the performance of taxi drivers, and providing insights into transportation patterns in a given area.

Table 6.2 shows the input and output variables. Input variables are influenced by other variables in the system on which the output variable depends.

Table 6.2. Input and output data description

Input variable	Type and Description
Taxi ID	ID of the taxi
Latitude	Geographic coordinate expressing the north-south position of the taxi at the initial timestamp of the event
Longitude	Geographic coordinate expressing the east-west position of the taxi at the initial timestamp of the event
Service time	In hours and minutes
Early Time	In hours and minutes
Late time	In hours and minutes
Service cost	In Yen
The early taxi penalty cost	In Yen
The late pass penalty cost	In Yen
Late taxi penalty cost	In Yen
Demand	Continuous
Output Variable	**Description**
Number of taxis	Continuous
Cost	In Yen

For the entire process of taxi sharing, these studies focus on two key aspects: matching, and fare.

Table 6.3. The procedure, classification, and methods of taxi sharing.

<table>
<tr><th>Procedure</th><th>Classification</th><th>Methods</th></tr>
<tr><td>Taxi sharing matching</td><td>Passenger-based</td><td>Client agent calling a taxi together</td></tr>
<tr><td rowspan="2">Taxi sharing Fare</td><td>Number of passengers based</td><td rowspan="2">The travel distance as the decision variable</td></tr>
<tr><td>Travel distance based</td></tr>
</table>

6.1.3 Methodology

In order to solve the shared taxi fare problem using the Mixed-Integer Nonlinear Programming (MINLP) formulation described in Section 5.2.1, we use the Gurobi-9.0.1 environment via the Windows platform. The experiments were performed in a machine with an Intel(R) Core (TM) i7-5500U processor at 2.40 GHz, 16 GB of RAM, and a Windows 10 operating system.

Algorithm 6.1 solves the MINLP problem (5.1) at every iteration. We develop a Branch and Cut algorithm that generates valid inequalities dynamically during the Branch & Bound procedure. This functionality is offered by the Branch and Cut algorithm in Python using a Gurobi solver as a state-of-the-art computational feature. In a minimization problem, at each node, this problem would like to add cuts (obtained from a separation problem) to cut off fractional solutions. After the solving process, the taxi routes and service systems are found.

Algorithm 6.1 *The branch-and-cut algorithm for solving MINLP problems*

Step 0 (Initialization)

Generate an initial feasible solution $\boldsymbol{x_0}$ & compute the objective function $\boldsymbol{f}(\boldsymbol{x_0})$.

Set as the best solution $\boldsymbol{x}^* \leftarrow \boldsymbol{x_0}$ and $\boldsymbol{f}(\boldsymbol{x}^*) \leftarrow \boldsymbol{f}(\boldsymbol{x_0})$

Choose a set, $\boldsymbol{N_s}(\boldsymbol{s} = \boldsymbol{1}, \boldsymbol{2} .. \boldsymbol{s_{max}})$ and $\boldsymbol{N_k}(\boldsymbol{k} = \boldsymbol{1}, \boldsymbol{2} .. \boldsymbol{k_{max}})$

Step 1 Set $\boldsymbol{s} \leftarrow \boldsymbol{1}$

Step 2 MINLP formulation

Step 3 (Callback Procedure)

Step 3.1 (Cut Generation) Randomly create a solution $\boldsymbol{x}'$ from the $\boldsymbol{s^{th}}$ neighborhood $\boldsymbol{N_s}(\boldsymbol{x}^*)$ of the current solution.

Step 3.2 (Local Search)

Set $\boldsymbol{k} \leftarrow \boldsymbol{1}$

If the objective function decrease $\boldsymbol{f}(\boldsymbol{x}'') \leftarrow \boldsymbol{f}(\boldsymbol{x}')$

$\boldsymbol{x}' \leftarrow \boldsymbol{x}''$ and $\boldsymbol{k} \leftarrow \boldsymbol{1}$

Else

Set $\boldsymbol{k} \leftarrow \boldsymbol{k} + \boldsymbol{1}$

Until $\boldsymbol{k} = \boldsymbol{k_{max}}$

Step 3.3 (Move or not) the local optimum is better than the current solution.

$\boldsymbol{f}(\boldsymbol{x}') < \boldsymbol{f}(\boldsymbol{x}^*)$

Return to Step 1.

6.1.4 Results and Discussion

In this section, we present an optimization solution and sensitivity analysis of mathematical modeling.

The model is applied to the case study of Sanjo City. Sanjo City, Niigata Prefecture, is located in Niigata Prefecture and is classified in the Chuetsu region. The area is about 432 km^2. The number of households is about 36,611, the population is about 95,811, and about one in three citizens is elderly[6]. Figure 6.6 shows the simplified City location of Sanjo city based on Google Maps.

6 https://www.city.sanjo.niigata.jp/soshiki/shimimbu/shimimmadoguchika/madoguchi/madoguchi/4473.html

Figure 6.6. The case study area of Sanjo City

The shared taxi has been operating in Sanjo City, Niigata Prefecture, since June 2011. Accordingly, the fares have increased. It has been indicated that the cause is that about 80% of users use it alone.

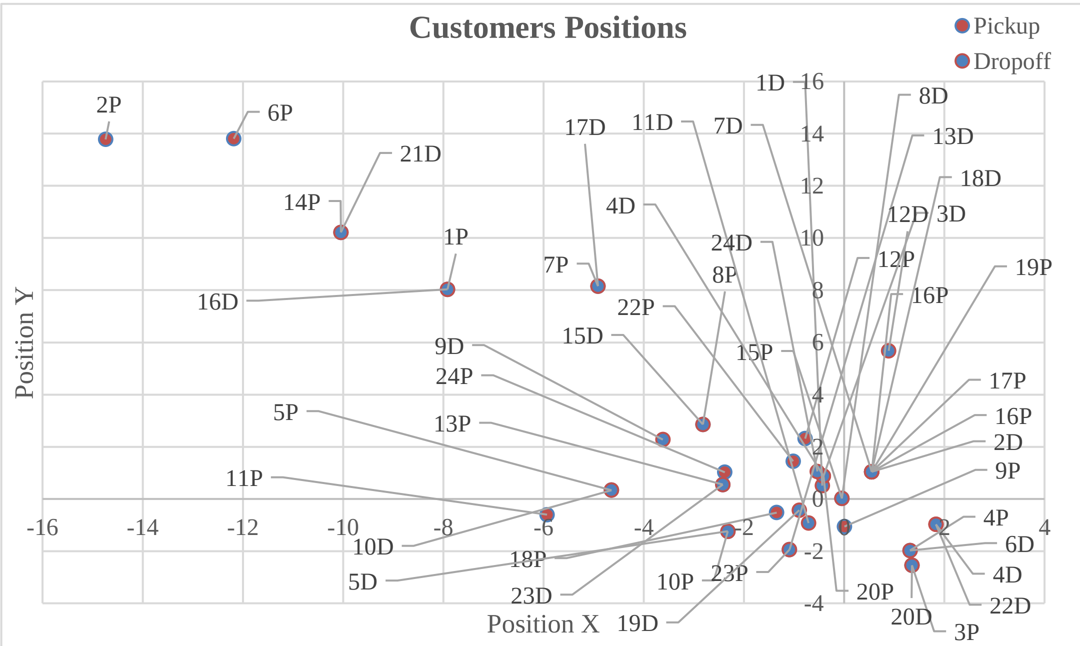

Figure 6.7. Customers positions

Figure 6.7 displays customers' position in Sanjo city. It is represented on a two-dimensional XY plateau; we have used 24 pickup nodes and 24 drop-off nodes. There are a total of 48 nodes for instance of customers in Sanjo city.

All passengers' pickup and drop-off needs, in particular, are arbitrarily manufactured. Although each whole instance contains 50 passengers, the initial few customers might be used to explore lesser sizes of the concerned representative. The cost of travel on an arc connecting two places was calculated using Euclidian distance calculations and rounded down to the closest integer if the difference was small. The total of the arc's travel cost and service time at the beginning point was used to calculate the travel time on the arc.

6.1.5 Optimization Results

There are three dimensions dependent on taxi speed, the number of taxis, and the seating capacity of each taxi. These are used for evaluating model performance. The time windows, initial fare rate, and distance-based fare are varied to conduct a sensitivity analysis concerning the cost components. The optimization results can be seen in Figure 6.9 by solving the optimal model and obtaining the optimal solution. The results have been carried out using Sanjo City taxi data. The experiment's major goal is to guarantee that all taxi rides are serviced at the same time and place as the data suggests. Meanwhile, we determine the best technique for (1) minimizing the cost of waiting time, distance-based cost, and total trip cost for all taxis, and (2) locating the fewest taxis.

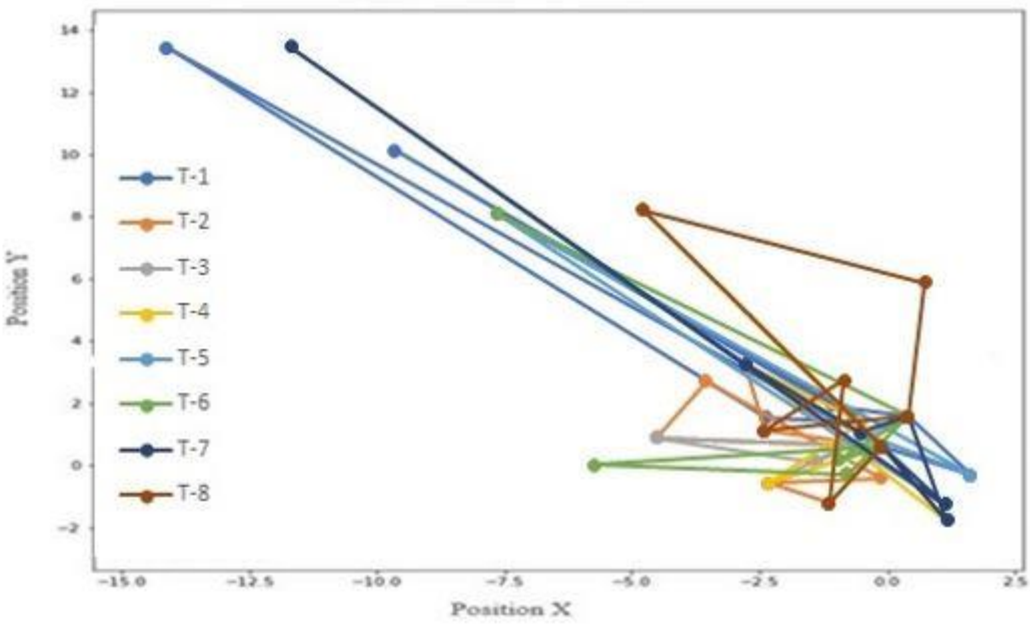

Figure 6.8. The original route before optimization

Figure 6.8 represents the regular taxi route of Sanjo City. After the taxi trip route experiments, we consider the time window of 10 min and taxi speed of 20 Km/h, and the distanced -based fares of 630 yen up to 1200 meters and additional fares of 90 yen every 294 meters after that.

After the optimal solution of the mathematical model (5.1) in the case study, we minimize the number of taxis and taxi shortest distances in Figure 6.9.

From Figure 6.9, we can confirm for the model that the total taxi takes the minimum number of taxis and shortest distance.

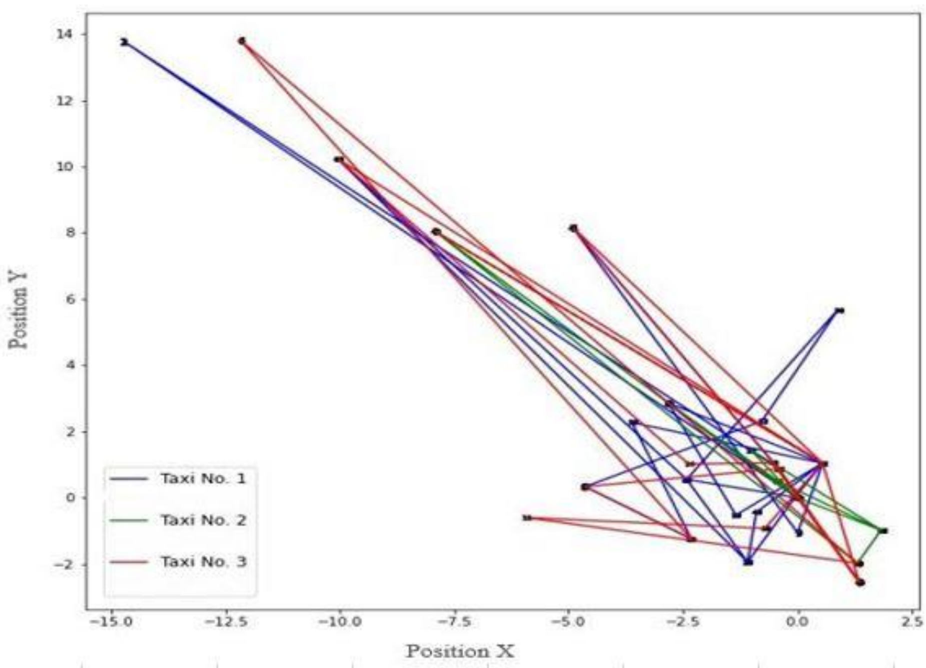

Figure 6.9. The optimal solution of MILP programs.

Table 6.4. Solution for one-day taxi trips

No of passengers (Trips)	**Time windows**		**Initial Fare (Yen)**	**Discount of waiting time (%)**	**Distance -based fare (Yen)**	**Initial distance in fixed fare (Km)**	**Number of taxis uses**	**Waiting time cost (Yen)**	**Distance based cost (Yen)**	**Total cost (Yen)**
	ITW (Min.)	LTW (Min.)								
50(24)	10	10	630	[0.1,0.3]	90	1.2	3	4,001	12,440	13,975

6.1.6 Sensitivity Analysis

This section discusses the sensitivity of significant parameters, including time windows, initial fare, and distance-based fare. Analysis is vital in assessing the validity of the model. Sensitivity analysis is a technique used to determine how changes in the input variables of a model affect the output of the model. In the context of a shared taxi fare model, sensitivity analysis can help determine the impact of changes in key input variables on the overall fare. Some key input variables for a shared taxi fare model might include time windows, initial fare, and distance-based fare.

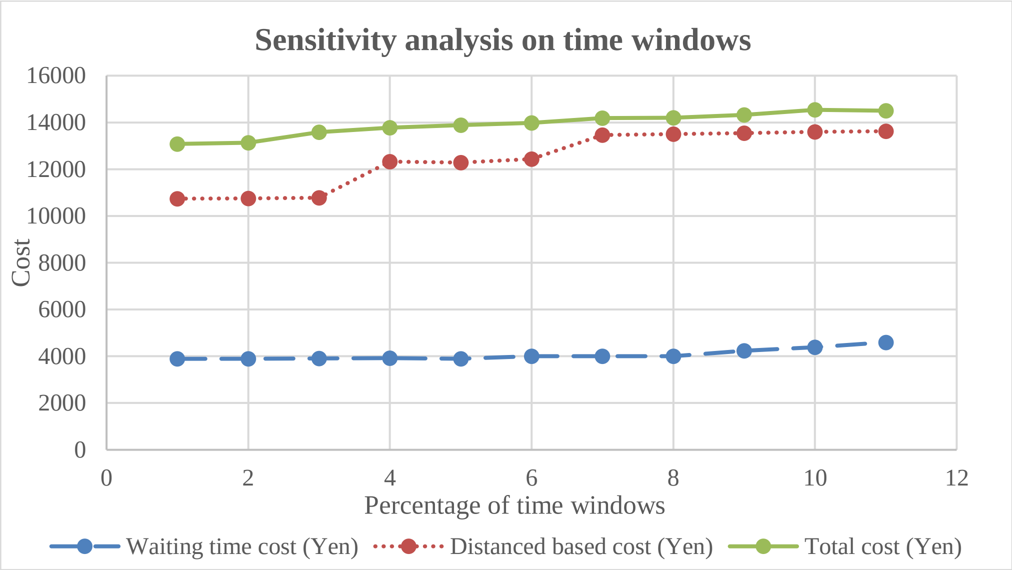

Figure 6.10. Sensitivity analysis on time windows

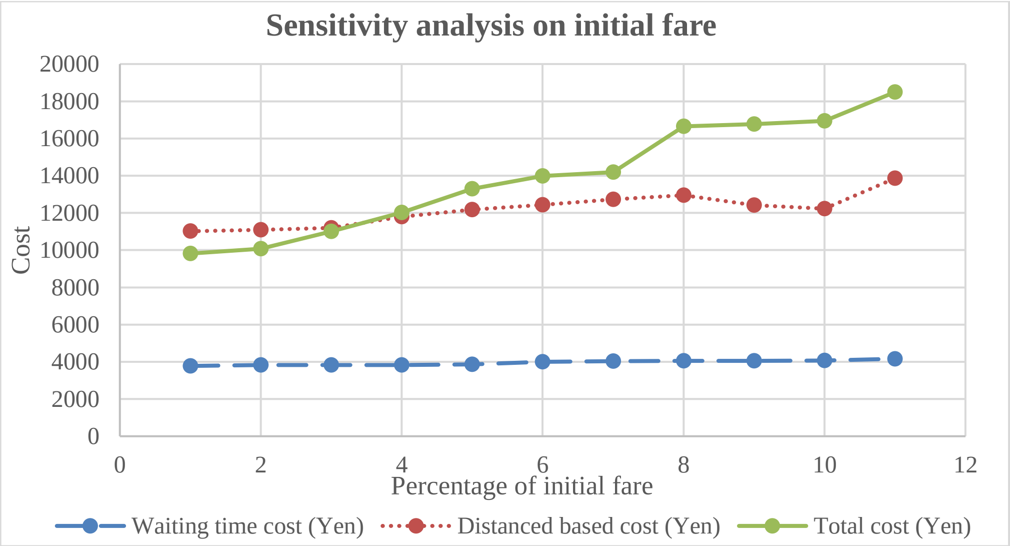

Figure 6.11. Sensitivity analysis on initial fare

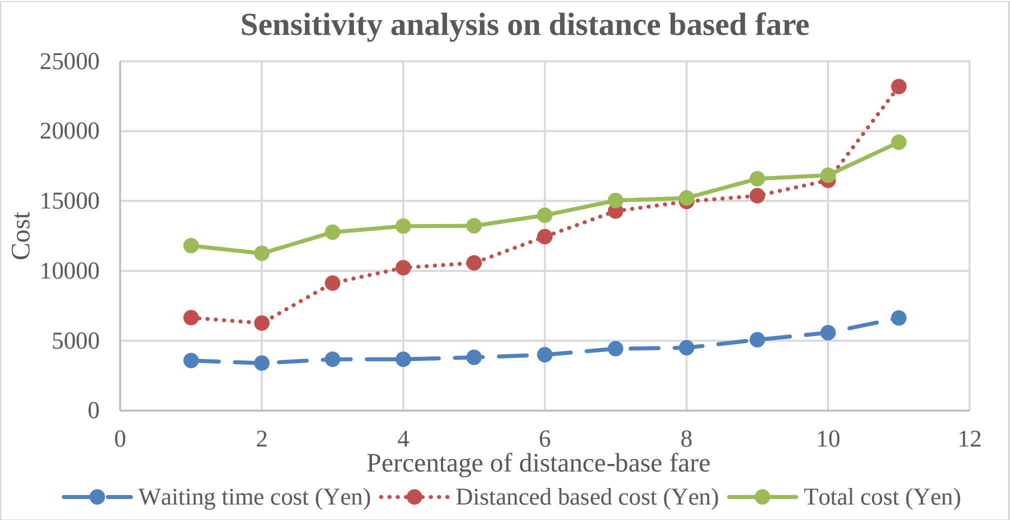

Figure 6.12. Sensitivity analysis on distance-based fare

The sensitivities are carried out by increasing and decreasing the value of the parameters by 10%. From the sensitivity, it is clear that the waiting time cost, distance-based cost, and total cost increase in Figure 6.10 if the time windows increase. Figure 6.11 depicts that the initial fare positively influences the waiting time cost, distance-based cost, and total cost. Figure 6.12 illustrates that the influence of distance-based fare is proportional to the waiting time cost, distance-based cost, and total cost.

6.1.7 Summary of the Shared Taxi Model

This chapter proposes a Mixed Integer Non-Linear Programming formulation model to investigate the DARP of shared taxis during congested travel periods caused by the ride sharers. The model's main goal is to reduce the overall cost of a passenger's journey as well as the number of taxis used.

To begin, we used GPS data to determine the condition of the Sanjo city ridesharing taxi. The concept and solution technique was tested in the case study of Sanjo, Japan. The experimental findings demonstrate that the MINLP solver can identify the best solution to the issue in an acceptable amount of time for scenarios with 50 passengers. The model was solved using a Gurobi solver and the Branch and Cut procedure in Python. This research is significant because it addresses a genuine issue that was handled using an accurate approach of MINLP for consumers with capacity, demand, and time frame constraints. It can be seen that there is a lot of demand on weekdays. It's natural to assume that when journey time increases, the suitability time would grow as well.

The model has validated the model sensitivity of different parameters such as time windows, initial fare, and distance-based fare are vital. The sensitivities have been carried out by increasing and decreasing the value of the parameters by 10%. The sensitivity shows that the waiting time cost, distance-based cost, and total cost increase if the different parameters increase.

This formulation may be used to create heuristics for the issue of shared taxi fares, perhaps speeding up the optimization process. More work may be done into enhancing the metaheuristic method and reducing the computing gap in the suggested model. A future study might lower the driver's idle time and take into account performance improvements.

6.2 A Case Study of the Nagaoka Taxi Sector for Taxi Allocation

Nagaoka is located in the Niigata area of Japan, as well as the adjacent Chetsu region. The city has a population density of 300 residents per square kilometer (780/sq mi) as of August 4, 2021, with a population of 264,611 people living in 109,283 homes.

Table 6.5. Number of taxis in the case study area (Nagaoka)

Name of the taxi company	**Number of vehicles**
Asahi Taxi	29
kankô Taxi	45
Sôgo Taxi	48
Chûetsu Taxi	21
Tsubame Taxi	36
Nagaoka Taxi	39
Mitsukoshi Taxi	50

Demand for taxis in local cities is sparse in terms of time and space (see Table 6.5), which shows the number of taxis owned by each taxi company in Nagaoka City, the number of taxis owned by Mitsukoshi Taxi, the largest taxi company, is Nagaoka. It is about 20% of the city as a whole, and the share rate of Mitsukoshi taxis is considered to be about 20%. The number of actual vehicle trips in Nagaoka City is considered to be about 450 trips per day and the share rate is 20% even for the largest taxi companies. Nagaoka City as a whole is only about 2,300 trips), and it is very difficult to find new passengers near the place where the passengers were dropped off with the demand of only one company. Therefore, this study focused on analyzing idle time and reducing taxi numbers to optimize taxi operations thus reducing CO_2 emissions. This study aims primarily to decrease the amount of CO_2 emissions by optimizing the allocation of taxis. The study proceeds in three sequential steps, which are as follows::

1. We analyze taxi Global Positioning System (GPS) data to detect taxi hotspots where drivers and passengers appear frequently in suburban areas. This research easily and formally identifies areas with high demand and quantifies those hotspots.
2. We compare the performance of the proposed algorithm with two heuristic algorithms, namely greedy and simulated annealing. Further, we evaluate the optimality of the model by applying it to a real case study in Nagaoka City, Japan.
3. Even though the recent articles (Zhang *et al.* (2020); Arabani and Balooch Sirgani (2022)) analyzed the CO2 emissions in taxi operations, the optimization of the idle time and the number of taxis still needed to be considered. Therefore, this study focused on analyzing idle time and reducing taxi numbers to optimize taxi operations thus reducing CO2 emissions.

6.2.1 Taxi Demand Hotspots

For reducing CO2 emissions, this study focuses on reducing taxi's idle time by optimizing taxi allocation in hotspot locations based on the high taxi demand. Therefore, taxi hotspot detection is necessary to consider appropriate taxi allocation. This paper detects taxi hotspots by using real data. The data collection starts by considering taxi usage patterns, including pick-up and drop-off locations, trip distances, and times of the day. This data will allow us to identify high-traffic areas and peak usage times that can be utilized to optimize taxi allocation.

In this study, Nagaoka sub-urban area was selected as the case study area. Nagaoka is in the center of Niigata and the nearby Chūetsu economic region of Japan and is on the national highway network. According to population data from August 4, 2021, there are 264,611 people living here and 109,283 households, with 300 people per square kilometer of population density (780 square miles). The total land extent of the city is 891.06 square kilometers. We selected Nagaoka because alternative public transportation is frequently unavailable in this location. After 9:30 p.m., there are no buses inside the city. Further, since the majority of individuals are elderly and unable to drive in this Nagaoka sub-urban area, they rely on public transportation. In order to solve these kinds of problems, the general taxis work 24 hours a day. A taxi driver works at a specific time every day and sometimes works overtime. The dynamic properties of the identified taxi demand hotspots are analyzed in this study. The selected city of Nagaoka is in the center of Niigata and the nearby Chūetsu eco-nomic region of Japan and is on the national highway network.

6.2.2 Research Data

Mitsukoshi Taxi Co., Ltd. is the largest taxi company with a business area in Nagaoka City. The Nagaoka Taxi Company provided the taxi data for this study. Their 60 taxis were equipped with the Global Positioning System (GPS), which enables vehicle identification in a location-specific manner with date, and time (i.e., 1/1/2019 11:32:00 unique key), latitude and longitude (i.e., 37.4634 and 138.8147 degrees), and other relevant real-time information as depicts in Table 6.6. The origin and destination are considered places at the beginning of the ride and the arrival. The data was collected over a year in 2019, and 163,532 valid data were identified. This data can be utilized for detecting taxi demand information with dynamic features.

Table 6.6. Selected data for the study.

No	**Pick-up Date Time**	**Drop-off Date Time**	**Pickup Longitude** (degree)	**Pickup Latitude** (degree)	**Drop-off Longitude** (degree)	**Drop-off Latitude** (degree)
1	1/1/2019 11:32	1/1/2019 11:46	138.8247	37.4634	138.8488	37.44448
2	1/1/2019 12:28	1/1/2019 12:39	138.8233	37.44263	138.8528	37.44718
3	1/1/2019 14:22	1/1/2019 14:38	138.8289	37.46078	138.8526	37.42944
4	1/1/2019 15:09	1/1/2019 15:21	138.829	37.46065	138.8388	37.44359
5	1/1/2019 16:01	1/1/2019 16:21	138.7792	37.44842	138.8517	37.43882
.	.	.	.	.	.	.
.	.	.	.	.	.	.
163,532	12/31/2019 5:19	12/31/2019 5:56	138.8525	37.44682	138. 999	37.47461

After preprocessing data, we can study taxi journeys from different perspectives. Here identified two specific types of characteristics: the geographic characteristics of positions (i.e., longitude, latitude) and trip characteristics (i.e., trip demand, distance, day of the week, time of day, day of the week). The taxi trip data includes average daily passenger travel times, distances, and waiting times. The analysis of the travel characteristics between the taxi demand locations and the times is also helpful for presenting events in the future.

6.2.3 Hourly Taxi Demand Locations

Analyzing hourly taxi demand locations in a suburban area can provide valuable insights for taxi companies, transportation planners, and policymakers. This information can help them identify patterns and trends in taxi usage, which can inform decisions related to pricing, service areas, and infrastructure development. After identifying the hotspots, we analyze the data to determine the reasons for the high demand. For example, we might find that certain areas have high demand during rush hour because of their proximity to major employment centers or transportation hubs. Once we have the data, we start by visualizing the hourly taxi demand locations on a map to identify the hotspots or areas with the highest demand. This can be done using geographic information system (GIS) tools, which help us create density maps of taxi pickups and drop-offs.

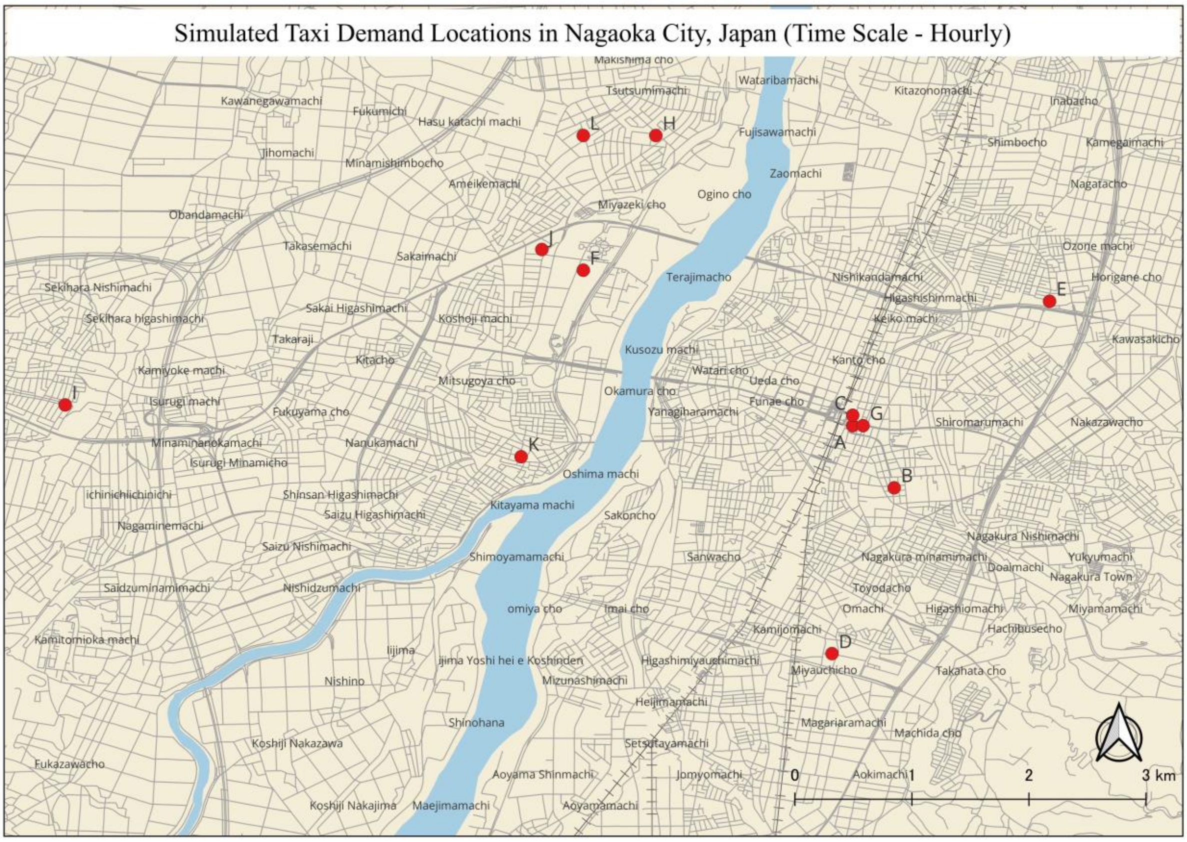

*Figure 6.13***.** Locations of taxi demand per hour.

*Location details for the taxi demand areas are listed in Table 6.7 (Locations ID).

Alternatively, we might find that certain areas have high demand during late-night hours because of their proximity to entertainment venues or residential areas. With this information, taxi companies can adjust their pricing and service areas to better meet the needs of their customers.

Table 6.7. Hourly taxi demand locations.

No	Time of day	Number of trips	Total trips	Address	Locations ID
1	0:00-1:00	7,901	13,153	LiVE MAX, Higashiguchi Street, Nagaoka, Niigata, 940-0033, Japan	A
2	1:00-2:00	5,252			
3	2:00-3:00	3,018	3,018	Shiromaru 2-chome, Nagaoka, Niigata, 940-0041, Japan	B
4	3:00-4:00	1,381	4,250	E-PLAZA, East Exit Street, Nagaoka, Niigata, 940-0033, Japan	C
5	4:00-5:00	613			
6	5:00-6:00	812			
7	6:00-7:00	1,444			
8	7:00-8:00	2,922	2,922	Kamijo-cho, Nagaoka, Niigata, 940-8621, Japan	D
9	8:00-9:00	7,517	7,517	Horikane 1-chome, Nagaoka, Niigata, 940-8653, Japan	E
10	9:00-10:00	9,883	28,253	Nagaoka Red Cross Hospital, Nagaoka, Niigata, 940-2085, Japan	F
11	10:00-11:00	9,846			
12	11:00-12:00	8,524			
13	12:00-13:00	7,733		E-PLAZA, East Exit Street, Nagaoka, Niigata, 940-0033, Japan	C
14	13:00-14:00	6,941		Daishi Bank, East Exit Street, Nagaoka, 940-0033, Japan	G
15	14:00-15:00	7,814	30,953	LiVE MAX, Higashiguchi Street, Nagaoka, Niigata, 940-0033, Japan	A
16	15:00-16:00	7,429			
17	16:00-17:00	7,295			
18	17:00-18:00	8,415			
19	18:00-19:00	8,973	8,973	Hasugata 4-chome, Nagaoka, Niigata, 940-2088, Japan	H
20	19:00-20:00	9,004	9,004	Nagaoka Red Cross Hospital, Nagaoka, Niigata, 940-2085, Japan	F
21	20:00-21:00	9,633	9,633	Kamiwake-cho Nishi 1-chome, Nagaoka, Niigata, 940-2035, Japan	I
22	21:00-22:00	11,245	11,245	Hasugata 5-chome, Nagaoka, Niigata, 940-2093, Japan	J
23	22:00-23:00	11,243	11,243	Oshima Honmachi 4-chome, Nagaoka, Niigata, 9402104, Japan	K
24	23:00-00:00	8,694	8,694	Shimoyanagi 3-chome, Nagaoka, Niigata, 940-2088, Japan	L

Therefore, analyzing hourly taxi demand locations in a suburban area can provide valuable insights that can inform decisions related to pricing, service areas, and infrastructure development. By understanding the patterns and trends in taxi usage, taxi companies, transportation planners, and policymakers can better meet the needs of their customers and improve the overall transportation system. Table 6.7Table *6.7* depicts the hourly aggregated data for clarifying hotspot changes in the morning, afternoon, and night. In the morning (9:00-12:00), taxi demand is high in the Nagaoka Red Cross Hospital area. Taxi trips were covered in the supermarket area in the afternoon (13:00-18:00). In the night, station and restaurant areas were high taxi demand for service.

The consistent hotspots were determined mainly by passenger traffic in various time segments. Railway stations, being major hubs for moving people between cities, handled massive traffic. However, the majority of the hotspots seemed to work only at certain times. Therefore, hotspots were concentrated around stations, hospitals, restaurants, and business areas, as depicted in Figure 6.13.

6.2.4 Weekly Taxi Demand Locations

Table 6.8 depicts the taxi demand variation pattern on different days within the week. The number of daily trips was counted at taxi pick-up/drop-off points. The analysis showed that the highest and lowest demand for taxis per week occurred on Fridays and Sundays, respectively. This was consistent with the experience that Friday is the last working day of the week, when people tend to go out for recreation after work, thus resulting in more taxi trips. In contrast, Sunday is the second day of the weekend, and people prefer to rest at home to prepare for the workweek ahead, thus resulting in fewer taxi trips. Kernel density analysis was performed on the 24-hour data from Friday to quantify the dense grade of taxi demand and visualize taxi travel hotspots at the spatial level.

The results indicate five hotspots, including No. 1, 2, 4, and 5 at the railway station or supermarket, while no. 3 located in Nagaoka Red Cross Hospital areas, see Table 6.8. Large public spaces where people and vehicles converge and deliver, stations are usually equipped with taxi locations at specific locations to meet passenger travel needs. Table 6.8 depicts the network of hotspot no. 1, covering 29,541 taxi trips on Friday. But three days (Monday, Wednesday, and Thursday) are high taxi demand covering 65,868 trips in the Nagaoka Red cross Hospital area.

Table 6.8. The top 5 busiest locations on the day of the week.

No.	Address	Number of Trips	Days of the Week	Locations ID
1	Hasugata 5-chome, Nagaoka, Niigata Prefecture, 940-2093, Japan	29,541	Friday	J
2	Ojimahoncho 4-chome, Nagaoka, Niigata Prefecture, 940-2104, Japan	27,437	Saturday	K
3	Nagaoka Red Cross Hospital, Nagaoka, Niigata, 940-2085, Japan	23689 22,914 19,265	Wednesday Thursday Monday	F
4	Kaminozokimachi Nishi 1-chome, Nagaoka, Niigata, 940-2035, Japan	22,642	Tuesday	I
5	Nagaoka Station, Nagaoka City, Niigata Prefecture, 940-0061, Japan	18,044	Sunday	M

In contrast, the city has several major routes where the high traffic leads to a significant need for taxi spots. Such hotspots with a dispersed network state are acceptable as a study area for taxi location citations (Figure 6.14).

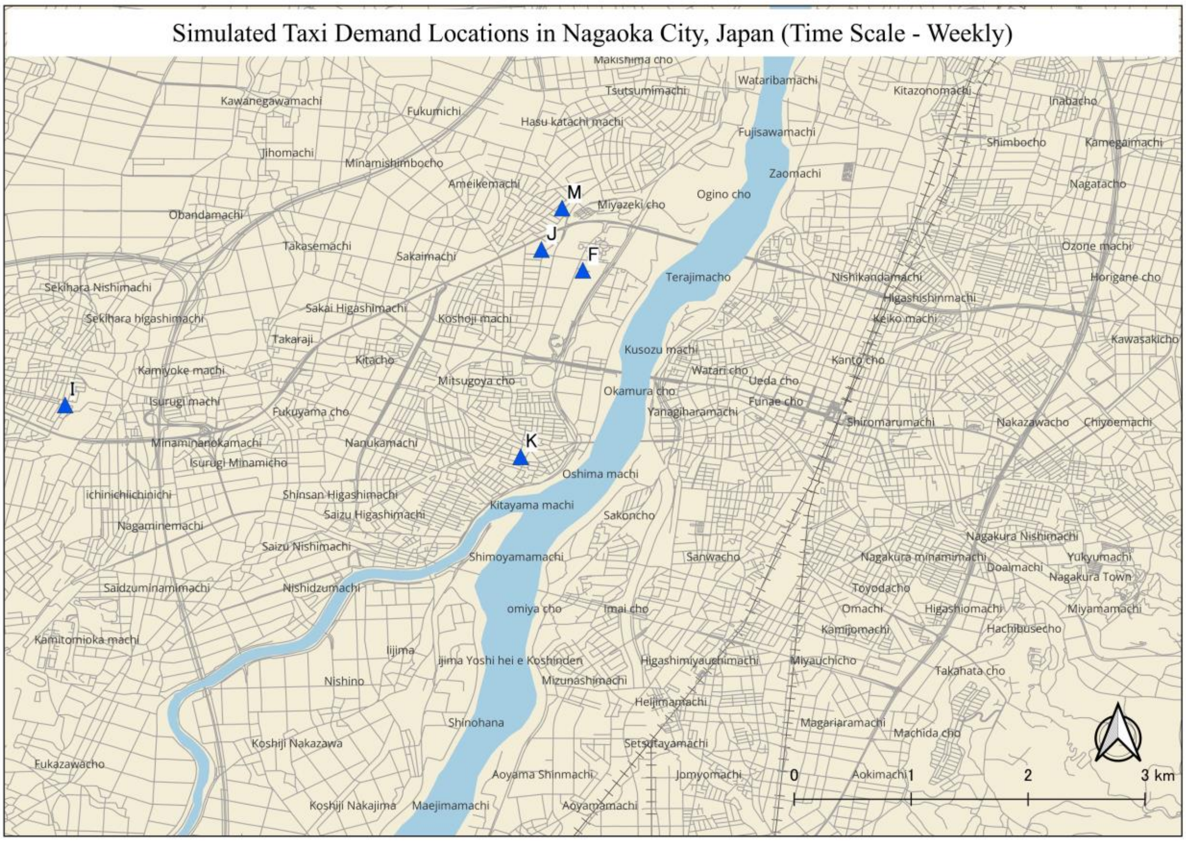

Figure 6.14. Taxi locations of the days of the week.

*Location details for the taxi demand areas are listed in Table *6.7* and Table 6.8 (Locations ID).

Represented as the blue area in Figure 6.14, the constant hotspots are mainly at Nagaoka Red Cross Hospital (F), Nagaoka Railway Station (M), supermarket (I), and so on. The constant hotspots mainly depend on passenger demand with respect to different days. In this study, taxi data are used to determine the potential hotspot area for taxi stand locations. In suburban areas, the drivers and passengers' behaviors affect the new taxi stand. So, this paper has formulated the mathematical model of the taxi fleets utilization problem to optimize taxi driver idle time by using taxi data, which may be a benefit for passengers and drivers.

Table *6.9* shows the input and output variables. Input variables are influenced by other variables in the system on which the output variable depends.

Table 6.9. Input and output data description

Input variable	Type and Description
Month (12 months)	Categorical: Jan, Feb,⋯, Dec.
Day of month (365 days)	Categorical: #1, #2,⋯, #31.
Time of day	In hours and minutes
Demand	Continuous
Idle time cost	In yen
Output Variable	Description
Number of taxis	Continuous
Cost	In Yen

For the entire process of taxi utilization, these studies focus on time aspects. Table *6.10*. summarizes these aspects.

Table 6.10. The procedure, classification, and methods of taxi utilization

Procedure	**Classification**	**Methods**
Taxi driver idle time cost	Travel distance based	The travel distance as the decision variable

6.2.5 Methodology

The pick-up and drop-off problem is well-established as a non-deterministic polynomial-time hardness (NP-hard) problem, which means that the computational time required to solve it increases rapidly as the size of the expansion Ghilas *et al.* (2016); Lu *et al.* (2018); Santos and Xavier (2013). This issue can be solved using two different kinds of algorithms. The first kind, known as an exact algorithm, is based on a mathematical model that guarantees the optimal result in every situation. However, the biggest disadvantage is the computing time necessary to solve the issue, particularly if it is NP-hard. This is referred to as a set of heuristic algorithms, which use advanced mathematical optimization techniques such as branch-by-branch and neighborhood search.

Consequently, heuristic algorithms take less time to execute than accuracy algorithms. Although it has a quick processing time, the optimal solution is not guaranteed, and the results may not be satisfactory. Construction heuristics, local search heuristics, and metaheuristics are the three types of heuristic techniques to make a new heuristic algorithm.

In this research were used a linear mixed-integer programming model. The model ((5.2)-(5.9)) can be solved by heuristic and metaheuristic algorithms. We have devised a technique for solving the taxi issue that entails the establishment of (i) a greedy heuristic (Daniel Fuentes and Sano (2020)), (ii) a local search strategy (Daniel Fuentes and Sano (2020)), (iii) a simulated annealing (Yu *et al.* (2021)) metaheuristic to solve the optimization problems, and (iv) dynamic greedy algorithm to improve the greedy algorithm.

6.2.5.1 Greedy Algorithm

The number of drivers, assignment costs, and the area that the first solution must have been all inputs to the heuristic method function. It begins by initializing all global parameters, such as the total number of passengers still requiring pick-up and drop-off, the global time, global distance, global service time, global waiting time, and total costs. The first iteration process begins when these variables are established, and it continues until all of the taxis available have been utilized. Once within the first iteration, the taxi driver parameters (i.e., total time, travel time, service time, waiting time, and costs) are set to their starting states, including the places that the taxi has previously chosen and its initial position and pick-up location. After collecting all pick-up sites from the passenger positions, the heuristic technique enters the second iterative step. This process continues until the present taxi runs out of drop-off sites or it is time to return to the depot.

It is common to use greedy heuristics to create effective initial solutions to difficult optimization problems. One can first, and sometimes intuitively, find an initial possible solution, which is subsequently improved by heuristics. To obtain a possible initial solution to the problem, an attractively structured heuristic is developed that considers the demand for all taxi stands and periods, prioritizes demand with high profit, and checks which taxis can be moved empty when necessary.

In the pseudocode presented for the greedy algorithm, the proposed greedy heuristic method is described as generating a relatively good initial solution to the taxi allocation problem. The basic idea is to check if there is a taxi that does not meet every pick-up that "arrives on time". "Arrival time" refers to the travel time from the current location of pick-up within the requested time. It is also necessary to check whether the taxi route (empty or not) can be used to meet the demand.

We have taxi ride data for a set of drivers, and we want to optimize the idle time cost.

Here's the greedy algorithm would work:

Step 1: Start with an initial solution. This could be the initial sequence of rides for each taxi driver.

Step 2: Evaluate the cost of the initial solution. This would involve calculating the total idle time cost of all the taxi drivers.

Step 3: Determine the set of candidate solutions that can be obtained by making a small modification to the current solution. For example, we could consider swapping the order of two consecutive rides for a particular taxi driver.

Step 4: Evaluate the cost of each candidate solution. This would involve calculating the taxi drivers' total idle time cost for each candidate solution.

Step 5: Select the candidate solution that provides the best immediate improvement in the objective function value and make it the new current solution. In this case, we want to minimize the total idle time cost of all the taxi drivers, so we would select the candidate solution that has the lowest idle time cost.

Step 6: If the stopping criterion is not met, repeat steps 3-5 until no more candidate solutions can be found or until the stopping criterion is met. The stopping criterion could be a maximum number of iterations or a minimum improvement in the objective function value.

Step 7: Return the current solution. The solution returned by the Greedy Algorithm would be the sequence of rides for each taxi driver that has the lowest total idle time cost.

The greedy works step-by-step based on **Algorithm 1** for optimizing idle time cost in taxi ride data.

Algorithm 1: Greedy Heuristic

```
Function greedy
    Greedy solution = Solution()
    For each client in the total clients
        Minimum costs = np.inf
        Minimum costs driver = 0
        For each driver in the total drivers
            Current costs = calculate costs (client, driver, Greedy
            solution)
                If the current costs < minimum costs
                    Minimum costs driver = driver
                    Minimum costs = current costs
        Assert minimum costs driver != 0
        Assign client (client, minimum costs driver, Greedy solution)
    Return all to depot (Greedy solution)
    Costs = total costs (Greedy solution)
    Greedy solution (total costs) = costs
    Assigned drivers = 0
    For each driver in the total drivers
        If the Greedy solution in the current driver assigned for each client
            Assigned drivers += 1
Return Greedy solution
```

Taxi data (trip data) is input into memory. All empty and onboard passenger movement of the taxi is recorded in a data structure, including route, time, and cost. All passenger pickups are selected in sequential taxi calling, and if there are multiple pickups simultaneously, they are prioritized depending on the distance. The purpose of the problem is the value of the function. In the initial solution, the taxi driver calculates the cost of waiting time. It is the responsibility of the taxi driver if the current cost is less than the minimum cost. Finally, the optimal solution is evaluated by the greedy algorithm.

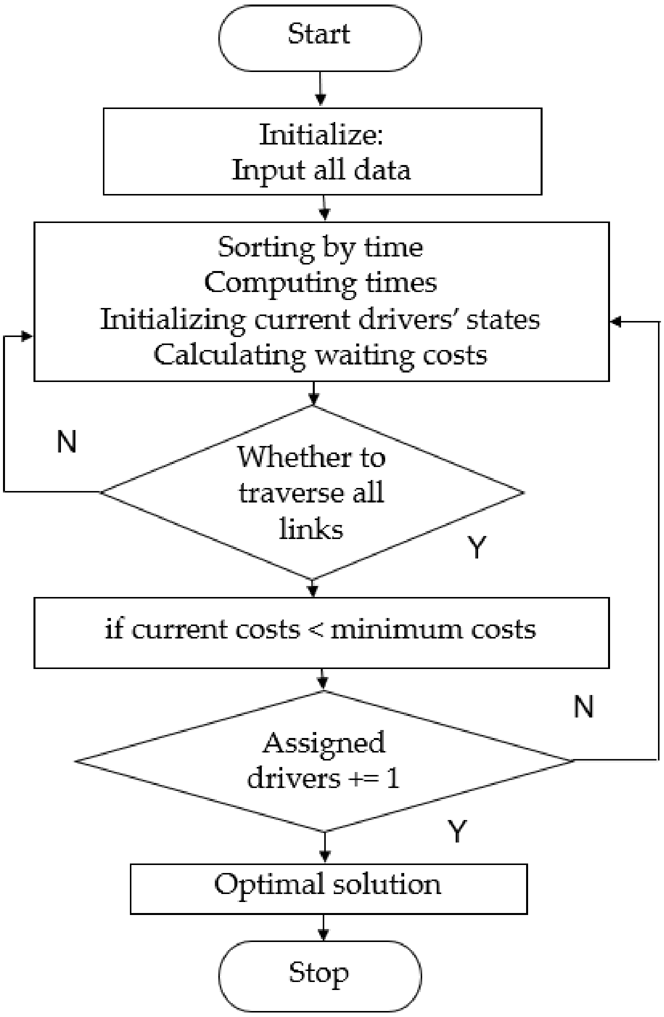

Figure 6.15. Greedy algorithm flowchart

The Greedy algorithm is presented in the flowchart in Figure 6.15. This is a diagrammatic representation of the algorithm. Taxi data (trip data) is input into memory. In the initial solution, the taxi driver calculates the cost of waiting time. It is the responsibility of the taxi driver if the current cost is less than the minimum cost. Finally, the optimal solution is evaluated by the greedy algorithm.

6.2.5.2 Local Search

Greedy developed an advanced method to perform from a greedy starting solution to investigate unmet pick-ups, which would be profitable. The local search method looks like the following. Here the steps of a local search algorithm are:

Step 1: Initialize the current solution to a random or heuristic solution.

Step 2: Evaluate the quality of the current solution using an objective function.

Step 3: Repeat the following steps until a stopping criterion is met:

a. Generate a neighboring solution by making a small perturbation to the current solution.

b. Evaluate the quality of the neighboring solution.

c. If the neighboring solution is better than the current one, accept it as the new one.

d. If the neighboring solution is worse than the current solution, accept it with a probability that depends on the difference in quality between the two solutions and the current temperature.

e. Update the temperature according to a cooling schedule.

Step 4: Return the best solution found during the search.

The following **Algorithm 2** illustrates a basic outline of a local search algorithm, which can be used to solve various optimization problems. However, the specific implementation of the algorithm and the choice of the objective function, neighborhood structure, and cooling schedule depend on the problem being solved.

Algorithm 2: Local Search

```
Function ImproveSolutions(Initial Route):
    Get local (no iteration)
    Current solution = randomly
    NUMBER ITERATION = no iteration
    Minimum cost = current solution (total costs)
    Values = zeros (no iteration)
    For each position in NUMBER ITERATION
        New solution = neighbor (current solution)
        If new solution (total costs) < minimum cost
            Minimum cost = new solution (total costs)
            Current solution = new solution
            Each position values = new solution (total costs)
    Assigned drivers = 0
    For each position in the Number of drivers
        If current solution (each position of current driver assigned)
            Assigned drivers += 1
Return Current solution, Values
```

The Local search algorithm is presented in the flowchart in Figure 6.16. Let the current drivers' states be complete, and drivers are returned to the depot. The first step is to input the relevant data into the formulation. Then, if necessary, the objective function is emptied. Next, randomly generate drivers and insert the copied passengers into a random location on the path of the second driver for the new solution. A model is formulated and calculated for all categories that should be included in the current problem neglecting the greedy solution.

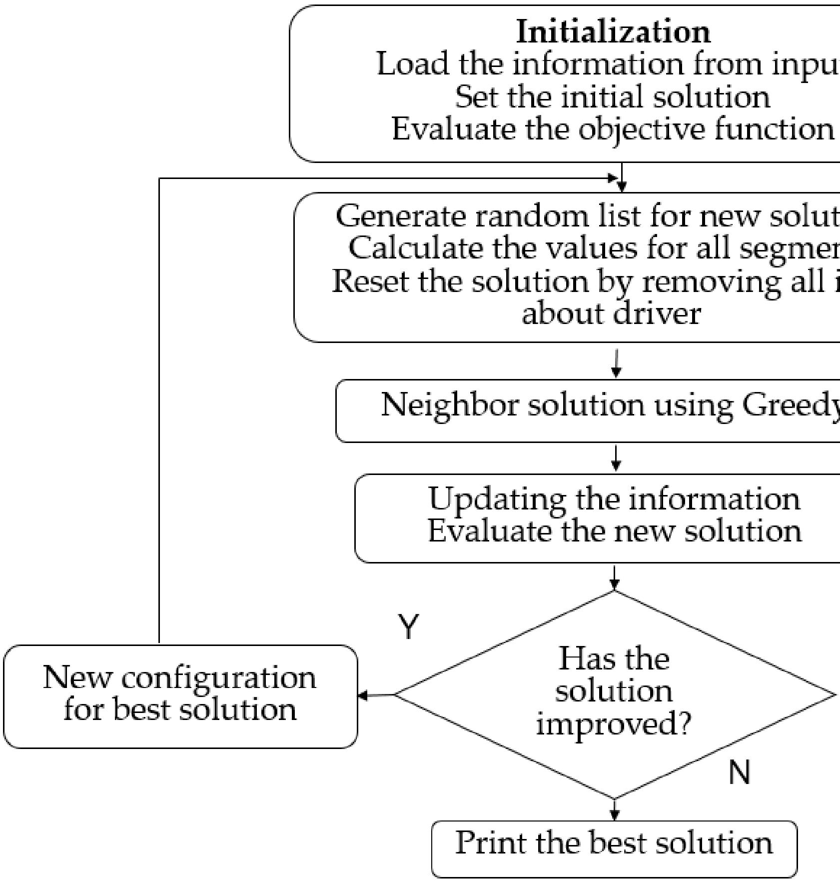

Figure 6.16. Local search flowchart

According to the solution of the model, a short-term scheduling model is developed using taxi driver information. Then it re-evaluates the costs for the two mentioned drivers, as this conversion does not affect the other drivers.

6.2.5.3 Simulated Annealing

The simulated annealing metaheuristic algorithm that was developed to solve the taxi allocation problem is described in the following steps.

Step 1: Start the program.

Step 2: Define the initial state or solution to the problem.

Step 3: Set the initial values for the algorithm's parameters, such as the cooling rate and the number of iterations.

Step 4: Define the temperature schedule, which determines how the temperature changes over time.

Step 5: Generate a neighbor state or solution based on the current one. This is done by making a small change to the current state or solution.

Step 6: Calculate the energy or difference in cost between the current and neighbor solutions.

Step 7: Accept or reject the neighbor solution based on the energy or difference in cost and temperature. The neighbor solution is accepted if the energy is lower than the current energy. If the energy is higher, the neighbor solution is accepted with a certain probability that depends on the temperature.

Step 8: Repeat steps 5-7 until the temperature reaches a minimum value.

Step 9: End the program.

Simulated annealing is a stochastic relaxation method based on an iterative procedure starting from an initial "high temperature" with the system in a known configuration. The simulated annealing **Algorithm 3** is a stochastic optimization algorithm that finds global optima by allowing uphill moves early in the optimization process.

It achieves this by using a temperature parameter to determine the probability of accepting a worse solution. The iterative method of simulated annealing improves the cost function until the current temperature cools down. At high temperatures, atoms can become unstable from initial positions, meaning that the algorithm is allowed to have flexibility in searching potential space. In contrast, at decreasing temperatures, it is more likely to improve with local searches than with initial conditions.

Algorithm 3: Simulated Annealing

```
Function Simulated Annealing (no iteration, low temperature, step, iteration step)
    Current solution = greedy solution
    Initial temperature
    For each step in the no iteration
        Temperature solution = current solution
        Cost delta = (temperature solution (total costs) – current solution (total costs)) / current solution (total costs)
        If cost delta <= 0 or random < exp (-cost delta / temperature)
            If temperature != 0.1 or cost delta <= 0
                Current solution = temperature solution
                Values (current solution (total costs))
        If each step % iteration step == 0 and temperature > low temperature + 1e-10:
            Temperature - = each step
Return Current solution, values
```

Start the solution with the greedy one, then always go to a solution with a good intention. The probability of jumping into a solution with a bad objective depends on temperature, decreasing as the algorithm progresses.

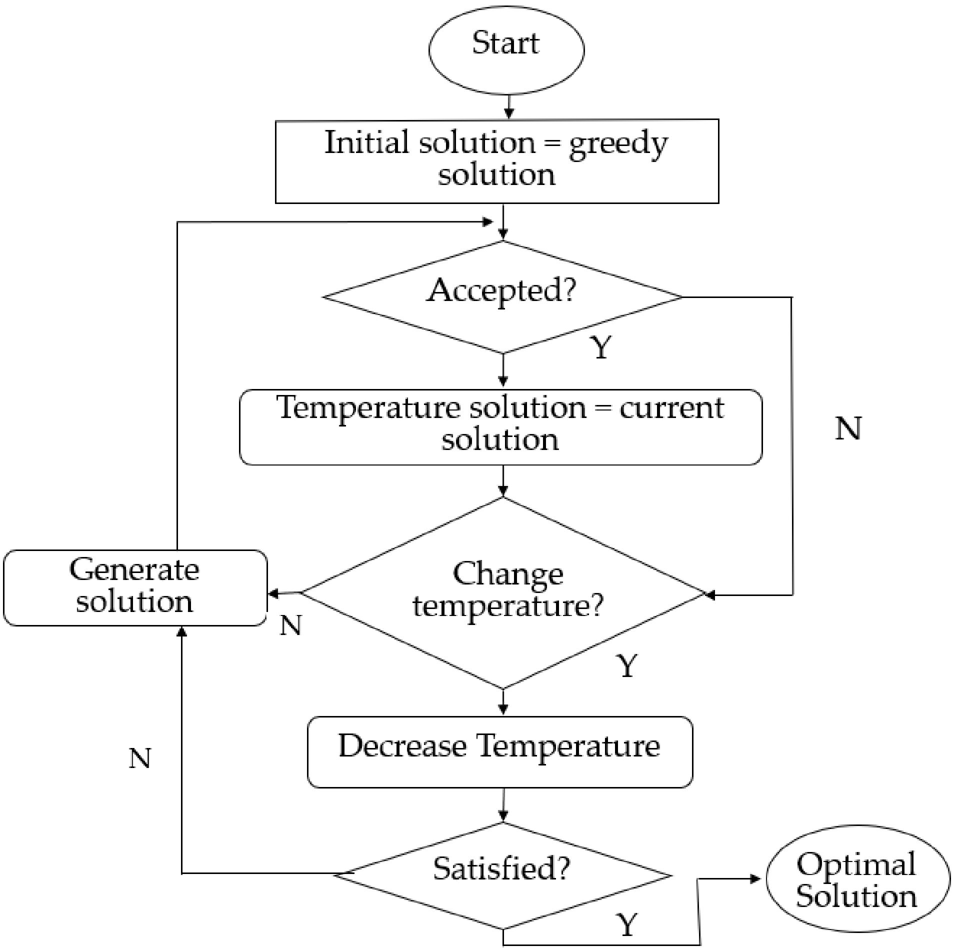

*Figure 6.17***.** Simulated annealing flowchart.

Simulated annealing is a potential method that is proposed to find a worldwide minimum cost function that may contain several minimums Dimitris Bertsimas ; John Tsitsiklis (2007). The objective is to minimize the cost function of the taxi system, while the search for the best solution relates to the cooling process. The Simulated annealing algorithm is presented in the flowchart in Figure 6.17. Simulated annealing is a stochastic relaxation method based on an iterative procedure starting from an initial "high temperature" with the system in a known configuration. The iterative method of SA improves the cost function until the current temperature cools down. At high temperatures, atoms can become unstable from initial positions, meaning that the algorithm is allowed to have flexibility in searching potential space, while at decreasing the temperatures it is more likely to improve with local searches than initial conditions.

6.2.5.4 Dynamic Greedy Programming

In dynamic greedy programming, decisions are made by considering the current problem and the solutions of previously solved subproblems to compute the best solution. This makes it the best solution because it considers all possible cases and chooses the best. For dynamic programming implementation, an additional data structure must be created. Here's a step-by-step explanation of the dynamic greedy algorithm based on the flowchart:

Step 1: Start the program.

Step 2: Define the problem that needs to be solved. This involves specifying the inputs, the objective function to be optimized, and any constraints that need to be satisfied.

Step 3: Divide the problem into smaller subproblems that can be solved independently. This is done using dynamic programming. The subproblems are usually defined in terms of a sequence of decisions or states that need to be made or reached to obtain the optimal solution.

Step 4: Solve each subproblem using the greedy algorithm. The greedy algorithm makes the locally optimal choice at each decision point in the subproblem. This means that it chooses the option that provides the maximum benefit at that particular point in time without considering the long-term consequences of that choice.

Step 5: Combine the solutions to the subproblems to obtain a solution to the original problem. This is done using dynamic programming. The solutions to the subproblems are combined to obtain the optimal solution to the original problem. This is achieved by choosing the sequence of decisions or states that provide the maximum benefit overall.

Step 6: Repeat steps 3-5 until an optimal solution is found. The dynamic greedy algorithm iteratively solves the subproblems and combines the solutions until an optimal solution is obtained. This involves solving the subproblems in a sequence that allows the maximum benefit to be achieved overall.

Step 7: End the program. The dynamic greedy algorithm terminates once the optimal solution has been obtained.

In summary, the dynamic greedy **Algorithm 4** is an iterative process that uses dynamic programming and the greedy algorithm to obtain an optimal solution to a problem that involves a sequence of decisions or states.

Algorithm 4: Dynamic Greedy

```
Function DynamicGreedySolution
    For each client in the total clients
        Minimum cost = np.inf
        Minimum driver = 0
        Time of minimum driver = 0
        For each driver in the total driver
            Distance to client = distance (client, driver)
            Time on arrive to the client = maximum (client desired
            time, distance  to client / speed + driver time)
            Lateness = time on arrive to the client - client desired
            time
            Idle cost = lateness ** 2 + time on arrive to the client +
            client segment
                If drivers == 0:
                    Idle += driver opening cost
                If minimum cost > idle cost
                    Minimum cost = idle cost
                    Minimum driver = driver
                    Time minimum driver = time on arrive to the
                    client
Driver cost[minimum driver] += minimum cost
Drivers on[minimum diver] = 1.0
Driver position[minimum driver] = client position
Driver time [minimum driver] = time min driver + client
segment / speed
Minimum costs of drivers
```

The Dynamic greedy algorithm is presented in the flowchart in Figure 6.18. The greedy solution methodology involves the initial solution. For this purpose, greedy and SA are combined into the dynamic greedy algorithm. The next step is the main solution. The optimization method is applied the dynamic programming and found the minimum total cost.

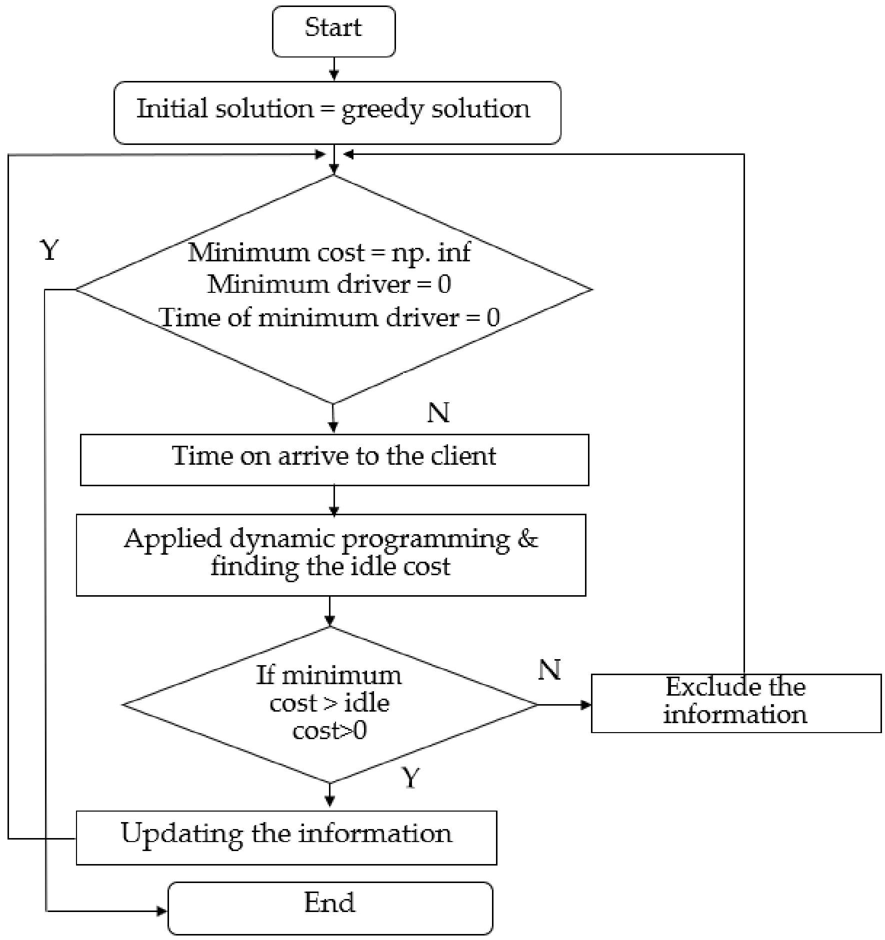

Figure 6.18. Dynamic greedy algorithm flowchart

6.2.6 Explain (Clarify) Each Algorithm

Greedy Algorithm: The Greedy Algorithm (Figure 6.19) is a simple and widely used approach in solving optimization problems. It starts with an empty solution and gradually builds it up by selecting the best possible option at each step. The Greedy Algorithm selects the local optimal solution at each step, without considering the global optimal solution. For taxi data, the Greedy Algorithm can be used to find the shortest path between two points. The algorithm would select the next intersection with the shortest distance to the current intersection until it reaches the destination. This approach is efficient for small datasets, but it may not always produce the most optimal solution.

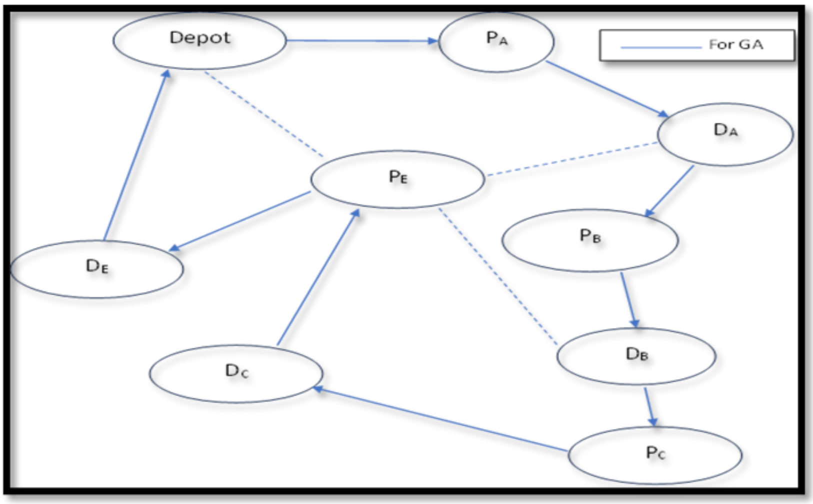

Figure 6.19. Clarify the Greedy Algorithm

Simulated Annealing: Simulated Annealing (Figure 6.20) is a probabilistic approach for finding the global optimal solution for an optimization problem. The algorithm starts with an initial solution and gradually modifies it to improve the solution quality. Simulated Annealing accepts suboptimal solutions to avoid getting stuck in local optima. For taxi data, Simulated Annealing can be used to find the optimal route for a taxi to reach the destination, considering various factors such as traffic, road conditions, and time of day. The algorithm would modify the initial route by swapping intersections and comparing the new route with the current route. The algorithm will accept the new route if it improves the solution, and it would also accept suboptimal solutions with a certain probability.

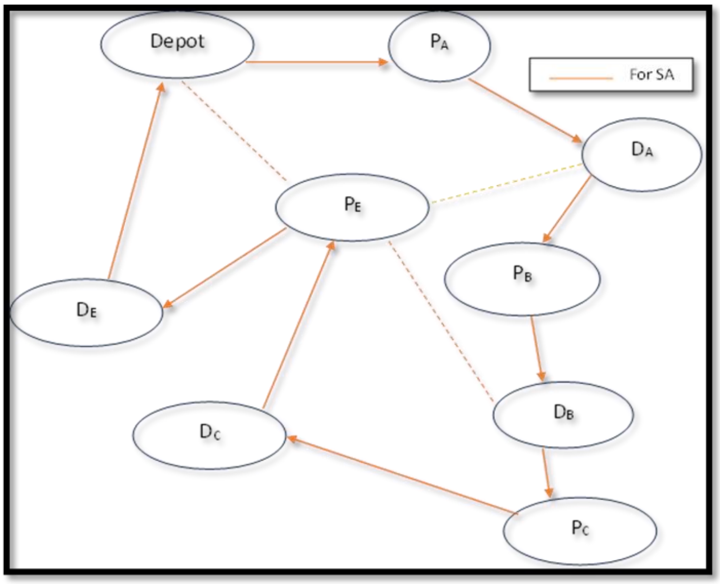

Figure 6.20. Clarify the Simulated Annealing

Dynamic Greedy Algorithm: The Dynamic Greedy Algorithm (Figure 6.21) for taxi data is a variation of the Greedy Algorithm that considers additional factors such as traffic, road conditions, and time of day, but it updates the solution in real time. The algorithm starts with the initial position of the taxi and selects the next intersection that minimizes the travel time to the destination. The algorithm would also consider the traffic and road conditions, and it would update the route dynamically based on the current conditions. For example, if there is a traffic jam on the current route, the algorithm would select an alternate route to avoid congestion. This approach is efficient for real-time applications, as it can quickly find the best route based on the current traffic and road conditions. The dynamic nature of this algorithm makes it more suitable for large datasets and more complex scenarios.

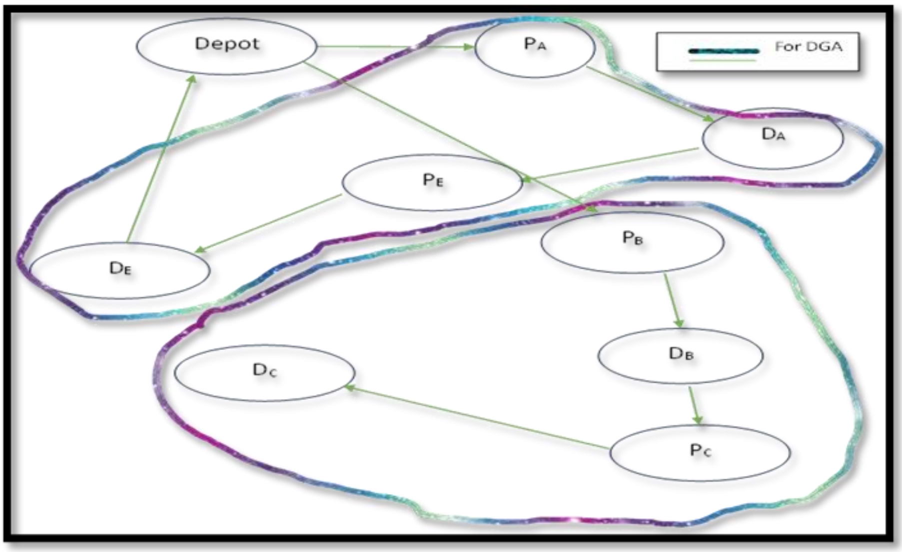

Figure 6.21. Clarify of the Dynamic Greedy Algorithm

Table 6.11. Relation between three algorithms

Greedy Algorithm	Simulated Annealing Algorithm	Dynamic Greedy Algorithm
Makes locally optimal choices at each step.	A probabilistic optimization algorithm inspired by annealing in metallurgy.	Combines the principles of dynamic programming and greedy algorithms.
Does not consider the global optimization.	Uses a randomization process to escape local optima and explore the solution space.	Utilizes memorization to store and reuse solutions to subproblems.
May lead to suboptimal solutions.	Starts with an initial solution and iteratively improves it by accepting or rejecting neighboring solutions based on a probability criterion.	Considers previously computed solutions to make greedy choices at each step.

Typically has a complexity compared to dynamic programming and simulated annealing.	Explores less optimal solutions to avoid getting stuck in local optima.	Achieves better optimization compared to a pure greedy approach.
Best suited for problems where a locally optimal choice leads to an acceptable overall solution.	Can converge to the global optimal solution under certain conditions.	Computes an optimal solution by considering the global optimization.
Does not require a cooling schedule or probability criterion.	Requires a cooling schedule and a mechanism to control the acceptance probability of worse solutions.	Requires iterative calculations and storage of intermediate results.
Example: Selecting the shortest available taxi at each pickup point without considering the overall optimization.	Example: Optimizing the taxi problem by iteratively perturbing the current solution and accepting or rejecting it based on the simulated annealing probability criterion.	Example: Calculating the optimal routes for taxis by considering the shortest distances between pickup and drop-off points.

6.2.7 Computational Experiments

This section details the experiments that have been carried out to solve the given model utilizing (Section 5.3) heuristics algorithms (greedy, simulated annealing, and dynamic greedy programming algorithms). The original taxi business data was the address from the Nagaoka taxi allocation issue in 2019, with average demand. The data has been started from 00:00 to 23:00 in a week in 2019. The actual operation has analyzed the GPS data of all vehicles and 42 taxis belonging to Mitsukoshi Taxi Co., Ltd., the largest taxi company with a business area in Nagaoka city. We have used the number of rides by the hour and the number of waiting for units by time zone in the taxi operation data.

The beginning and finishing temperatures for the simulated annealing were set as 0.21. The decay rate was 0.001. These parameters were maintained for all problems and independent runs. Each problem was solved 10 times by using an iteration step of 500, with the initial step of 0.1 and a low temperature of 0.1. The lengthy simulated annealing schedule may take a while to complete. Further, simulated annealing has a lot of adjustable parameters. Therefore, simulated annealing takes comparatively longer CPU time for simulation. On the other hand, the greedy algorithm also consumes comparatively longer time since it breaks the issue into its components and re-members the answer for each element to apply it later when a similar component occurs again. Compared to other algorithms, dynamic greedy algorithms have lower temporal complexity and identify the most practical solution to arrive at the best solution at every stage. Therefore, dynamic greedy algorithms consume comparatively less time for the simulation process. The heuristics were written in the Python programming language. The tests were carried out using a laptop with an Intel(R) Core (TM) i7-5500U CPU running at 2.40GHz, with 16 GB of RAM, and using the Windows 10 operating system.

6.2.7.1 Results and Discussions

In this section, we present several optimal solutions for mathematical modeling. We present the results of the experiments of the three algorithms in solving the random taxi demand problems. The total number of taxis (42) means the number of taxis in the company. The number of allocated taxis (20) is the number of taxis per day that the company used. Moreover, taxi demand (approximately 4) was the optimized number of taxis that was sufficient for a day according to the model.

The dataset has three instances, 01.01.19, 01.03.19, and 15.03.19, representing taxi problems in Nagaoka City. These instances contain between 333 and 658 passenger demand. Additionally, there are problems where only a subset of the taxi fleet can serve the passenger requests. To make it easier for the heuristic to find solutions, we set the number of available vehicles to one more than required. According to the model, taxi demand was the optimized number of taxis (approximately 8) that was sufficient for a day. Table 6.12 presents the average passenger demand, the number of taxis, the total cost, and the CPU time calculated for each algorithm. The outcomes are based on randomly generated taxi demand situations. The greedy, simulated annealing, and dynamic greedy heuristics showed average total costs of JPY 21,433.93, JPY 23,730.75, and JPY 20,573.98, respectively, in the optimal solution. The dynamic greedy heuristic required an average CPU time of 1.19 seconds to find the optimal solutions. The findings show it was able to identify better solutions in a shorter period on the CPU.

Table 6.12**.** Results obtained from the heuristics algorithms**.**

Instance	**Passenger Demand**	**Greedy algorithm**			**Simulated annealing**			**Dynamic greedy**		
		Taxi	**Total cost (yen)**	**CPU time (sec.)**	**Taxi**	**Total cost (yen)**	**CPU time (sec.)**	**Taxi**	**Total cost (yen)**	**CPU time (sec.)**
01.01.19	333	5	10,548.54	47.2	5	13,883.37	55.66	5	9,175.24	0.85
01.03.19	537	9	25,461.13	37.25	9	26,303.6	68.67	8	24,602.08	1.32
15.03.19	658	10	28,292.11	59.77	10	31,005.28	119.11	10	27,944.63	1.42
Average	**509**	**8**	**21433.93**	**48.07**	**8**	**23,730.75**	**81.15**	**8**	**20,573.98**	**1.19**

(1 yen = 0.0072 dollar)

The Passenger Demand/per day for the suburban area of Nagaoka city is, on average, 509, with the average number of taxis required as 8. The optimization results of the taxi allocation problem with different demands are summarized in Table 6.12. As the demand increased, the total idle time costs increased from JPY 10,548.54 with five taxis to JPY 28,292.11 with ten taxis in the greedy algorithm, and in the simulated annealing algorithm from JPY 13,883.37 to JPY 31,005.28. However, the proposed algorithm (dynamic greedy algorithm) increased the cost from JPY 9,175.24 to JPY 27,944.63.

In addition, a dataset with information on taxi trips, including pick-up and drop-off locations, was examined. This investigated the pickup and drop-off coordinates from the dataset and used them to compile the taxi trips for each location. We combined the trips based on these coordinates to determine how frequently taxis picked up and dropped off passengers at each distinct place. Figure 6.22 represents the regular taxi trips of Nagaoka City. We consider the average of 387 taxi trips per day.

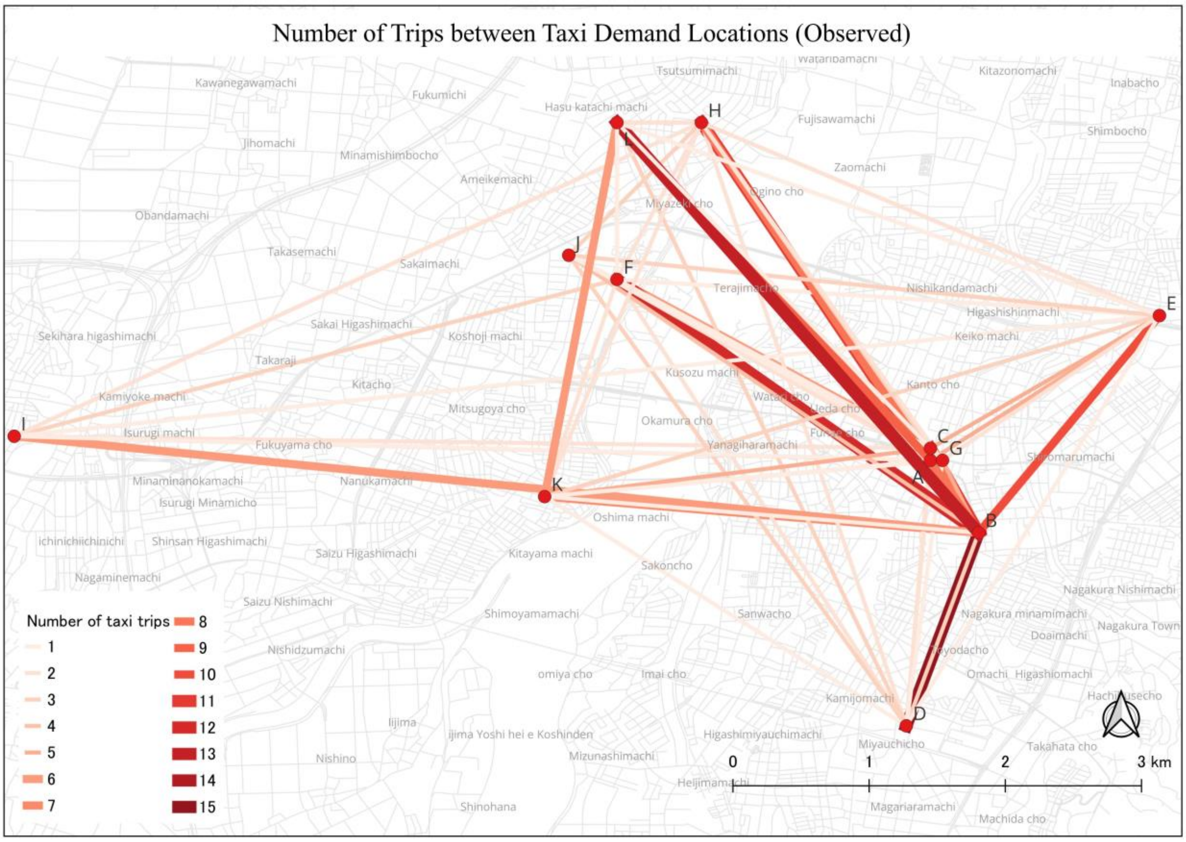

Figure 6.22. Number of taxi trips to Nagaoka City (Observed)

The number of journeys was reduced, and it was determined which trips showed notable drops. We created a complete dataset that reflects taxi journeys across several places by integrating and preprocessing these various datasets. The proposed model was obtained to optimize taxi driver idle time, taxi routes, and allocations while taking into account parameters including trip distances, traffic conditions, and passenger demand (Figure 6.23). We were able to come up with a solution that reduced the overall number of trips while preserving effective transportation services by redefining the issue as an optimization challenge. The fact that the number of trips fell by a statistically significant amount shows that the model was successful in optimizing the taxi service and maximizing resource usage. This result implies that the model was successful in identifying chances for trip consolidation, whereby a number of passengers with comparable itineraries were put together, resulting in fewer individual journeys.

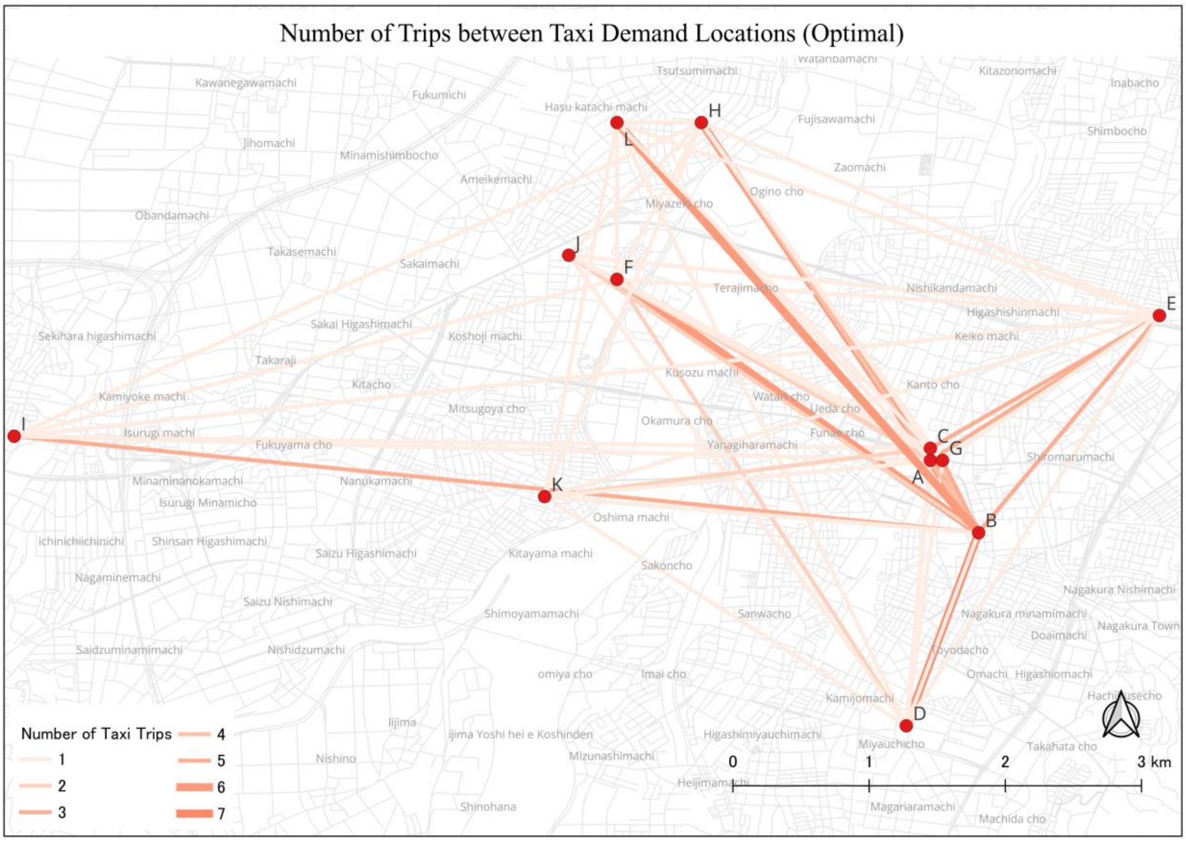

Figure 6.23. Number of taxi trips to Nagaoka City (Optimal)

The results of this study have implications that the reduction in taxi trips indicates the potential for improved efficiency and sustainability in suburban transportation systems. For general taxi number of trips 387 per day in Nagaoka City (Figure 6.22). One the other hand, the proposed model reduced the trip number to 174 (Figure *6.23*). By optimizing route planning and allocation of resources, we have minimized the number of unnecessary trips, leading to reduced fuel consumption, and CO2 emissions.

Calculate the total CO2 emissions:

The emission factor for taxis can vary depending on the type of vehicle, the fuel type, and the average distance traveled. According to the U.S. Environmental Protection Agency (EPA), the average CO2 emissions for a gasoline-powered taxi is approximately 0.404 kg/mile Environmental Protection Agency (2005). Using the carbon footprint model Rao, H. S., Hettige, H., Singru, N., Lumain, R., & Roldan (2010) provides a useful starting point for understanding the environmental impact of the taxi fleet and identifying ways to reduce emissions. The carbon footprint model is as follows:

$$CO2\ Emissions = N_{at} \times A_{Ce} \times A_{tt} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (6.1)$$

Where N_{at}: the number of allocated taxis per day,

A_{Ce}: the average CO2 emissions per taxi,

A_{tt}: the taxi travels on average per day.

Minimizing the number of taxis required to serve all passenger demand is typically the top priority in this problem. The current heuristic approach addresses this objective by proposing a simple algorithm that minimizes the number of vehicles needed. It's worth noting that this algorithm is only suitable for problems involving a homogeneous fleet of taxis. We also assume that the number of taxis available is 42, so constructing an initial feasible solution can always be done. This paper focused on minimizing the CO2 emissions from taxi idle time, and we reduced the number of taxis that idle. This can be done by implementing policies or practices that encourage drivers to turn off their engines when not driving, such as turning off the engine during short breaks or when waiting for passengers.

The number of allocated taxis per day is 42 taxis that the company used. and the average CO2 emissions per taxi are 0.404 kg/mile. Each taxi travels an average of 64.10 miles per day, then from the equation (6.1), the total daily CO2 emissions from the taxi would be:

$$CO2\ emissions\ =\ 42\ taxis\ \times\ 0.404\ kg\ CO2/mile\ \times\ 64.10\ miles\ per\ taxi$$

= 1,087.65 kg CO2 per day.

After the optimal solution of the mathematical model in the case study, we minimize the number of taxis and costs. On average, the number of taxis per day used is only 8 in Table *6.12*. So, after solving the model, the total daily CO2 emission is:

$CO2\ emissions\ =\ 8\ taxis\ \times\ 0.404\ kg\ CO2/mile\ \times 64.10\ miles\ per\ taxi\ =$ 197.48 kg CO2 per day.

The taxis in the suburban area of Nagaoka have reduced their CO2 emissions by 81.84%, which amounts to 890.17 kg per day.

We present a novel model for optimizing the operations of a taxi firm using a simulation-based approach to evaluate the performance of our model. It shows that our model can effectively capture the real-world situation of the taxi market. By using the proposed model, the best number of taxi hours can be made available to customers while retaining driver benefits, which means reducing customer waiting time and making it one of the best solutions. The effects on taxi customers must be taken into account when researching reducing taxi operating costs. Table 4 shows the optimal allocation of taxi demand. This confirms that the model can minimize the total idle time costs of taxi drivers while minimizing CO2 emission. Dynamic greedy heuristics in a relatively short runtime successfully solved the taxi allocation problem, making this an effective tool for taxi companies.

Taxi companies should know that establishing a decision support system will benefit fleet management and empty taxi repositioning. The results of our case study on optimizing taxi allocation and minimizing CO2 emissions in a suburban area were very encouraging. By collecting and analyzing data on taxi usage patterns, it is able to identify areas of high demand and optimize taxi allocation to reduce idle time, which in turn leads to a reduction in CO2 emissions. Optimization algorithms to identify the most efficient routes for taxis also significantly reduced the distance traveled by taxis and the amount of CO2 emissions produced. This was achieved by minimizing the number of empty trips and optimizing routes to avoid traffic congestion. The existing studies (Zhang *et al.* (2020); Zhang and Nian (2013)) described their model could reduce CO_2 emissions from 36% to 47.62% in urban areas, whereas this proposed model could reduce CO_2 emissions up to 81.84 % in suburban areas. Therefore, this model can make a sufficient contribution to the reduction of CO_2 emissions.

Taxi drivers and passengers saved time and money, while the suburban area became more sustainable and environmentally friendly. The success of this case study highlights the potential for similar programs to be implemented in other regions and cities. Using data analysis, optimization algorithms, and incentives can create a more sustainable and efficient transportation system that benefits everyone involved. It also highlights the importance of taking action to reduce CO2 emissions and mitigate the impact of climate change.

6.2.7.2 Sensitivity Analysis

This section discusses the sensitivity analysis based on taxi demand. Such analysis is vital in assessing the validity of the objective function. Sensitivity analysis is employed to identify relationships between input parameters and model outputs. For example, one can select initial concentrations of modeled species or reaction rates as input parameters. Sensitivity analysis is very helpful in mathematical modeling be-because it explains how different model components are interdependent. The developed model in this study was also subjected to a sensitivity analysis with regard to passenger demand. When performing a sensitivity analysis of expenses, the taxi demand was altered (increased or decreased) by 10% and 50% while only altering one parameter at a time, leaving the other parameters the same. The formula for calculating the percentage inaccuracy was (measured value-actual value) *100/actual value. Here the actual value is the average ideal time cost from the model, and the measured value is obtained after alternating the demand.

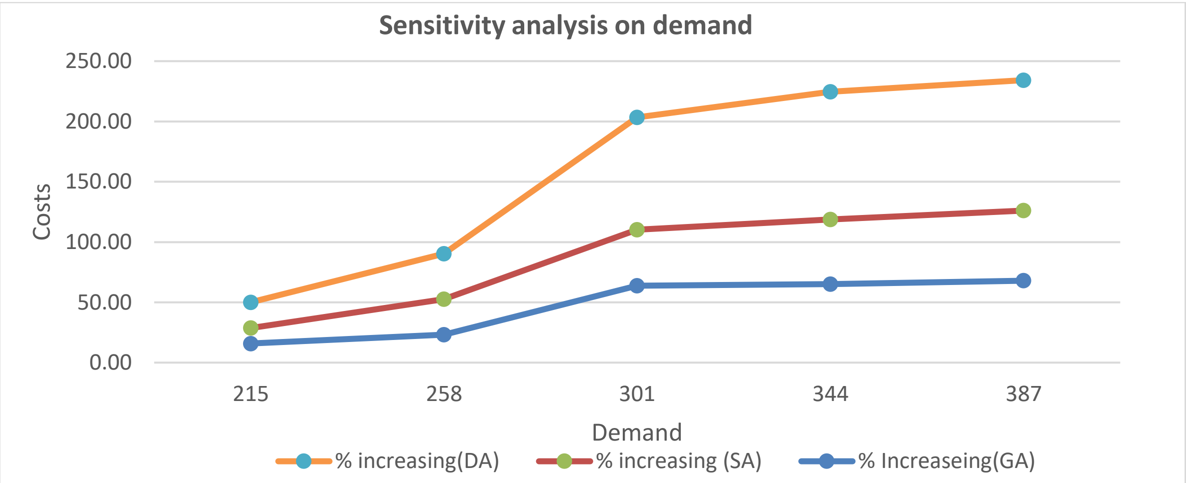

Figure 6.24. Sensitivity analysis results for the model based on the demand.

The sensitivity analysis was carried out by increasing and decreasing the passenger demands by the daily passenger volume. From the sensitivity, it was clear that the total idle time cost increased if the demand increased, as depicted in Figure 6.24. The operating costs (yen) increased significantly when taxi demand increased from 215 to 387 trips per day.

6.2.8 Summary of the Taxi Allocations Problem

Efficient transportation systems are necessary to reduce CO2 emissions in the context of climate change. Taxi services are an essential part of the transportation system in urban and rural areas. Inefficient taxi services cause problems such as increased idle times, resulting in increased CO2 emissions. Optimizing taxi allocation can help reduce CO2 emissions and improve the sustainability of urban transportation. This study proposed a mathematical model of taxi allocation optimization problem for minimizing taxi driver idle time costs which help to reduce CO2 emission in suburban contexts. For solving the proposed model, we proposed three heuristic algorithms: greedy algorithm, Simulated annealing algorithm, and dynamic greedy. Further, a case study has been carried out to demonstrate the proposed methods and show its implication by using real taxi data from Nagaoka, Japan. For the case study, we first identified the taxi travel hotspots as potential locations for taxi spots by investigating the pickup and drop-off locations and times by analyzing the GPS data of the taxi. After that, we found the optimal taxi allocations by applying the proposed model and solution algorithms. To summarize, dynamic greedy heuristics successfully obtained excellent solutions in relatively short runtimes for the taxi issue, making strategic decisions and feasible choices for taxi markets to adopt. The case study application in this study demonstrates the potential for using data analysis, optimization of taxi allocations and the number of taxis, and reducing approximately 81.84% of CO2 emissions in the transportation sector. Finally, sensitivity analysis was applied to validate the model for passenger demand. The sensitivity analysis showed that the total idle time cost increased if the demand increased.

To conclude, the optimization of taxi allocations can have a significant impact on creating carbon-free and environmentally friendly cities. By strategically placing taxi spots and optimizing their routes, cities can reduce the number of vehicles on the road, decreasing harmful emissions and improving air quality. Therefore, it is essential for cities to invest in and prioritize taxi allocation optimization as a crucial step toward achieving sustainable transportation. Future research able to be conducted utilizing ML algorithms to solve the allocation issue may provide promising results. This model can also be adapted to improve the performance of taxis in various cities and other urban public transport networks.

Chapter 7. COST-EFFICIENCY ANALYSIS ON SHARED TAXI

Shared taxi operations, also known as shared-ride or demand-responsive transit, have gained popularity in recent years due to their potential for cost savings and improved efficiency. Shared taxi operations involve multiple passengers sharing a vehicle for a trip that may have different origins and destinations. In this chapter, we will discuss the cost-efficiency analysis of shared taxi operations, highlighting their potential cost savings and improved efficiency.

7.1 Introduction

Japan has a well-developed transportation system, with a range of public and private options for individuals to get from place to place (Yudhistira *et al.* (2015)). In recent years, the use of shared taxis (paratransit) has emerged as a popular alternative to general taxi services, offering a more flexible and cost-effective solution for those who need to get around (Takeuchi *et al.* (2022)). In order to understand the cost-efficiency of shared taxi operations compared to general taxi operations in Japan, which can provide valuable insights into the movement of vehicles and drivers and help to identify strategies for improving the cost-efficiency of these services.

The goal of this cost-efficiency analysis is to compare the operational performance of shared taxis and general taxis in suburban areas, in Japan, with a focus on factors such as vehicle utilization, driver expenses, and operational costs. By analyzing Global Positioning System (GPS) data, this study can identify how different operational strategies, such as route optimization and vehicle pooling, impact the cost-efficiency of shared taxi operations compared to general taxi operations. This information can be used to develop best practices for improving the cost-efficiency of shared taxi operations in Japan and to better meet the transportation needs of individuals and communities. The shared taxi model has gained popularity in some areas due to its cost-effectiveness and convenience, particularly for short trips.

In contrast, general taxi services in Japan have been in operation for many decades and are a well-established mode of transportation. So, the cost of general taxi services can be a concern, particularly for longer trips or during peak hours when the cost per ride can be high. This has led to increased interest in shared taxi services as a potential solution to the cost inefficiencies of general taxi services. The shared taxi, also known as ridesharing, is an innovative form of transportation that has gained significant attention in recent years. Shared taxi services have been found to provide a flexible and convenient alternative to traditional public transportation, particularly for people living in suburban or rural areas where public transportation options are limited.

In the article, Zhai *et al.* (2019) proposed a bottom-up approach to measuring transportation network efficiency, with a specific focus on taxi efficiency in New York City. The approach involved using GPS data from taxis to identify key network segments and calculated performance metrics such as travel time, delay, and accessibility. The authors argued that this approach provided a more comprehensive and accurate assessment of network efficiency compared to traditional top-down approaches that rely on aggregate data. The study used real-world data from over 100,000 taxi trips in New York City to demonstrate the effectiveness of the proposed approach. The results showed that the approach can identify areas of the network that are experiencing high levels of congestion and delay and can be used to inform policy decisions aimed at improving network efficiency. The authors highlight the potential of bottom-up approaches to transportation network analysis and suggest that this approach could be applied to other modes of transportation as well. Inturri *et al.* (2021) compared the performance of general taxi systems with demand-responsive shared transport systems using agent-based simulation. The study aimed to investigate the potential of demand-responsive shared transport systems to reduce travel costs, improve accessibility, and enhance environmental sustainability. The study used a simulation model based on agent-based modeling (ABM) to compare the performance of general taxi systems with demand-responsive shared transport systems in terms of efficiency, cost-effectiveness, and environmental impact. The model considered various factors such as travel demand, vehicle availability, passenger preferences, and trip costs. The study found that demand-responsive shared transport systems have the potential to significantly reduce travel costs and improve accessibility, particularly in low-demand areas. The simulation results also showed that demand-responsive shared transport systems have the potential to significantly reduce carbon emissions compared to general taxi systems. Finally, the study suggested that demand-responsive shared transport systems can provide a more efficient, cost-effective, and environmentally sustainable alternative to general taxi systems, particularly in low-demand areas.

Wang *et al.* (2022) examined the relationship between urban road network centrality and taxi travel in Shenzhen, China. The authors used data on road network centrality at multiple scales and taxi GPS data to model the impact of network structure on taxi travel times and travel distances. They found that higher network centrality at the city and district levels is associated with shorter taxi travel times and distances, while higher centrality at the street level is associated with longer travel times and distances. The study provided insights into the complex relationship between urban road networks and travel behavior and suggested the importance of considering network structure when designing transportation policies and systems. Whitmore *et al.* (2022) explored the potential for integrating shared autonomous mobility (SAM) services with existing public transportation systems to provide more equitable transit coverage. The study used a cost-efficiency analysis to compare different integration scenarios and assess their impact on accessibility, travel time, and cost. The authors argued that integrating SAM services with public transportation can provide significant benefits in terms of increased transit coverage and reduced travel time and cost, particularly for low-income and underserved communities. They also suggested that this approach can be more cost-effective than expanding traditional public transportation services alone. Elting and Ehmke (2021) investigated the feasibility and potential benefits of introducing shared taxi services in rural areas in China. The authors conducted a survey to understand the travel behavior and preferences of local residents and used simulation models to analyze the potential impacts of shared taxi services on travel time, cost, and accessibility. The results showed that shared taxi services could significantly improve travel efficiency and reduce travel costs for local residents, while also providing additional income opportunities for taxi drivers. However, the success of shared taxi services in rural areas depends on various factors such as population density, road infrastructure, and regulatory frameworks. Čulík *et al.* (2022) analyzed the impact of technological changes and taxi market regulations on the taxi vehicle fleets in Slovakia. The study used data collected from taxi companies and regulators in Slovakia to analyze changes in the taxi industry over time, including changes in the number and age of vehicles, as well as changes in the types of vehicles used by taxi companies. The authors also analyzed the impact of regulations on the taxi industry, including regulations related to vehicle emissions and safety standards. Therefore, the study provided insights into how technological changes and regulations can impact the taxi industry, and how taxi companies can adapt to these changes to remain competitive. Peng *et al.* (2022) proposed a new algorithm for matching multiple passengers to a single taxi in a ride-sharing service. The algorithm aimed to balance the interests of the passengers, the taxi driver, and the ride-sharing service provider while minimizing the empty vehicle mileage and reducing congestion. The proposed algorithm was based on the stable matching theory and incorporates the concept of a price-based mechanism to incentivize the players to reveal their preferences truthfully. The authors evaluated the performance of the proposed algorithm through simulation experiments and demonstrate its effectiveness in achieving stable matching outcomes that satisfy the players' preferences and improve the operational efficiency of the taxi-sharing service. Qiu *et al.* (2022) explored various approaches to matching, routing, and pricing methods for taxi-sharing practices. The authors proposed a framework for matching and routing shared taxi services based on a shared taxi network model. They also examined different pricing methods for shared taxi services, including distance-based pricing, time-based pricing, and hybrid pricing. The article provided a comprehensive analysis of the key factors affecting the efficiency of taxi-sharing practices, including the network topology, passenger demand patterns, and pricing strategies. Several authors conducted studies on the cost of on-demand taxi services (Čulík *et al.* (2020); Mingolla and Lu (2021); Qu *et al.* (2014); Whitmore *et al.* (2022)).

Therefore, the literature review suggested that shared taxi services offer a number of benefits, but also face a range of challenges and limitations. Further research is needed to fully understand the potential benefits and limitations of shared taxi systems and to identify strategies for maximizing the positive impact of these services on communities and the environment. The growth of shared taxi services has also been accompanied by a number of challenges and concerns, including concerns about the safety of passengers, the impact on employment in the taxi industry, and the potential for increased congestion in densely populated areas. In terms of the cost-efficiency of shared taxi services, several studies have found that shared taxi systems can offer significant savings over general taxi services, due to the reduced costs of operating a single vehicle for multiple passengers. So, the cost-effectiveness of shared taxi services also depends on various factors such as the operating costs of the vehicle, the number of passengers sharing the ride, and the efficiency of the dispatch system.

7.2 Contributions

In general, a taxi's fare per passenger decreases as more people share it, and the driver makes more money as a result. Taxi sharing is a beneficial form of transportation that should receive greater attention in the future. Shared taxis can provide transportation services to areas where traditional public transit is limited or non-existent, increasing access to employment, healthcare, education, and other important destinations for individuals who may not have access to personal vehicles. Shared taxis can provide a more flexible and convenient mode of transportation, particularly for individuals who may have difficulty using traditional public transit, such as older adults, people with disabilities, and low-income individuals. By providing a complementary service to traditional public transit, shared taxis can help to increase the overall efficiency and effectiveness of the public transit system, reducing the need for larger and more expensive transit vehicles, and helping to meet the transportation needs of a wider range of individuals. So, by providing a more flexible, convenient, and cost-effective mode of transportation, shared taxis can play an important role in meeting the transportation needs of individuals and communities. This study aims to characterize the economic feasibility of public transportation for a shared taxi. Shared taxi services have the potential to improve the economic feasibility of taxi services in Japan by increasing ridership and reducing operational costs. By allowing multiple passengers to share a ride, shared taxis can reduce the number of taxis on the road and increase the utilization rate of each taxi, leading to reduced operational costs for taxi companies. The main contributions of this chapter are as follows:

Firstly, we propose a mathematical model of the shared taxi routing problem that considers the different time windows for passenger service, passenger behaviors, the capacity of the taxis, the locations of the pick-up and drop-off points, and other relevant variables. This model provides a framework for finding optimal solutions to the shared taxi routing problem.

Secondly, we developed a multiboot algorithm that combines heuristic and metaheuristic techniques to efficiently solve the shared taxi routing problem. The algorithm employs various optimization techniques, including local search, simulated annealing, and genetic algorithms, to identify near-optimal solutions in a short amount of time.

Thirdly, we compare the cost efficiency of shared and general taxi services by analyzing real-world data on travel time, distance, and cost, as well as environmental impacts such as emissions and congestion. Our results show that shared taxi services can potentially achieve significant cost savings compared to general taxi services, especially in areas with high demand for taxi services.

Therefore, the economic feasibility of shared taxis in Japan may depend on factors such as regulatory policies, market demand, and the availability of technology to support shared taxi services. Additionally, the cost-effectiveness of shared taxis may vary depending on factors such as the density of urban areas and the availability of alternative transportation options. Finally, making sure riders and drivers are properly matched will also help to maximize efficiency.

7.3 Analysis of Taxi Operations

In Japan, shared taxis are a form of public transportation where passengers share a taxi with other passengers who are traveling in the same direction. General taxis, on the other hand, are private taxis that are hired by a single passenger or group of passengers for a specific trip. To analyze the differences and similarities between shared taxis and general taxis in Nagaoka City, we would need to gather data on various factors, such as time of day, day of the week, weather conditions, and special events.

The overall number of trips taken from others (such as a home) is 220, but the total number of journeys taken to others (such as a home) is 331, resulting in an asymmetrical distribution. The outward route is frequently sent to a family member, and the return route is typically a cab, it is presumed. Moreover, there are 116 total trips coming into Nagaoka Station and 37 total journeys leaving, which is an unbalanced ratio. Yet, because the final bus leaves early, the outward route is referred to as a bus and the return route as a taxi. There are more trips on the return trip from the city center to others (home, etc.) than on the outbound trip, but it is assumed that taxis are used to return home after the last bus.

7.3.1 Analysis of Shared Taxi Operation

Shared taxi operation in the Niigata area, Japan, involves the analysis of the taxi data to understand the benefits and challenges of the service. Niigata area has been implementing a shared taxi service since 2016 to provide more affordable and convenient transportation options for residents and visitors (MINAMI *et al.* (2016); Sano *et al.* (2007); MONDAL *et al.* (2021)).

The analysis of shared taxi operations in Nagaoka City involves several factors, including the number of users, the cost savings, and the impact on the overall transportation system. Some of the key factors that need to be analyzed include:

a) The number of users: One of the key factors in analyzing shared taxi operations in Nagaoka City is the number of users. This information can be obtained by analyzing the number of reservations made and the number of passengers who used the service. By analyzing this data, we can understand the level of demand for the service and identify areas where the service may need to be expanded.

b) Cost savings: Another important factor to consider when analyzing shared taxi operations in Nagaoka City is cost savings. Since passengers are sharing a taxi with others, the cost of transportation is divided among them, making it more affordable than traveling alone.

c) Impact on the overall transportation system: The shared taxi service in Nagaoka City has the potential to impact the overall transportation system. For example, if more people use the shared taxi service, it could lead to a reduction in the number of cars on the road, reducing traffic congestion and improving air quality.

d) Challenges: Despite the benefits of shared taxi operations in Nagaoka City, several challenges need to be addressed. For example, the service may not be suitable for people who need to travel at specific times or have tight schedules. Additionally, the service may not be as reliable as other forms of transportation, especially during bad weather or traffic congestion.

Therefore, the analysis of shared taxi operations in Nagaoka City taxi data suggests that the service has the potential to provide more convenient and affordable transportation options for residents and visitors. There are several challenges that need to be addressed to make the service more accessible and reliable for passengers. By addressing these challenges, shared taxi services can play an important role in improving transportation options in suburban areas.

7.3.2 Analysis of General Taxi Operation

In Japan, general taxi operations are a vital part of the transportation industry. According to a report by the Japan Federation of Hire-Taxi Associations, the number of employees 283,193 in Japan has been steadily increasing in 2021(Kawanabe (2022)). The report also indicates that the majority of taxi drivers in Japan are self-employed and that the average working hours for taxi drivers are relatively long, ranging from 9 to 11 hours per day. One of the main issues facing general taxi operations in Japan is the aging population of drivers.

According to the same report by MLIT, the average age of taxi drivers in Japan is 56.7 years old, and the number of younger drivers is decreasing. This trend could lead to a shortage of drivers in the future, and it may become increasingly challenging to maintain high levels of service and reliability. Despite these challenges, general taxi operations in Japan continue to be an essential part of the transportation industry. With the increasing popularity of ride-hailing and ride-sharing services, general taxi companies are also exploring new business models and incorporating technology to stay competitive. For example, some companies have introduced mobile apps for booking rides, and others are experimenting with electric or hybrid vehicles to reduce their carbon footprint.

Therefore, general taxi operations in Japan have a strong foundation and continue to be a critical mode of transportation for many people. So, the industry will need to adapt and innovate to remain competitive and relevant in a rapidly changing transportation landscape. The cost structure of taxi transportation is a basic element in understanding the taxi operation cost. For the cost structure of taxi transportation, we use the value of the Hire Taxi Annual Report 2018 (Association (2018)). This cost is proportional to the number of taxis (personnel cost, vehicle depreciation cost, insurance fee: 76.4%) and proportional to mileage (fuel oil cost, vehicle repair cost, half of the other expenses: 16.6%).

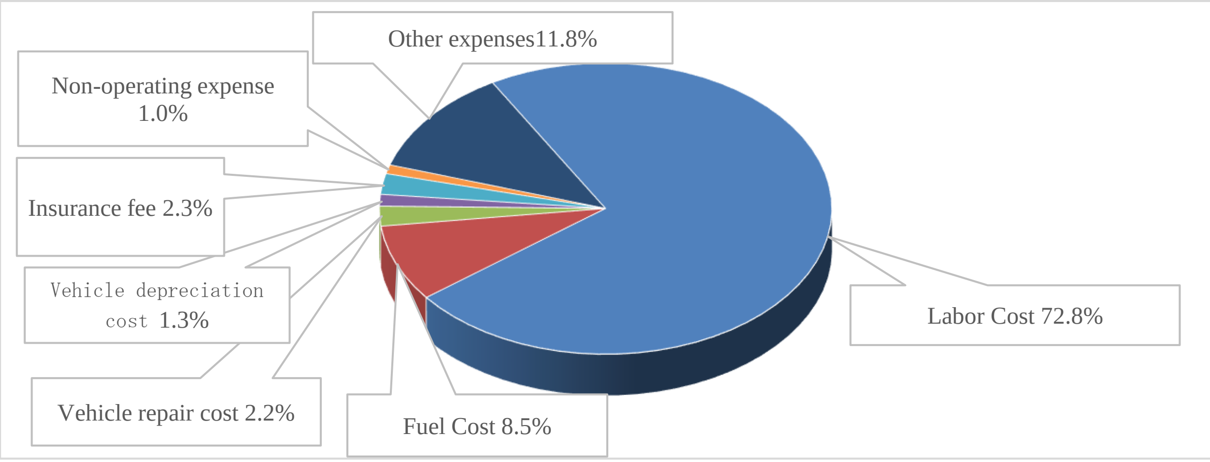

Figure 7.1. Taxi transportation cost structure

We calculated the reduction rate of the cost if all the rides were booked by the day before, with different time windows divided into three, which are not proportional to the number of vehicles and running (non-operating expenses, half of the other expenses: 6.9%). Figure 7.1 shows calculations based on the results of hearings that other expenses include car lease fees and head office expenses.

7.4 Materials

The shared taxi routing problem with different time windows is a transportation problem that involves finding an optimal routing plan for a fleet of shared taxis, taking into account the different time windows for pick-ups and drop-offs of passengers.

The problem involves determining the best routes for the shared taxis to pick up and drop off passengers at different locations within specific time windows while minimizing travel time and distance and maximizing the number of passengers served. This problem is particularly challenging because it requires balancing conflicting objectives, such as minimizing the costs, and number of taxis used while ensuring all passengers are served within their specified time windows. The problem can be further complicated by factors such as traffic congestion, varying demand for taxis at different times and locations, and the need to satisfy various operational constraints, such as vehicle capacity and driver availability. There are several ways to improve the efficiency of shared taxi operations. One would be to make sure that riders are properly matched with drivers so that they don't have to spend extra time going out of their way to pick up passengers. The metaheuristics algorithms can help to find efficient routing plans that meet the needs of both the taxi company and its passengers while taking into account the complex constraints and variables involved in the problem.

7.4.1 Assumptions

The following are some of the common assumptions made in the analysis of shared and general taxi operations:

- ✓ **Cost structure:** The cost structures of shared and general taxi operations are similar, including costs for vehicle maintenance, insurance, and driver salaries.
- ✓ **Demand:** The demand for shared and general taxi services is relatively stable and can be predicted with a certain degree of accuracy.
- ✓ **Utilization rate:** The utilization rate of shared taxi vehicles is higher compared to general taxis, as shared taxis are used by multiple passengers per trip, whereas general taxis are used by one passenger at a time.
- ✓ **Technology:** Both shared and general taxi operations have access to similar technology and data systems, such as GPS tracking, mobile apps, and real-time data analytics.
- ✓ **Economic conditions:** The overall economic conditions and market trends are similar for both shared and general taxi operations.

It is important to note that these assumptions may not hold in all cases, and the actual situation of shared and general taxi operations may vary based on specific market conditions and local factors.

7.4.2 Data Collection and Preparation

This section shows examples with distinct characteristics such as the arrangement of locations, passenger count, duration of time windows (whether short or long), and the number of taxis involved, which was introduced in Nagaoka on 15 March 2020. The driver's data acquisition method is input and recorded by the time, vehicle ID, direction, operating status, latitude, & longitude readings acquired from GPS. The acquirement of data was as follows:

- ➢ Pick-up time (unit: mints);
- ➢ Pick-up longitude (unit: degree);
- ➢ Pick-up latitude (unit: degree);
- ➢ Drop-off time (unit: mints);

- ➢ Drop-off longitude (unit: degree);
- ➢ Drop-off latitude (unit: degree);

In this research, after pre-processing, we have used taxi operation data such as id, the coordinates of the node (pick up and drop off longitude and latitude), service time (pick up and drop off time), the number of users, and operation status (departure and arrival time) in Table 7.1.

Table 7.1. Taxi data

Id	**X***	**Y***	**Service time (mint.)**	**Demand (Pass./Per trips)**	**TW start (mint.)**	**TW end (mint.)**
0	0	0	0	0	0	1440
1	-2.87603	0.22042	5	1	570	580
2	-2.11642	1.508638	5	1	627	637
3	-2.28939	1.490674	5	1	798	808
4	-2.13356	1.514184	5	1	881	891
5	-0.53049	-0.45244	5	-1	579	589
6	-6.76713	0.083942	5	-1	641	651
7	-3.58415	0.150614	5	-1	806	816
8	2.044301	-0.63208	5	-1	899	909

Where Id is the node id, X and Y are the coordinates of the node Customers position in Nagaoka city was represented on a Two-Dimensional XY plateau. *Customers' positions = (x,y), demand is positive for pickups and negative for drop-off, and TW start and TW end define the start and end of the time window for the node.

7.4.3 Mathematical Model

The shared taxi routing problem pertains to creating the most efficient routes for vehicles to serve a set of customer requests. Specifically, the service cannot be initiated before the start of the time window, nor extended beyond its end. Although a vehicle can arrive at a customer location earlier than the start of the time window, it must remain idle until the service's designated start time.

Let N be the set of the passenger pickup/drop-off locations. Let M be the set of all available taxis. Let T be the set of all periods. The graph $G = (V, A)$ consists of all passenger nodes, start depots, and end depots. Each taxi m has a limited capacity C_m. The cost and travel time from passenger i to the passenger are denoted by c_{ij} and T_{ij} respectively. The cost and the traveled are assumed times to satisfy the triangle inequality. Furthermore, S_i indicates the maximum number of passengers that passengers i is willing to share a taxi. Let P_i be the set of sharing preferences for passengers i. r_i: maximum wait time for passengers i. Let w_{ti} be the set passenger i by the waiting at time t. Let a_{tmi} be the taxi m arrives at passenger i 's location at time t. Let b_{tmi} be the taxi m drops off passenger i at time t.

There are three main types of decision variables in this mathematical model. The first one is x_{tij}, a binary variable, equal to 1 if passenger i is assigned to passenger j at time t and equal to 0 alternatively. It is possible that there may be insufficient taxis available to fulfill all the requests made. In such cases, these unfulfilled requests are added to a separate list known as the "request". While this problem prohibits the inclusion of any requests in the final solutions, the heuristic may employ the requests during a transitional phase. The second variable is y_{tmi} and is binary, it is set to 1 if taxi m is assigned to passenger i at time t, otherwise, it is set to 0. The third variable is s_{tmi} and is binary, it is set to 1, if taxi m serves passenger i at time t, otherwise it is set to 0. The other decision variables are $Share_{wanted_i}$ and is binary, it is set to 1, if passenger i wants to share a taxi, otherwise 0, $NO_{talking_i}$ and is binary, it is set to 1, if passenger i does not want to share a taxi with passengers who are talking on the phone, otherwise 0, and $talking_j$ and is also binary, it is set to 1, if passenger j is talking on the phone, otherwise 0. As a summary of all the above, the mathematical formulation of the problem is as follows.

Parameters:

N	:	the set of nodes (passenger pickup/drop off locations)
M	:	set of taxis
T	:	set of time periods
c_{ij}	:	the cost of assigning passenger i to passenger j
C_m	:	maximum capacity of taxi m

S_i	:	maximum number of passengers that passengers i is willing to share a taxi
P_i	:	set of sharing preferences for passenger i
$Share_{wanted_i}$	:	binary variable indicating if passenger i wants to share a taxi
$NO_{talking_i}$	:	binary variable indicating if passenger i does not want to share a taxi with passengers who are talking on the phone
$talking_j$	:	binary variable indicating if passenger j is talking on the phone
T_{ij}	:	travel time from node i to node j
r_i	:	maximum wait time for passenger i
w_{ti}	:	passenger i is waiting at time t
a_{tmi}	:	taxi m arrives at passenger i 's location at time t
b_{tmi}	:	taxi m drops off passenger i at time t

Decision Variables:

x_{tij}	:	binary variable indicating if passenger i is assigned to passenger j at time t
y_{tmi}	:	binary variable indicating if taxi m is assigned to passenger i at time t
s_{tmi}	:	binary variable indicating if taxi m serves passenger i at time t

Objective Function:

$$min \sum_{t \in T} \sum_{i \in N} \sum_{j \in N} c_{ij} x_{tij} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (7.1)$$

The objective function minimizes the total cost of all vehicles.

Constraints:

Passenger assignment constraints:

$$\sum_{i \in N, i \neq j} x_{tij} = 1 \qquad \forall t \in T, j \in N$$

$$\sum_{j \in N, i \neq j} x_{tij} = 1 \qquad \forall t \in T, i \in N$$

The constraints dictate that every vehicle must depart from its initial depot, reach its final depot, and when stopping at a customer node, it must both arrive at and depart from that node.

Taxi assignment constraints:

$$\sum_{m\in M} y_{tmj} = 1 \qquad \forall t \in T, j \in N$$

$$\sum_{j\in N} y_{tmj} \leq 1 \qquad \forall t \in T, m \in M$$

A constraint is in place to ensure that both the pickup and drop-off locations are serviced by the identical vehicle, while also verifying that the drop-off node is only visited if and when the pickup node is visited. And, constraints ensure that each passenger is assigned to a vehicle, each passenger is assigned to only one vehicle, passengers are assigned to the same vehicle at the same time, passengers who do not want to share a ride are assigned to a vehicle alone, and passengers who want to share a ride are assigned to a vehicle with other passengers.

Capacity constraints:

$$\sum_{j\in N} x_{tij} \leq \sum_{m\in M} y_{tmj} \, . \, C_m \qquad \forall t \in T, i \in N$$

The existence of constraints guarantees that the taxis are accurately established and meet capacity limitations.

Passenger behavior constraints:

$$x_{tij} \;\; \leq \; Share_{wanted_i} \; \forall t \in T, i \in N, j \in N$$

$$\sum_{j\in P_i} x_{tij} \;\; \leq \; Share_{wanted_i} . \, |P_i| \quad \forall t \in T, i \in N$$

Service constraints ensure that each passenger wants to share with other passengers the same taxi. For passengers who prefer not to share, the taxi should be assigned only to them.

$$x_{tij} \;\; \leq \left(1 - NO_{talking_i}\right) + talking_j \quad \forall t \in T, i \in N, j \in N$$

If a passenger requests a no-talking ride, then all other passengers in the same shared taxi must also request a no-talking ride. If a passenger requests a talking ride, then at least one other passenger in the same shared taxi must also request a talking ride.

Time constraints:

$$w_{ti} \leq w_{t+1,i} \; \forall t \in T, i \in N$$

$$w_{ti} \leq \sum_{m \in M} a_{tmi} \ \forall t \in T, i \in N$$

$$\sum_{j \in N} T_{ij} x_{tij} \leq w_{ti} + r_i \ \forall t \in T, i \in N$$

$$\sum_{m \in M} a_{tmi} \leq w_{ti} + r_i \ \forall t \in T, i \in N$$

$$\sum_{m \in M} T_{tmi} \leq w_{ti} + r_i \ \forall t \in T, i \in N$$

Arrival and drop-off time constraints:

$$a_{tmi} \leq \sum_{j \in N} T_{ij} x_{tij} \ \ \forall i \in N, m \in M, t \in T$$

$$b_{tmi} \leq \sum_{j \in N} T_{ij} x_{tij} \ \ \forall i \in N, m \in M, t \in T$$

$$a_{tmi} + T_{ij} \leq a_{t+1,mj} + B\left(1 - x_{tij}\right) \ \ \forall i \in N, \forall j \in N, m \in M, t \in T$$

$$b_{tmi} + T_{ij} \leq b_{t+1,mj} + B\left(1 - x_{tij}\right) \ \ \forall i \in N, \forall j \in N, m \in M, t \in T$$

The presence of constraints guarantees that a service will not commence at a particular location until the vehicle has the ability to reach that location. Additionally, the service cannot begin prior to the start of the time window or after the end of the time window at that specific location. The arrival time for a passenger at their destination should not be earlier than the lower bound of the time window specified for their drop-off time. Similarly, the arrival time should not be later than the upper bound of the time window. The drop-off time for a passenger at their destination should be after their pickup time plus the travel time from the pickup location to the drop-off location.

Service constraints:

$$\sum_{t=1}^{T} s_{tmi} \leq 1 \ \ \forall i \in N, m \in M$$

$$s_{tmi} \leq \sum_{j \in N} x_{tji} \ \ \ \forall i \in N, m \in M, t \in T$$

$$s_{tmi} \geq y_{tmi} + \sum_{j \in N} x_{tji} - 1 \quad \forall i \in N, m \in M, t \in T$$

Service constraints are a set of constraints that ensure that each passenger is served within a specific time window and that each taxi has a maximum number of passengers to serve.

7.4.4 Optimality Algorithm

In this section, we develop a metaheuristics algorithm for the shared taxi routing problem. Metaheuristics are generally more effective than optimization algorithms, iterative techniques, or basic heuristics in combinatorial optimization because they search through a larger range of feasible solutions.

One of the biggest challenges faced by heuristic algorithms is to prevent solutions from leading to a local optimum. A solution that is the best among a close-by group of potential solutions is known as the local optimum of an optimization problem. Most local search strategies are rendered worthless and unable to remove a solution from a local optimum when one occurs. One of the primary purposes for which metaheuristic techniques are utilized in algorithms is to avoid becoming trapped in a local optimum. Once again, the processing time presents a significant obstacle, as was stated in the metaheuristic technique. The likelihood that the algorithm will avoid local optimums increases with the number of initial solutions employed. So, with each new response, the amount of time needed to analyze them grows. This time-constrained delivery planning problem is resolved using the proposed method, one of the Metaheuristic methods.

The Metaheuristic procedure is a function that takes two inputs: the number of taxis required in the initial solution and the area radius. The function initializes global parameters such as the number of locations, distance, travel time, service time, and waiting time. Then, the first iteration begins and runs until all available taxis are used. Within the first iteration, for each taxi, the taxi parameters such as total time, travel time, service time, and waiting time are set to their initial states. Next, the heuristic procedure selects all pickup locations from the number of locations and starts the second iterative process. This process continues until the taxi has no more locations to pick up and drop off or has run out of time to return to the depot.

During the second iteration of the heuristic procedure, the search for the next optimal location begins. The procedure first selects locations where the taxi does not exceed the maximum capacity and places them in a list called possible locations. The open locations, where the current total time plus travel time falls between the initial and final time window, are classified as open possible locations. If the possible locations are empty, the next location to be visited will be the one with the least travel time plus waiting time from the current taxi position. If this list is also empty, the iteration process stops. After obtaining the open possible locations list, the procedure identifies the locations that are within the previously defined area radius and adds them to a new list called area locations. The next location is filled with the earliest closing location in the area locations list. So, if the area locations are empty, the next location variable is completed with the closest location from the open possible locations list. Once the next location is determined, the procedure checks the priority destinations. If one of the priority destinations becomes unavailable due to the assigned next location, the next location is switched to that specific priority destination.

After the total number of locations has been established, the taxi parameters are updated, and the current taxi's next location is removed from the total number of locations and pickup locations. If the taxi's next location is one of the total numbers of locations, its destination is added to the priority destinations list. The second iteration continues until all possible locations have been visited. At that point, a new taxi is set up, and the first iteration process proceeds until all locations have been visited or all taxis have been used. Once all locations have been visited or all taxis have been used, the global parameters are updated, and the function returns the final solution with the global and each taxi parameters. As a summary of all the above, the algorithm for this proposed method is as follows.

Step 1: Set the total passenger set,

Step 2: Find the first passenger from the passenger sets whose total cost from the pickup location to the passenger's departure point and the passenger's total cost on the return trip from the destination to the position is the lowest. If the set of passengers is empty, output the necessary number of taxis together with their routes, and then stop.

Step 3: Based on the previously defined operating path, compute the cost of all feasible insertion places for distributing cars from the second person onward. Then, insert the vehicle at the position with the lowest cost. Take this passenger out of the group of unmoved passengers. If there is no passenger, output the necessary number of taxis together with their routes, then stop.

Step 4: Until there are no more passengers that can be transported in the process.

Step 5: If there are no more passengers that can be transported in the same vehicle, prepare an additional vehicle and execute the process step 2.

7.5 Result and Discussion

In order to solve the problem using the Mixed-Integer Nonlinear Programming (MINLP) formulation described in equation (7.1), we solved the equation by metaheuristic algorithms and used hypothetical Nagaoka City general taxi data. The solving algorithms used in shared taxi operations can vary depending on the specific goals and objectives of the operation. Metaheuristic algorithms are optimization algorithms that can be applied to various problems, including those faced by shared taxi operations (Sano *et al.* (2020)).

These metaheuristic algorithms can be used in combination with other optimization techniques and data analysis tools to improve the efficiency and profitability of shared operations. To perform a cost-efficiency analysis between shared taxi and general taxi operations using GPS data, first, analyze the total operational costs incurred for both services. Then, compare the average fare rate of the two services, taking into account the differences in service scope, geographic coverage, and demand levels. Finally, using GPS data such as distance and time traveled, calculate the transportation cost (Table *7.2*) per km/trip for each service.

Table 7.2. Taxi transportation cost structure

Items	Percentage
Labor cost	72.8%
Fuel oil cost	8.5%
Vehicle repair cost	2.2%
Vehicle depreciation	1.3%
Insurance fee	2.3%
Non-operating expenses	1.0%
other expenses	11.8%

7.5.1 Comparison Shared and General Taxi in Nagaoka

The computational tests performed to evaluate the effectiveness of the heuristics are presented in this section. The algorithm was produced through the Google Collaboratory interface and was made using the Python programming language.

In this section, the number of vehicles required for two options namely, general and shared taxi services, in Nagaoka city has been compared for a specific number of passengers (400 passengers). This time, as a case study, we will target 400 actual vehicle trips on March 15, 2020, when there is an average demand for the taxi company, assuming the above-mentioned situation, and one of the heuristic solutions.

Table *7.3* shows 42 vehicles are required for 400 passengers when general taxi services are used by the passengers. On the other hand, only 8 vehicles are required if the passenger uses the shared taxi services. In terms of expenses, approximately 78.21%, 78.28%, and 86.53% of the cost are reduced for a company in the shared taxi service over the general taxi services in 10 mints, 20 mints, and 30-minute time windows respectively.

Table 7.3. Compare the shared and general taxi cost

	General Taxi	Shared Taxi		
Time Window	0-24hrs.	10min	20 min	30 min
Number of Taxis	42	8	8	4
Total of Distance	4333 km	1810.24 km	1784.24 km	1693.32 km
Fuel Cost	¥ 69,328	¥ 28,963.84	¥ 28,547.84	¥ 27,093.12
Driver Cost	¥ 403,200	¥76,800.00	¥ 76,800.00	¥ 38,400.00
Insurance Fee	¥11,507	¥ 2,191.78	¥ 2,191.78	¥ 1,095.89
Car maintains Cost	¥ 23,014	¥ 4,383.56	¥ 4,383.56	¥ 4,383.56
Car Buying	¥ 67,123	¥ 12,785.39	¥ 12,785.39	¥ 6,392.69
Total	¥ 574,172	¥ 125,124.57	¥ 124,708.57	¥ 77,365.27
Cost (%)		78.21%	78.28%	86.53%

The results of the study showed that the shared taxi service was more cost-efficient than the general taxi service. The shared taxis had a higher occupancy rate and traveled longer distances per trip compared to general taxis.

However, Table *7.3* represents the percentage difference in operational costs between shared taxis and general taxis, the results suggest that shared taxis may have a lower operational cost compared to general taxis in all three-time windows. The percentage differences may vary depending on the time window, with a smaller percentage difference in the 10-minute window and a larger percentage difference in the 30-minute window. This may be due to the fact that shared taxis, despite taking longer to reach their destination due to additional stops, may have a more optimized route and a higher passenger occupancy rate, which can reduce their operational cost per passenger. The actual operational cost of shared taxis and general taxis can vary widely depending on various factors such as fuel prices, maintenance costs, driver salaries, and demand for taxi services. Therefore, these results should be interpreted with caution and should be validated by a more thorough and context-specific analysis. This led to a lower cost per kilometer traveled for the shared taxis. Additionally, the study found that the shared taxi service could potentially reduce the number of vehicles needed to serve the same demand as the general taxi service. This could lead to a reduction in congestion and environmental impact.

Therefore, the cost-efficiency analysis based on GPS data suggests that shared taxi services can be a more cost-effective and environmentally friendly alternative to regular taxi services in certain settings.

7.5.2 The Impact of Operational Cost Efficiency on the Shared Taxi

The impact of operational cost efficiency for shared taxis can be significant in several ways:

Reduced Cost for Passengers: Shared taxi services have the potential to provide passengers with a more affordable transportation option than general taxis or private cars. By optimizing operational costs, shared taxi services can keep fares low and attract more passengers.

Increased Profit for Operators: Operational cost efficiency can also have a positive impact on the profit margins of shared taxi operators. By reducing the costs associated with operating the service, operators can increase their profits without increasing fares.

Reduced Traffic Congestion: Shared taxi services can help reduce traffic congestion by providing a more efficient means of transportation. By optimizing routes and encouraging shared rides, shared taxi services can reduce the number of vehicles on the road and alleviate traffic congestion.

Environmental Benefits: Shared taxi services can also have environmental benefits by reducing the number of vehicles on the road and decreasing carbon emissions. By optimizing routes and encouraging shared rides, shared taxi services can contribute to a more sustainable transportation system.

Therefore, operational cost efficiency is essential for the success of shared taxi services. It can help reduce costs for passengers, increase profits for operators, reduce traffic congestion, and provide environmental benefits.

7.6 Summary of the Chapter

In summary, this study aimed to analyze the cost efficiency of taxis as public transport. The use of GPS data and cost-efficiency analysis provides a valuable tool for understanding the operational performance of shared taxis and general taxi services. The hypothetical data shows that shared taxis can offer a more flexible, convenient, and cost-effective mode of transportation compared to general taxis, particularly when strategies such as route optimization and vehicle pooling are used. It is important to note that the specific results of this cost-efficiency analysis have depended on the particular context and market conditions in which the shared taxi services are operating.

Therefore, the use of GPS data and cost-efficiency analysis can provide a valuable starting point for policymakers, and transportation providers to understand the operational performance of shared taxi services and to identify strategies for improving their cost-efficiency and effectiveness. This study had certain limitations. For instance, the most comprehensive and current dataset available pertained only to a specific time frame. Furthermore, only direct expenses were taken into account during the analysis, and other costs such as cleaning, garaging, and administrative expenses were excluded. Additionally, the primary objective of this research was not to devise a new methodology; instead, existing tools were utilized to aggregate the transit-dependent population and transit coverage.

The current models only consider the economic viability, hence there is a need for future research to incorporate environmental concerns. In addition, the following general domains could be explored in upcoming research: (i) long-term analysis; (ii) geographic analysis; (iii) multi-modal analysis, social analysis, and (iv) technological analysis. These areas of research can help policymakers and investors make informed decisions about transportation infrastructure and investment to ensure efficient and sustainable transportation systems in the future.

Chapter 8. CONCLUSION AND FUTURE WORK

The study's accomplishments, contributions, limitations, and suggestions are now presented. It might be valuable for policy and new geometric design testing, among other things.

8.1 Summary of the Work

Taxi traffic is an important part of the urban area. If the number of taxis is too high, resulting in low per-taxi income and overall low revenue due to high operation costs since companies manage some excess number of fleets in waste of public resources, and if there are too few taxis it will cause inconvenience to the public. In areas like Niigata, we cannot lose operators as well as customers to overcome this bottleneck. The dissertation suggests a model that seamlessly brings benefits to both parties means reducing the fare of passengers and optimizing fleet planning on the operator's end.

This thesis developed two mathematical models for intractable optimization problems. The two models have been applied to a case study, using real data from Nagaoka and Sanjo City, Japan. We focused on three different areas where to find such problems: a) Machine Learning, b) the shared taxi fare model, and c) the taxi allocation model.

The first part of our study focused our attention on taxi demand for machine learning applications. Chapter 4 clarified the actual situation of taxi operation, GPS data of all vehicles, and 50 cars belonging to Mitsukoshi Taxi Co., Ltd., which is the largest taxi company with a business area of the entire Nagaoka city. Also, we developed a method to understand ML algorithms and established the usefulness of our method through various tests of real datasets. In the proposed method of clustering problems, we applied the approach to K-Near Neighbors, Linear Regression, Decision Trees, Random Forests, and Gradient Boosting problems in datasets with a moderate number of dimensions. We provided the first proof that the algorithm leads to the problems of one-dimensional data clustering. We developed an optimization-based approach for estimating Hybrid ML. We found that taxi demand has high predictability. The effectiveness of these models is evaluated using MAE, MAPE, MSE, RMSE, MdAE, and R^2. The HML algorithm consistently achieved the best classification performance.

We analyzed the actual taxi operation by using GPS data of all vehicles belonging to Sanjo Taxi Company, Ltd., the largest taxi company with a business area in the entire Niigata. The analysis period was in June 2015. Sanjo City, Niigata Prefecture, is in the center of Niigata Prefecture and is classified in the Chuetsu region. In the second part of this thesis, we developed a Mixed Integer Non-Linear Programming formulation model for the routing problem to study the Dial-A-Ride Problem (DARP) of shared taxis under busy travel times created by the ride shared themselves in Chapter 5. In Chapter 6, the model has its objective to minimize the total cost of passengers and the number of taxis utilized. This thesis targets the reduction of the passenger's effective fare because no attempts have been observed for the total taxi cost reduction. We have solved an effective Branch & Cut algorithm for the shared taxi fare problems. The problems have been solved in Python using a Gurobi solver showing that the proposed model can find the shortest route of taxi routing decisions that we compute the total cost of the shared taxi by developing a mathematical model. Establishing a generalized pricing method is necessary based on the service order, travel distance, number of shared passengers, and waiting time, the cost-sharing ratio. It can lower fare costs for each customer and reduce travel costs. Meanwhile, it increases the taxi company's revenue as it accepts drivers to pick up multiple passengers.

The third part of this study formulated the mathematical model for taxi allocation problems and routing problems in Chapter 5. The model has its to minimize taxi idle time, pick-up delays, and total costs in Chapter 6. We have solved the model by heuristic and metaheuristic methods such as the Greedy algorithm, Simulated Annealing, and Dynamic Greedy Programming that consider passengers, drivers, and time. We proposed three heuristic algorithms: greedy algorithm, simulated annealing algorithm, and dynamic greedy. Further, a case study has been carried out to demonstrate the proposed methods and show its implication by using real taxi data from Nagaoka, Japan. For the case study, we first identified the taxi travel hotspots as potential locations for taxi spots by investigating the pick-up and drop-off locations and times by analyzing the GPS data of the taxi. After that, we found the optimal taxi allocations by applying the proposed model and solution algorithms. To summarize, dynamic greedy heuristics successfully obtained excellent solutions in relatively short runtimes for the taxi issue, making strategic decisions and feasible choices for taxi markets to adopt. The case study application in this study demonstrates the potential for using data analysis, optimization of taxi allocations and the number of taxis, and reducing approximately 81.84% of CO_2 emissions in the transportation sector. Finally, sensitivity analysis was applied to validate the model for passenger demand. The sensitivity analysis showed that the total idle time cost increased if the demand increased.

Finally, in Chapter 7, the use of GPS data and cost-efficiency analysis provides a valuable tool for understanding the operational performance of shared taxi and general taxi services.

8.2 Recommendations for Future Works

In this study, we examined the possibility of taxi visualization, reduced operating costs, and improved productivity when a reservation system or joint dispatch business was introduced using GPS data of taxi vehicles, but the accuracy was further improved. However, it is necessary to determine the optimum value that is more realistic by using the distance on the road network and the link running speed by weather and time zone throughout the year and limiting the driver's working hours. Also, the solution used in this study is very primitive and not suitable for large-scale problems. Due to its value in decreasing the service failure rate, it is necessary to collect historical data and study ML methods to learn such distributions. Here are some potential recommendations for future works in the shared taxi market in Japan:

I. **Partnerships with local transportation companies**: Shared taxi companies can consider partnering with local transportation companies to expand their services and reach more customers. This can also help them to better integrate with existing transportation networks.

II. **Utilizing technology to improve efficiency:** Shared taxi companies can consider using technology to improve their efficiencies, such as real-time routing and scheduling algorithms. They can also consider implementing mobile apps that allow customers to book and track their rides. The idea can be extended to predict the fare and routes all over the mega City in Japan.

III. **Offering personalized services:** Shared taxi companies can differentiate themselves by offering personalized services, such as customized routes or in-vehicle amenities. This can help attract customers who value convenience and comfort.

IV. **Focusing on sustainability:** With growing concern for the environment, shared taxi companies can consider incorporating eco-friendly vehicles or implementing carbon offset programs to reduce their environmental impact.

V. **Engaging in community outreach:** Shared taxi companies can engage in community outreach initiatives to build goodwill and trust with customers. This can include partnerships with local businesses or community organizations or sponsoring local events.

VI. **Efficiency and utilization:** Investigating the efficiency and utilization of taxis and ride-sharing vehicles in a given area. This could involve analyzing data on trip durations, wait times, and vehicle occupancy rates to determine how these factors differ between traditional taxis and ride-sharing services.

VII. **Socioeconomic implications:** Examining the socioeconomic implications of the integration of ride-sharing and traditional taxi systems. This could involve studying the impact on employment, income distribution, and accessibility to transportation services in different socioeconomic groups.

VIII. **Environmental sustainability:** Assessing the environmental sustainability of ride-sharing services compared to traditional taxis. This could involve analyzing factors such as fuel consumption, emissions, and the potential for electric or hybrid vehicles in both taxi and ride-sharing fleets.

IX. **Exploring new markets:** Shared taxi companies can explore new markets, such as rural areas where public transportation options are limited. This can help them to reach a new customer base and expand their business.

Therefore, there are many potential areas for future growth and innovation in the shared taxi market in Japan. By focusing on customer needs, utilizing technology, and exploring new opportunities, shared taxi companies can continue to thrive and provide valuable transportation services to their customers.

BIBLIOGRAPHY

1. ABBAS, K. (2016). Estimation of Taxi Fleet Size: a Generic Algorithm. *Sigma Journal of Engineering and Natural Sciences,* **7**, 21–29.
2. ABDEL-BASSET, M., MOHAMED, R., JAMEEL, M. and ABOUHAWWASH, M. (2023). Spider wasp optimizer: a novel meta-heuristic optimization algorithm. *Artificial Intelligence Review*, **13**, 1–64 Available at: https://doi.org/10.1007/s10462-023-10446-y.
3. AGATZ, N., ERERA, A., SAVELSBERGH, M. and WANG, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, **223**, 295–303 Available at: http://dx.doi.org/10.1016/j.ejor.2012.05.028.
4. AGGOUNE-MTALAA, W. (2023). Analyzing Air Pollution and Traffic Data in Urban Areas in Luxembourg. *smart cities*, 929–943.
5. AINI, A., YAMANAKA, H., ISHIO, M. and SHIMADA, N. (2019). Planning of Flexible Transport Services for Rural Area Considering Operation Cost. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **75**, I_779-I_785.
6. AL-AYYASH, Z., ABOU-ZEID, M. and KAYSI, I. (2016). Modeling the demand for a shared-ride taxi service: An application to an organization-based context. *Transport Policy*, **48**, 169–182 Available at: http://dx.doi.org/10.1016/j.tranpol.2016.02.013.
7. ALI, S., MACIEJEWSKI, A.A., SIEGEL, H.J. and KIM, J.K. (2004). Measuring the robustness of a resource allocation. *IEEE Transactions on Parallel and Distributed Systems*, **15**, 630–641.
8. ALPAYDIN, E. (2004). *Introduction to Machine Learning.pdf*.
9. AN, S., HU, X. and WANG, J. (2011). Urban taxis and air pollution: a case study in Harbin, China. *Journal of Transport Geography*, **19**, 960–967 Available at: http://dx.doi.org/10.1016/j.jtrangeo.2010.12.005.
10. ANWAR, A., VOLKOV, M. and RUS, D. (2013). ChangiNOW: A mobile application for efficient taxi allocation at airports. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 694–701.
11. ARABANI, M. and BALOOCH SIRGANI, P. (2022). Evaluation of Performance and Moisture Sensitivity of Polymer-Modified Mixture Containing Glass Aggregates. *International Journal of Pavement Research and Technology*, **15**, 1166–1179 Available at: https://doi.org/10.1007/s42947-021-00080-8.
12. ASHKROF, P., HOMEM DE ALMEIDA CORREIA, G., CATS, O. and VAN AREM, B. (2022). Ride

acceptance behaviour of ride-sourcing drivers. *Transportation Research Part C: Emerging Technologies*, **142**, 103783 Available at: https://doi.org/10.1016/j.trc.2022.103783.

13. ASKARI, B., LE QUY, T. and NTOUTSI, E. (2020). Taxi Demand Prediction using an LSTM-Based Deep Sequence Model and Points of Interest. *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, 1719–1724.
14. ASSOCIATION, J.F. OF H.T. (2018). Optimizing and Revitalizing Business in the Taxi Industry. *Japan Federation of Hire-Taxi Associations*.
15. B.POWELL, W. (1988). A Comparative review of alternative algorithms for dynamic vehicle allocation problem.pdf. *Vehicle Routing: Methods and Studies*.
16. BAI, J., SO, K.C., TANG, C.S., CHEN, X. and WANG, H. (2019). Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing and Service Operations Management*, **21**, 556–570.
17. BALDACCI, R., MANIEZZO, V. and MINGOZZI, A. (2004). An exact method for the car pooling problem based on Lagrangean column generation. *Operations Research*, **52**, 422–439.
18. BEN-SMIDA, H.E., B, F.C. and KRICHEN, S. (2020). Iterated Granular Neighborhood Algorithm for the Taxi Sharing Problem. *Applications of Evolutionary Computation. EvoApplications 2020. Lecture Notes in Computer Science, vol 12104. Springer, Cham*, **1**, 291–304 Available at: https://doi.org/10.1007/978-3-030-43722-0_19.
19. BENTÉJAC, C., CSÖRGŐ, A. and MARTÍNEZ-MUÑOZ, G. (2021). *A comparative analysis of gradient boosting algorithms*. Springer Netherlands, 1937–1967 pp Available at: https://doi.org/10.1007/s10462-020-09896-5.
20. BIAU, G., CADRE, B. and ROUVIÈRE, L. (2019). Accelerated gradient boosting. *Machine Learning*, **108**, 971–992 Available at: https://doi.org/10.1007/s10994-019-05787-1.
21. BREIMAN, L. (2001). Random forests. *Machine Learning*, 5–32.
22. CHEN, C.Y., YAN, S. and WU, Y.S. (2019). A model for taxi pooling with stochastic vehicle travel times. *International Journal of Sustainable Transportation*, **13**, 582–596 Available at: https://doi.org/10.1080/15568318.2018.1490468.
23. CHEN, J., LIU, Z., ZHU, S. and WANG, W. (2015). Design of limited-stop bus service with capacity constraint and stochastic travel time. *Transportation Research Part E: Logistics and Transportation Review*, **83**, 1–15 Available at: http://dx.doi.org/10.1016/j.tre.2015.08.007.
24. CHEN, Y.C., LOH, C.H., WANG, F.C., CHEN, Z.J., FU, S.H. and WANG, C.Y. (2022). Application of Generative Adversarial Network to Optimize Vehicle Allocation at Dispatch

Stations of Paratransit Services. *Electronics (Switzerland)*, **11**.

25. CHEN, YINGGUO, CHEN, M., WEN, J., CHEN, YINGWU and XIANG, W. (2020). An Adaptive Large Neighborhood Search Algorithm for the Satellite Data Transmission Scheduling Problem. *International Journal of Aerospace Engineering*, **2020**.
26. COCHRANE, J. (1997). Time Series for Macro and Finance. *Spring*.
27. CORDEAU, J.F. (2006). A branch-and-cut algorithm for the dial-a-ride problem. *Operations Research*, **54**, 573–586.
28. CORDEAU, J.F. and LAPORTE, G. (2007). The dial-a-ride problem: Models and algorithms. *Annals of Operations Research*, **153**, 29–46.
29. CRAMA, Y. and PIRONET, T.L.A. (2019). Vehicle allocation problem with uncertain transportation requests over a multi-period rolling horizon. *Logistics Research*, **12**, 1–22.
30. CRUZ, C.A., MUNARI, P. and MORABITO, R. (2020). A branch-and-price method for the vehicle allocation problem. *Computers and Industrial Engineering*, **149**.
31. ČULÍK, K., HRUDKAY, K., KALAŠOVÁ, A., ŠTEFANCOVÁ, V. and NEDELIAKOVÁ, E. (2022). Impact of Technological Changes and Taxi Market Regulation on the Taxi Vehicle Fleets—The Case Study of Slovakia. *Vehicles*, **4**, 1158–1175.
32. ČULÍK, K., KALAŠOVÁ, A. and OTAHÁLOVÁ, Z. (2020). Alternative Taxi Services and their Cost Analysis. *Transportation Research Procedia*, **44**, 240–247 Available at: https://doi.org/10.1016/j.trpro.2020.02.047.
33. DANIEL FUENTES, P. and SANO, D.K. (2020). Metaheuristic algorithm for pickup and delivery vehicle route problems with time windows.
34. DE LA TORRE, R., CORLU, C.G., FAULIN, J., ONGGO, B.S. and JUAN, A.A. (2021). Simulation, optimization, and machine learning in sustainable transportation systems: Models and applications. *Sustainability (Switzerland)*, **13**, 1–21.
35. DE OLIVEIRA, A.A.M., SOUZA, M.P., PEREIRA, M. DE A., REIS, F.A.L., ALMEIDA, P.E.M., SILVA, E.J. and CREPALDE, D.S. (2015). Optimization of taxi cabs assignment in geographical location-based systems. *Proceedings of the Brazilian Symposium on GeoInformatics*, **2015-Novem**, 92–104.
36. DIMITRIS BERTSIMAS ; JOHN TSITSIKLIS. (2007). Simulated Annealing. *Statistical Science*, **8**, 10–15.
37. DREISEITL, S. and OHNO-MACHADO, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, **35**, 352–359.

38. ELTING, S. and EHMKE, J.F. (2021). Potential of Shared Taxi Services in Rural Areas-A Case Study. *Transportation Research Procedia*, **52**, 661–668 Available at: https://doi.org/10.1016/j.trpro.2021.01.079.
39. ENVIRONMENTAL PROTECTION AGENCY. (2005). Emission Facts. *United States Environmental Protection Agency*, 1–3.
40. ESLAMIPOOR, R. (2022). A fuzzy multi-objective model for supplier selection to mitigate the impact of vehicle transportation gases and delivery time. *Journal of Data, Information and Management*, **4**, 231–241 Available at: https://doi.org/10.1007/s42488-022-00076-4.
41. ESLAMIPOOR, R. (2023a). A Reliable and Sustainable Design of Supply Chain in Healthcare under Uncertainty Regarding Environmental Impacts. *Journal of Applied Research on Industrial Engineering*, **10**, 256–272.
42. ESLAMIPOOR, R. (2023b). An optimization model for green supply chain by regarding emission tax rate in incongruous vehicles. *Modeling Earth Systems and Environment*, **9**, 227–238 Available at: https://doi.org/10.1007/s40808-022-01470-y.
43. FAGÚNDEZ, G., MASSOBRIO, R. and NESMACHNOW, S. (2014). Online taxi sharing optimization using evolutionary algorithms. *Proceedings of the 2014 Latin American Computing Conference, CLEI 2014*.
44. FITRIA. (2013). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 1689–1699 pp.
45. FRANCKX, L. and MAYERES, I. (2016). Future trends in mobility: the rise of the sharing economy and automated transport Available at: http://www.mind-sets.eu/wordpress/wp-content/uploads/2015/11/D3.3-Future_Trends_in_Mobility_The_rise_of_the_sharing_economy_and_automated_transport_Annex_A.pdf.
46. FRIGUI, H. (2008). Clustering: Algorithms and applications. *2008 1st International Workshops on Image Processing Theory, Tools and Applications, IPTA 2008*.
47. FUKUMOTO, M., MATSUO, K. and MATSUMOTO, Y. (2017). A Study into the Factors Affecting the Number of Taxi Trips in Toyohashi, Japan. *Journal of the Eastern Asia Society for Transportation Studies*, **12**, 1434–1447.
48. FURUHATA, M., DESSOUKY, M., ORDÓÑEZ, F., BRUNET, M.E., WANG, X. and KOENIG, S. (2013). Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, **57**, 28–46 Available at: http://dx.doi.org/10.1016/j.trb.2013.08.012.

49. GENG, Y. and CASSANDRAS, C.G. (2013). New "smart parking" system based on resource allocation and reservations. *IEEE Transactions on Intelligent Transportation Systems*, **14**, 1129–1139.
50. GHAHRAMANI, M., GALLE, N.J., DUARTE, F., RATTI, C. and PILLA, F. (2021). Leveraging artificial intelligence to analyze citizens' opinions on urban green space. *City and Environment Interactions*, **10**, 100058 Available at: https://doi.org/10.1016/j.cacint.2021.100058.
51. GHAHRAMANI, M. and PILLA, F. (2021a). Analysis of Carbon Dioxide Emissions from Road Transport Using Taxi Trips. *IEEE Access*, **9**, 98573–98580.
52. GHAHRAMANI, M. and PILLA, F. (2021b). Leveraging artificial intelligence to analyze the COVID-19 distribution pattern based on socio-economic determinants. *Sustainable Cities and Society*, **69**.
53. GHAHRAMANI, M., ZHOU, M. and WANG, G. (2020). Urban sensing based on mobile phone data: Approaches, applications, and challenges. *IEEE/CAA Journal of Automatica Sinica*, **7**, 627–637.
54. GHILAS, V., DEMIR, E. and VAN WOENSEL, T. (2016). An adaptive large neighborhood search heuristic for the Pickup and Delivery Problem with Time Windows and Scheduled Lines. *Computers and Operations Research*, **72**, 12–30.
55. GLASCHENKO, A., IVASCHENKO, A., RZEVSKI, G. and SKOBELEV, P. (2009). Multi-agent real time scheduling system for taxi companies. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, **2**, 1320–1326.
56. HE, F., WANG, X., LIN, X. and TANG, X. (2018). Pricing and penalty/compensation strategies of a taxi-hailing platform. *Transportation Research Part C: Emerging Technologies*, **86**, 263–279 Available at: https://doi.org/10.1016/j.trc.2017.11.003.
57. HOSNI, H., NAOUM-SAWAYA, J. and ARTAIL, H. (2014). The shared-taxi problem: Formulation and solution methods. *Transportation Research Part B: Methodological*, **70**, 303–318 Available at: http://dx.doi.org/10.1016/j.trb.2014.09.011.
58. HOU, Q., GUAN, Y. and YU, S. (2020). Stochastic differential game model analysis of emission-reduction technology under cost-sharing contracts in the carbon trading market. *IEEE Access*, **8**, 167328–167340.
59. HOU, Y. (2010). Forecast on consumption gap between cities and countries in China based on ARMA model. *Proceedings - 3rd International Conference on Intelligent Networks and Intelligent Systems, ICINIS 2010*, 342–345.

60. HSU, J.Y. (2010). Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining*, **5**, 3–18.

61. HU, C. and THILL, J.C. (2019). Predicting the upcoming services of vacant taxis near fixed locations using taxi trajectories. *ISPRS International Journal of Geo-Information*, **8**.

62. HUA, S., ZENG, W., LIU, X. and QI, M. (2022). Optimality-guaranteed algorithms on the dynamic shared-taxi problem. *Transportation Research Part E*, **164**, 102809 Available at: https://doi.org/10.1016/j.tre.2022.102809.

63. HUANG, K., LIU, F., HU, Y. and LIU, Z. (2018). An analysis of the taxi-sharing organizing and pricing. *Lecture Notes in Electrical Engineering*, **419**, 263–276.

64. HUANG, Y. and POWELL, J.W. (2012). Detecting Regions of Disequilibrium in Taxi Services. 139–148.

65. INTURRI, G., GIUFFRIDA, N., IGNACCOLO, M., LE PIRA, M., PLUCHINO, A., RAPISARDA, A. and D'ANGELO, R. (2021). Taxi vs. demand responsive shared transport systems: An agent-based simulation approach. *Transport Policy*, **103**, 116–126 Available at: https://doi.org/10.1016/j.tranpol.2021.01.002.

66. JAMES COOPER, R.M. AND J.N. (2010). *Taxi! Urban Economies and the Social and Transport Impacts of the Taxicab*.

67. JAPAN METEOROLOGICAL AGENC. (2019). Japan Meteorological Agenc.

68. JUNG, J., JAYAKRISHNAN, R. and PARK, J.Y. (2013). Design and Modeling of Real-time Shared-Taxi Dispatch Algorithms. *TRB Annual Meeting*, **8**, 1–20.

69. JUNG, J., JAYAKRISHNAN, R. and PARK, J.Y. (2016). Dynamic Shared-Taxi Dispatch Algorithm with Hybrid-Simulated Annealing. *Computer-Aided Civil and Infrastructure Engineering*, **31**, 275–291.

70. KANG, S. (2021). K-nearest neighbor learning with graph neural networks. *Mathematics*, **9**.

71. KAWANABE, I. (2022). Taxi Today in Japan 2022 (japanese). *Japan Federation of Hire-Taxi Associations*, 1–7.

72. KAYA, C.B., KAYA, E. and GÖKKUŞ, G. (2023). Training Neuro-Fuzzy by Using Meta-Heuristic Algorithms for MPPT. *Computer Systems Science and Engineering*, **45**, 69–84.

73. KE, J., CEN, X., YANG, H., CHEN, X. and YE, J. (2019). Modelling drivers' working and recharging schedules in a ride-sourcing market with electric vehicles and gasoline vehicles. *Transportation Research Part E: Logistics and Transportation Review*, **125**, 160–180 Available at: https://doi.org/10.1016/j.tre.2019.03.010.

74. KE, J., YANG, H., LI, X., WANG, H. and YE, J. (2020). Pricing and equilibrium in on-demand

ride-pooling markets. *Transportation Research Part B: Methodological*, **139**, 411–431 Available at: https://doi.org/10.1016/j.trb.2020.07.001.

75. KEITH W. HIPEL, A.I.M. (1994). TIME SERIES MODELLING OF WATER RESOURCES AND ENVIRONMENTAL SYSTEMS. *Elsevier*.
76. KONG, X., XIA, F., WANG, J., RAHIM, A. and DAS, S.K. (2017). Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data. *IEEE Transactions on Industrial Informatics*, **13**, 1202–1212.
77. LABBÉ, M. and LAPORTE, G. (1986). Maximizing user convenience and postal service efficiency in post box location. *Belgian Journal of Operations Research, Statistics and Computer Science*, **26**, 21–35.
78. LAHARIYA, C. (2008). The State of the World Population 2007: Unleashing the potential of urban growth. *Indian Pediatrics*, **45**, 481–482.
79. LAZAROIU, G.C. and ROSCIA, M. (2012). Definition methodology for the smart cities model. *Energy*, **47**, 326–332.
80. LI, B., ZHANG, D., SUN, L., CHEN, C., LI, S., QI, G. and YANG, Q. (2011). Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. *2011 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2011*, 63–68.
81. LI, J., BOONAERT, J., DONIEC, A. and LOZENGUEZ, G. (2021). Multi-models machine learning methods for traffic flow estimation from Floating Car Data. *Transportation Research Part C: Emerging Technologies*, **132**, 103389 Available at: https://doi.org/10.1016/j.trc.2021.103389.
82. LI, J., LIU, Z. and WANG, X. (2022). Public charging station localization and route planning of electric vehicles considering the operational strategy: A bi-level optimizing approach. *Sustainable Cities and Society*, **87**, 104153 Available at: https://doi.org/10.1016/j.scs.2022.104153.
83. LI, N., SUN, Y., YU, J., LI, J.C., ZHANG, H.F. and TSAI, S. (2019). An Empirical Study on Low Emission Taxiing Path Optimization of Aircrafts on Airport Surfaces from the Perspective of Reducing Carbon Emissions. *Energies*, **12**.
84. LI, X., PAN, G., WU, Z., QI, G., LI, S., ZHANG, D., ZHANG, W. and WANG, Z. (2012). Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science in China*, **6**, 111–121.
85. LI, Y., SHAHABI, C., FU, K., YE, J., WANG, Z. and LIU, Y. (2018). Multi-task representation

learning for travel time estimation. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1695–1704.

86. LIAO, S., ZHOU, L., DI, X., YUAN, B. and XIONG, J. (2018). Large-scale short-term urban taxi demand forecasting using deep learning. *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*, **2018-Janua**, 428–433.
87. LIN, C.K.Y. (2014). Solving a location, allocation, and capacity planning problem with dynamic demand and response time service level. *Mathematical Problems in Engineering*, **2014**.
88. LIU, S., WANG, S., LIU, C. and KRISHNAN, R. (2015). Understanding taxi drivers' routing choices from spatial and social traces. *Frontiers of Computer Science*, **9**, 200–209.
89. LOKHANDWALA, M. and CAI, H. (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. *Transportation Research Part C: Emerging Technologies*, **97**, 45–60 Available at: https://doi.org/10.1016/j.trc.2018.10.007.
90. LU, C.C., YAN, S. and HUANG, Y.W. (2018). Optimal scheduling of a taxi fleet with mixed electric and gasoline vehicles to service advance reservations. *Transportation Research Part C: Emerging Technologies*, **93**, 479–500.
91. LUO, H., CAI, J., ZHANG, K., XIE, R. and ZHENG, L. (2021). A multi-task deep learning model for short-term taxi demand forecasting considering spatiotemporal dependences. *Journal of Traffic and Transportation Engineering (English Edition)*, **8**, 83–94 Available at: https://doi.org/10.1016/j.jtte.2019.07.002.
92. MA, S., ZHENG, Y. and WOLFSON, O. (2015). Real-Time City-Scale Taxi Ridesharing. *IEEE Transactions on Knowledge and Data Engineering*, **27**, 1782–1795.
93. MAHMOUDI, M. and ZHOU, X. (2016). Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state-space-time network representations. *Transportation Research Part B: Methodological*, **89**, 19–42.
94. MIAO, F., HAN, S., LIN, S., WANG, Q., STANKOVIC, J.A., HENDAWI, A., ZHANG, D., HE, T. and PAPPAS, G.J. (2019). Data-Driven Robust Taxi Dispatch under Demand Uncertainties. *IEEE Transactions on Control Systems Technology*, **27**, 175–191.
95. MINAMI, R., SANO, K. and NISHIUCHI, H. (2016). Actual Condition of the Sanjo City'S Shared-Taxi Focusing on Rideshare (Japanese). *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **72**, I_743-I_750.
96. MINGOLLA, S. and LU, Z. (2021). Carbon emission and cost analysis of vehicle technologies

for urban taxis. *Transportation Research Part D: Transport and Environment*, **99**, 102994 Available at: https://doi.org/10.1016/j.trd.2021.102994.

97. MIWA, T., ISHIGURO, Y., YAMAMOTO, T. and MORIKAWA, T. (2013). Allocation planning for probe taxi devices based on information reliability. *Transportation Research Part C: Emerging Technologies*, **34**, 55–69 Available at: http://dx.doi.org/10.1016/j.trc.2013.05.004.
98. MOHAMMED, M., KHAN, M.B. and BASHIER, E.B.M. (1985). *Machine learning: Algorithms and Applications*, 83–132 pp.
99. MONDAL, M., SANO, K., WATARI, T., PUPPATERAVANIT, CHONNIPA, PERERA, F. and SILVA, C.K. DE. (2021). An Efficient Modeling Approach on Passenger Demand with Fare in Shared Taxis: A Case Study of Sanjo City, Japan. *Journal of the Eastern Asia Society for Transportation Studies*, **14**, 684–701 Available at: https://www.jstage.jst.go.jp/result/journal/-char/en?cdjournal=easts&globalSearchKey=manik.
100. MOREIRA-MATIAS, L., GAMA, J., FERREIRA, M., MENDES-MOREIRA, J. and DAMAS, L. (2013). Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, **14**, 1393–1402.
101. MUKAI, N. and YODEN, N. (2012). Taxi demand forecasting based on taxi probe data by neural network. *Smart Innovation, Systems and Technologies*, **14**, 589–597.
102. NEOH, J.G., CHIPULU, M. and MARSHALL, A. (2017). What encourages people to carpool? An evaluation of factors with meta-analysis. *Transportation*, **44**, 423–447.
103. NGUYEN, N.Q., TUAN, K.L.E., NGHIEM, N.V.D., NGUYEN, M.S., THUAN, P.D.O. and MUKAI, N. (2015). People and parcels sharing a taxi for Tokyo city. *ACM International Conference Proceeding Series*, **03-04-Dece**, 90–97.
104. NICOLAIDES, D., CEBON, D. and MILES, J. (2019). An Urban Charging Infrastructure for Electric Road Freight Operations: A Case Study for Cambridge UK. *IEEE Systems Journal*, **13**, 2057–2068.
105. NIE, Y. (MARCO). (2017). How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. *Transportation Research Part C: Emerging Technologies*, **79**, 242–256 Available at: http://dx.doi.org/10.1016/j.trc.2017.03.017.
106. NILRIT, S., SAMPANPANISH, P. and BUALERT, S. (2018). Carbon dioxide and methane emission rates from taxi vehicles in Thailand. *Carbon Management*, **9**, 37–43 Available at: https://doi.org/10.1080/17583004.2017.1412232.

107. O’Keeffe, K., Anklesaria, S., Santi, P. and Ratti, C. (2021). Using reinforcement learning to minimize taxi idle times. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, **0**, 1–16 Available at: https://doi.org/10.1080/15472450.2021.1897803.

108. OECD and European Conference of Ministers of Transport. (2007). *(De)Regulation of the Taxi Industry*.

109. Omrani, H. (2015). Predicting travel mode of individuals by machine learning. *Transportation Research Procedia*, **10**, 840–849 Available at: http://dx.doi.org/10.1016/j.trpro.2015.09.037.

110. Ouyang, W. and Zhu, X. (2023). Meta-Heuristic Solver with Parallel Genetic Algorithm Framework in Airline Crew Scheduling. *Sustainability*, **15**, 1506.

111. Paraboschi, A., Santi, P. and Ratti, C. (2015). Modeling urban-level impact of a shared taxi market. *CUPUM 2015 - 14th International Conference on Computers in Urban Planning and Urban Management*.

112. Peng, Z., Shan, W., Zhu, X. and Yu, B. (2022). Many-to-one stable matching for taxi-sharing service with selfish players. *Transportation Research Part A: Policy and Practice*, **160**, 255–279 Available at: https://doi.org/10.1016/j.tra.2022.04.012.

113. Pepić, L. (2018). The sharing economy: Uber and its effect on taxi companies. *Acta Economica*, **16**, 123–136.

114. Postorino, M.N., Mantecchini, L. and Paganelli, F. (2019). Improving taxi-out operations at city airports to reduce CO2 emissions. *Transport Policy*, **80**, 167–176 Available at: https://doi.org/10.1016/j.tranpol.2018.09.002.

115. Pourakbari-Kasmaei, M., Lehtonen, M., Contreras, J. and Mantovani, J.R.S. (2020). Carbon footprint management: A pathway toward smart emission abatement. *IEEE Transactions on Industrial Informatics*, **16**, 935–948.

116. Qian, S., Cao, J., Le Mouël, F., Sahel, I. and Li, M. (2015). SCRAM: A sharing considered route assignment mechanism for fair taxi route recommendations. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, **2015-Augus**, 955–964.

117. Qiu, J., Huang, K. and Hawkins, J. (2022). The taxi sharing practices: Matching, routing and pricing methods. *Multimodal Transportation*, **1**, 100003 Available at: https://doi.org/10.1016/j.multra.2022.100003.

118. Qu, M., Zhu, H., Liu, J., Liu, G. and Xiong, H. (2014). A cost-effective recommender

system for taxi drivers. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 45–54.

119. RAHIMI, F., SADEGHI-NIARAKI, A., GHODOUSI, M. and CHOI, S.M. (2021). Modeling population spatial-temporal distribution using taxis origin and destination data. *Sustainability (Switzerland)*, **13**, 1–13.

120. RAJENDRAN, S., SRINIVAS, S. and GRIMSHAW, T. (2021). Predicting demand for air taxi urban aviation services using machine learning algorithms. *Journal of Air Transport Management*, **92**, 102043 Available at: https://doi.org/10.1016/j.jairtraman.2021.102043.

121. RAJENDRAN, S. and ZACK, J. (2019). Insights on strategic air taxi network infrastructure locations using an iterative constrained clustering approach. *Transportation Research Part E: Logistics and Transportation Review*, **128**, 470–505 Available at: https://doi.org/10.1016/j.tre.2019.06.003.

122. RAMEZANI, M. and NOURINEJAD, M. (2018). Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach. *Transportation Research Part C: Emerging Technologies*, **94**, 203–219 Available at: https://doi.org/10.1016/j.trc.2017.08.011.

123. RAO, H. S., HETTIGE, H., SINGRU, N., LUMAIN, R., & ROLDAN, C. (2010). Reducing Carbon Emissions from Transport Projects. *Evaluation Knowledge Brief*, **16**, 1–96.

124. REIHANEH, M. and GHONIEM, A. (2019). A branch-and-price algorithm for a vehicle routing with demand allocation problem. *European Journal of Operational Research*, **272**, 523–538 Available at: https://doi.org/10.1016/j.ejor.2018.06.049.

125. RODRIGUES, F., MARKOU, I. and PEREIRA, F.C. (2019). Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. *Information Fusion*, **49**, 120–129 Available at: https://doi.org/10.1016/j.inffus.2018.07.007.

126. RONG, K., XIAO, F. and WANG, Y. (2019). Redundancy in the sharing economy. *Resources, Conservation and Recycling*, **151**, 104455.

127. ROPKE, S., CORDEAU, J.F. and LAPORTE, G. (2007). Models and branch-and-cut algorithms for pickup and delivery problems with time windows. *Networks*, **49**, 258–272.

128. ROPKE, S. and PISINGER, D. (2006). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, **40**, 455–472.

129. SALANOVA, J.M., ROMEU, M.E. and AMAT, C. (2014). Aggregated Modeling of Urban Taxi Services. *Procedia - Social and Behavioral Sciences*, **160**, 352–361.

130. SANO, K., KAI, W., HATOYAMA, K., TAKAHASHI, T. and WATARI, T. (2020). Analysis on Taxi Operation Using GPS Data and Possibility of Its Efficiency Improvement. *JSTE Journal of Traffic Engineering*.

131. SANO, K., WISETJINDAWAT, W., SUGA, Y. and RAOTHANACHONKUN, P. (2007). A Study on the Benefits of Improving Local Bus Service. *Journal of the Eastern Asia Society for Transportation Studies*, **7**, 1575–1583.

132. SANTOS, D.O. and XAVIER, E.C. (2013). Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem. *IJCAI International Joint Conference on Artificial Intelligence*, 2885–2891.

133. SANTOS, D.O. and XAVIER, E.C. (2015). Taxi and ride sharing: A dynamic dial-a-ride problem with money as an incentive. *Expert Systems with Applications*, **42**, 6728–6737 Available at: http://dx.doi.org/10.1016/j.eswa.2015.04.060.

134. SARICA, A., CERASA, A. and QUATTRONE, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Frontiers in Aging Neuroscience*, **9**, 1–12.

135. SCHALLER, B. (2007). Entry controls in taxi regulation: Implications of US and Canadian experience for taxi regulation and deregulation. *Transport Policy*, **14**, 490–506.

136. SHAH, K., PATEL, H., SANGHVI, D. and SHAH, M. (2020). A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. *Augmented Human Research*, **5**, 1–16 Available at: https://doi.org/10.1007/s41133-020-00032-0.

137. SHEN, W. and LOPES, C. (2015). Managing autonomous mobility on demand systems for better passenger experience. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **9387**, 20–35.

138. SHI, Y. and LIAN, Z. (2016). Optimization and strategic behavior in a passenger-taxi service system. *European Journal of Operational Research*, **249**, 1024–1032 Available at: http://dx.doi.org/10.1016/j.ejor.2015.07.031.

139. SINGH, R., SHARMA, C. and AGRAWAL, M. (2017). Emission inventory of trace gases from road transport in India. *Transportation Research Part D: Transport and Environment*, **52**, 64–72 Available at: http://dx.doi.org/10.1016/j.trd.2017.02.011.

140. SONG, Y., ZHANG, M. and SHAN, C. (2019). Research on the decoupling trend and mitigation potential of CO2 emissions from China's transport sector. *Energy*, **183**, 837–843 Available at: https://doi.org/10.1016/j.energy.2019.07.011.

141. SUJATH, R., CHATTERJEE, J.M. and HASSANIEN, A.E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, **34**, 959–972 Available at: https://doi.org/10.1007/s00477-020-01827-8.

142. SUN, H., WANG, H. and WAN, Z. (2019). Model and analysis of labor supply for ride-sharing platforms in the presence of sample self-selection and endogeneity. *Transportation Research Part B: Methodological*, **125**, 76–93 Available at: https://doi.org/10.1016/j.trb.2019.04.004.

143. SUN, M. and FAN, H. (2021). Detecting and analyzing urban centers based on the localized contour tree method using taxi trajectory data: A case study of shanghai. *ISPRS International Journal of Geo-Information*, **10**.

144. TAKEUCHI, R., MINAMI, S., INOI, H. and YOSHIDA, I. (2022). Productivity and cost structure of transportation passenger services using private cars (paratransit) in Japan. *Asian Transport Studies*, **8**, 100085 Available at: https://doi.org/10.1016/j.eastsj.2022.100085.

145. TANIZAKI, T. (2013). Optimum allocation method of standby taxi vehicles at taxi stands. *IFIP Advances in Information and Communication Technology*, **415**, 3–10.

146. TAYLOR, T.A. (2018). On-Demand Service Platforms. *Manufacturing & Service Operations Management*, **20**, 704–720.

147. TREVOR HASTIE, ROBERT TIBSHIRANI, J.F. (2008). The elements of statistical Learning Data Mining, Inference, and Prediction. *Spring Series in Statistics*.

148. VANICHRUJEE, U. (2017). *TAXI DEMAND PREDICTION USING ENSEMBLE MODEL AND TAXI GPS DATA ANALYSIS*.

149. VANICHRUJEE, U., HORANONT, T., THEERAMUNKONG, T., PATTARA-ATIKOM, W. and SHINOZAKI, T. (2018). Taxi Demand Prediction using Ensemble Model Based on RNNs and XGBOOST. *2018 International Conference on Embedded Systems and Intelligent Technology and International Conference on Information and Communication Technology for Embedded Systems, ICESIT-ICICTES 2018*, 1–6.

150. VASCO, R.A. and MORABITO, R. (2016). The dynamic vehicle allocation problem with application in trucking companies in Brazil. *Computers and Operations Research*, **76**, 118–133 Available at: http://dx.doi.org/10.1016/j.cor.2016.04.022.

151. VAZIFEH, M.M., SANTI, P., RESTA, G., STROGATZ, S.H. and RATTI, C. (2018). Addressing the minimum fleet problem in on-demand urban mobility. *Nature*, **557**, 534–538.

152. WALLSTEN, S. (2015). The Competitive Effects of the Sharing Economy: How is Uber

Changing Taxis? *Technological Policy Institute*, 1–22.

153. WANG, D., GONG, Y., ZHANG, X. and LIN, Y. (2022). The Impacts of Multiscale Urban Road Network Centrality on Taxi Travel: A Case Study in Shenzhen. *Complexity*, **2022**.

154. WANG, F. and LI, J. (2020). The Application of Taxis Time-Sharing Pricing Under the Influence of Sharing Economy | SpringerLink. *Springer Proceedings in Mathematics and Statistics*, 621–634.

155. WANG, Q., HUANG, Z., HU, J. and LIU, Z. (2020). A carbon emission evaluation method for remanufacturing process of a used vehicle CVT gearbox. *IEEE Access*, **8**, 193257–193267.

156. WANG, RAN, CHOW, C.Y., LYU, Y., LEE, V.C.S., KWONG, S., LI, Y. and ZENG, J. (2018). TaxiRec: Recommending road clusters to taxi drivers using ranking-based extreme learning machines. *IEEE Transactions on Knowledge and Data Engineering*, **30**, 585–598.

157. WANG, Y., WEI, C. and CHEN, G. (2017). The Simulation of Dynamic Shared Taxi Routing Problem. *DEStech Transactions on Computer Science and Engineering*, 382–391.

158. WANG, YAZHE, ZHENG, B. and LIM, E.P. (2018). Understanding the effects of taxi ride-sharing — A case study of Singapore. *Computers, Environment and Urban Systems*, **69**, 124–132 Available at: https://doi.org/10.1016/j.compenvurbsys.2018.01.006.

159. WANG, Z., CHEN, F. and FUJIYAMA, T. (2015). Carbon emission from urban passenger transportation in Beijing. *Transportation Research Part D: Transport and Environment*, **41**, 217–227 Available at: http://dx.doi.org/10.1016/j.trd.2015.10.001.

160. WATEL, D. and FAYE, A. (2018). Taxi-sharing: Parameterized complexity and approximability of the dial-a-ride problem with money as an incentive. *Theoretical Computer Science*, **745**, 202–223.

161. WHITMORE, A., SAMARAS, C., HENDRICKSON, C.T., SCOTT MATTHEWS, H. and WONG-PARODI, G. (2022). Integrating public transportation and shared autonomous mobility for equitable transit coverage: A cost-efficiency analysis. *Transportation Research Interdisciplinary Perspectives*, **14**.

162. WONG, K.I., WONG, S.C. and YANG, H. (2001). Modeling urban taxi services in congested road networks with elastic demand. *Transportation Research Part B: Methodological*, **35**, 819–842.

163. WONG, K.I., WONG, S.C., YANG, H. and WU, J.H. (2008). Modeling urban taxi services with multiple user classes and vehicle modes. *Transportation Research Part B: Methodological*, **42**, 985–1007.

164. WONG, R.C.P., SZETO, W.Y. and WONG, S.C. (2014a). Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: Modeling methodology and policy implications. *Transport Policy*, **33**, 73–81 Available at: http://dx.doi.org/10.1016/j.tranpol.2014.02.011.

165. WONG, R.C.P., SZETO, W.Y., WONG, S.C. and YANG, H. (2014b). Modelling multi-period customer-searching behaviour of taxi drivers. *Transportmetrica B: Transport Dynamics*, **2**, 40–59.

166. XIE, C., LU, J. and PARKANY, E. (2003). Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. *Transportation Research Record*, 50–61.

167. XIE, P., LI, T., LIU, J., DU, S., YANG, X. and ZHANG, J. (2020). Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion*, **59**, 1–12 Available at: https://doi.org/10.1016/j.inffus.2020.01.002.

168. XU, J., RAHMATIZADEH, R., BÖLÖNI, L. and TURGUT, D. (2021). A taxi dispatch system based on prediction of demand and destination. *Journal of Parallel and Distributed Computing*, **157**, 269–279 Available at: https://doi.org/10.1016/j.jpdc.2021.07.002.

169. YANG, H., LAU, Y.W., WONG, S.C. and LO, H.K. (2000). A macroscopic taxi model for passenger demand, taxi utilization and level of services. *Transportation*, **27**, 317–340.

170. YANG, H., LEUNG, C.W.Y., WONG, S.C. and BELL, M.G.H. (2010). Equilibria of bilateral taxi-customer searching and meeting on networks. *Transportation Research Part B: Methodological*, **44**, 1067–1083 Available at: http://dx.doi.org/10.1016/j.trb.2009.12.010.

171. YANG, H., QIN, X., KE, J. and YE, J. (2020a). Optimizing matching time interval and matching radius in on-demand ride-sourcing markets. *Transportation Research Part B: Methodological*, **131**, 84–105.

172. YANG, H., SHAO, C., WANG, H. and YE, J. (2020b). Integrated reward scheme and surge pricing in a ridesourcing market. *Transportation Research Part B: Methodological*, **134**, 126–142 Available at: https://doi.org/10.1016/j.trb.2020.01.008.

173. YANG, H. and WONG, S.C. (1998). A network model of urban taxi services. *Transportation Research Part B: Methodological*, **32**, 235–246.

174. YANG, H. and YANG, T. (2011). Equilibrium properties of taxi markets with search frictions. *Transportation Research Part B: Methodological*, **45**, 696–713 Available at: http://dx.doi.org/10.1016/j.trb.2011.01.002.

175. YANG, H., YE, M., TANG, W.H. and WONG, S.C. (2005). Regulating taxi services in the

presence of congestion externality. *Transportation Research Part A: Policy and Practice*, **39**, 17–40.

176. YANG, J., KWON, Y. and KIM, D. (2021). Regional Smart City Development Focus: The South Korean National Strategic Smart City Program. *IEEE Access*, **9**, 7193–7210.

177. YAO, B., JIN, L., CAO, Q., GAO, J. and ZHANG, M. (2016). Fleet size and fare optimization for taxi under dynamic demand. *Journal of Transport Literature*, **10**, 45–50.

178. YAO, H., HUANG, Z., YE, X., GRIFOLL, M., LIU, G. and ZHENG, P. (2022). Analysis of Taxi Travels during an Epidemic Period Using System Dynamics Method. *Sustainability*, **14**, 3457.

179. YAO, H., WU, F., KE, J., TANG, X., JIA, Y., LU, S., GONG, P., LI, Z., YE, J. and CHUXING, D. (2018). Deep multi-view spatial-temporal network for taxi demand prediction. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2588–2595.

180. YAO, S., DONG, Z., WANG, X. and REN, L. (2020). A Multiobjective multifactorial optimization algorithm based on decomposition and dynamic resource allocation strategy. *Information Sciences*, **511**, 18–35 Available at: https://doi.org/10.1016/j.ins.2019.09.058.

181. YE, Q., MA, C., HE, R., XIAO, Q. and ZHANG, W. (2015). Multi-objective optimisation for taxi ridesharing route based on non-dominated sorting genetic algorithm. *International Journal of Wireless and Mobile Computing*, **8**, 262–270.

182. YOU, L., GUAN, Z., LI, N., ZHANG, J., CUI, H., CLARAMUNT, C. and CAO, R. (2021). A spatio-temporal schedule-based neural network for urban taxi waiting time prediction. *ISPRS International Journal of Geo-Information*, **10**.

183. YU, B., SONG, X., GUAN, F., YANG, Z. and YAO, B. (2016). K-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. *Journal of Transportation Engineering*, **142**, 1–10.

184. YU, J.J., TANG, C.S., SHEN, Z.J.M. and CHEN, X.M. (2020). A balancing act of regulating on-demand ride services. *Management Science*, **66**, 2975–2992.

185. YU, K.A.I., JI, L. and ZHANG, X. (2002). Kernel Nearest-Neighbor Algorithm. *Kluwer Academic Publishers*, 147–156.

186. YU, V.F., INDRAKARNA, P.A.Y., REDI, A.A.N.P. and LIN, S.W. (2021). Simulated annealing with mutation strategy for the share-a-ride problem with flexible compartments. *Mathematics*, **9**.

187. YUAN, J., ZHENG, Y., ZHANG, L., XIE, X. and SUN, G. (2011). Where to Find My Next Passenger? *In Proceedings of the 13th international conference on Ubiquitous computing*,

109–118.

188. YUDHISTIRA, G., FIRDAUS, M.I. and AGUSHINTA, L. (2015). Transportation System in Japan: a Literature Study. *Jurnal Manajemen Transportasi & Logistik (JMTRANSLOG)*, **2**, 333.

189. ZANTALIS, F., KOULOURAS, G., KARABETSOS, S. and KANDRIS, D. (2019). A review of machine learning and IoT in smart transportation. *Future Internet*, **11**, 1–23.

190. ZHA, L., YIN, Y. and DU, Y. (2018). Surge pricing and labor supply in the ride-sourcing market. *Transportation Research Part B: Methodological*, **117**, 708–722.

191. ZHAI, W., BAI, X., PENG, Z. REN and GU, C. (2019). A bottom-up transportation network efficiency measuring approach: A case study of taxi efficiency in New York City. *Journal of Transport Geography*, **80**.

192. ZHAN, X., LAFAYETTE, W., LAFAYETTE, W., UKKUSURI, S. V and LAFAYETTE, W. (2014). Measuring the Efficiency of Urban Taxi Service System. *UrbComp'14*.

193. ZHAN, X., QIAN, X. and UKKUSURI, S. V. (2016). A Graph-Based Approach to Measuring the Efficiency of an Urban Taxi Service System. *IEEE Transactions on Intelligent Transportation Systems*, **17**, 2479–2489.

194. ZHANG, C. and NIAN, J. (2013). Panel estimation for transport sector CO2 emissions and its affecting factors: A regional analysis in China. *Energy Policy*, **63**, 918–926 Available at: http://dx.doi.org/10.1016/j.enpol.2013.07.142.

195. ZHANG, D., HE, T., LIN, S., MUNIR, S. and STANKOVIC, J.A. (2016). Taxi-Passenger-Demand Modeling Based on Big Data from a Roving Sensor Network. *IEEE Transactions on Big Data*, **3**, 362–374.

196. ZHANG, H., WU, S., FENG, O., TIAN, T., HUANG, Y. and ZHONG, G. (2023). Research on Demand-Based Scheduling Scheme of Urban Low-Altitude Logistics UAVs. *Applied Sciences (Switzerland)*, **13**.

197. ZHANG, L., LONG, R., LI, W. and WEI, J. (2020). Potential for reducing carbon emissions from urban traffic based on the carbon emission satisfaction: Case study in Shanghai. *Journal of Transport Geography*, **85**, 102733 Available at: https://doi.org/10.1016/j.jtrangeo.2020.102733.

198. ZHANG, W., HE, R., CHEN, Y., GAO, M. and MA, C. (2019). Research on Taxi Pricing Model and Optimization for Carpooling Detour Problem. *Journal of Advanced Transportation*, **2019**.

199. ZHANG, Y. and XIE, Y. (2008). Travel mode choice modeling with support vector

machines. *Transportation Research Record*, 141–150.

200. ZHAO, K., KHRYASHCHEV, D., FREIRE, J., SILVA, C. and VO, H. (2016). Predicting taxi demand at high spatial resolution: Approaching the limit of predictability. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, **3**, 833–842.

201. ZHAO, K., KHRYASHCHEV, D. and VO, H. (2019). Predicting Taxi and Uber Demand in Cities: Approaching the Limit of Predictability. *IEEE Transactions on Knowledge and Data Engineering*, **33**, 2723–2736.

202. ZHAO, L., WEN, X., WANG, Y. and SHAO, Y. (2022). A novel hybrid model of ARIMA-MCC and CKDE-GARCH for urban short-term traffic flow prediction. *IET Intelligent Transport Systems*, **16**, 206–217.

203. ZHAO, P.X., QIN, K., ZHOU, Q., LIU, C.K. and CHEN, Y.X. (2015). DETECTING HOTSPOTS from TAXI TRAJECTORY DATA USING SPATIAL CLUSTER ANALYSIS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **2**, 131–135.

204. ZHENG, X., LIANG, X. and XU, K. (2012). Where to wait for a taxi? *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 149–156.

205. ZHENG, Z., RASOULI, S. and TIMMERMANS, H. (2018). Modeling taxi driver anticipatory behavior. *Computers, Environment and Urban Systems*, **69**, 133–141 Available at: https://doi.org/10.1016/j.compenvurbsys.2018.01.008.

206. ZHU, Z., QIN, X., KE, J., ZHENG, Z. and YANG, H. (2020). Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. *Transportation Research Part A: Policy and Practice*, **132**, 713–727 Available at: https://doi.org/10.1016/j.tra.2019.12.018.

APPENDIX A. PUBLICATIONS

1. MONDAL M. SANO K., WATARI T, PUPPATERAVANIT C., PERERA F., DE SILVA C., An Efficient Modelling Approach on Passenger Demand with Fare in Shared Taxis: A Case Study of Sanjo City, Japan, *Journal of the Eastern Asia Society for Transportation Studies,* 2022 Volume 14 p. 684-701. https://doi.org/10.11175/easts.14.684.
2. Mondal M., Sano K. Teppei K, and Chonnipa P., Optimization of Taxi Allocation for Minimizing CO2 Emissions Based on Heuristics Algorithms, *Smart Cities,* 2023, Volume 6, Issue 3, 1589-1611, https://doi.org/10.3390/smartcities6030075
3. PERERA F., SANO K., DE SILVA C., PUPPATERAVANIT C., MONDAL M. and WATARI T., Determining the Impact of Runway Configuration on an Airport Taxiing Process with Developing Airport Categorizing Method, *Journal of the Eastern Asia Society for Transportation Studies,* 2022 Volume 14 p. 2345-2364. https://doi.org/10.11175/easts.14.2345.
4. PUPPATERAVANIT C., SANO K., HATOYAMA K., Frank PERERA K.P.D, MONDAL M., WATARI T., Impact of COVID-19 on Residential Self-selection and Travel Behavior Change, *Journal of the Eastern Asia Society for Transportation Studies,* 2022 Volume 14 p. 1532-1552. https://doi.org/10.11175/easts.14.1532
5. NGUYEN T., SANO K., MONDAL M., Linh T. T., Factors affect on the use of bicycle in Hanoi, Vietnam, *The 14th International Conference of Eastern Asia Society for Transportation Studies in Hiroshima,* Japan on September 12-15, 2021, Hiroshima, Japan.
6. Q. T. Nhu Phan, M. Mondal and S. Kazushi, "Hybrid Model to Predict the Arrival Time at Tollgate of Vehicles on Expressway," *2022 10th International Conference on Traffic and Logistic Engineering (ICTLE),* Macau, China, 2022, pp. 41-45, doi: 10.1109/ICTLE55577.2022.9901808.
7. Q. T. N. Phan, M. Mondal and S. Kazushi, "Application of LSTM and ANN Models for Traffic Time Headway Prediction in Expressway Tollgates," *2022 Moratuwa Engineering Research Conference (MERCon),* Moratuwa, Sri Lanka, 2022, pp. 1-6, doi: 10.1109/MERCon55799.2022.9906226.

APPENDIX B. MAIN CODE-1 (Shared taxi Fare)

```
# In[1]:
import sys
import math
from gurobipy import *
from gurobipy import Model, GRB, quicksum
import scipy.sparse as sp
import pandas as pd
import numpy as np
import datetime
import matplotlib.pyplot as plt
from itertools import combinations
stime = datetime.datetime.now()
# In[2]:
path = 'I:\\25.11.20\\New gurobi code\My code gurobi\RD\Book06_RD_4_10.csv'
#path = 'H:\\New\\Book20_5.csv'
my_data = pd.read_csv(path, header=None, index_col=None, names=['Id', 'x', 'y', 'serviceTime', 'demand', 'early_limit', 'late_limit', 'Destination'])
my_data = np.array(my_data)
# In[3]:
L = my_data.shape[0]
K = int((L - 2) / 2)

# initial fare
capacity = 3
Ki = 2
Mf = 630
a2 = 90
Id = 1.2
# x = my_data[:, 1]
# y = my_data[:, 2]
s = my_data[:, 3]
```

```
d = my_data[:, 4]
e = my_data[:, 5]
l = my_data[:, 6]
O_D = my_data[:, 7]

# plt.plot(x[0], y[0], c='r', marker='s')
# plt.scatter(x[:], y[:], c='b')

speed = 1
timestamp = np.zeros((L, L))
DistanceMatrix = np.zeros((L, L))
dN = [1, 1/2, 1/3]
res = np.zeros((K, 1))
tnum = {}
  for i in list(range(0, L, 1)):
   for j in list(range(0, L, 1)):
      DistanceMatrix[i, j] = ((my_data[j, 1] - my_data[i, 1])**2 + (my_data[j, 2] - my_data[i,
2])**2)**0.5
      timestamp[i,j]= DistanceMatrix[i, j]/ speed

# In[4]:
# Variable
model = Model("FST")
# model.params.timelimit = 5*60
# In[5]:
x = {}
for i in range(0, L, 1):
   for j in range(0, L, 1):
      for k in range (0, K, 1):
         x[i,j,k] = model.addVar(vtype=GRB.BINARY, name = "x(%s,%s,%s)" %  (i, j, k))
model.update()
y = {}
for i in range(0, L, 1):
```

```
    for j in range(0, L, 1):
        for k in range (0, K, 1):
            y[i,j,k] = model.addVar(vtype=GRB.BINARY, name = "y(%s,%s,%s)" %  (i, j, k))
model.update()
z = {}
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range (0, K, 1):
            z[i,j,k] = model.addVar(vtype=GRB.BINARY, name = "z(%s,%s,%s)" %  (i, j, k))
model.update()
w = {}
for i in range(0, L, 1):
    for k in range(0, K, 1):
        for r in range(0, capacity, 1):
            w[i,k,r] = model.addVar(vtype=GRB.BINARY, name = "w(%s,%s,%s)" % (i, k, r))
model.update()
 # a1= model.addVar(vtype=GRB.CONTINUOUS, name = "a1" )
# model.update()
S = model.addVar(vtype=GRB.BINARY, name = "S" )
model.update()


C = {}
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range (0, K, 1):
            C[i,j,k] = model.addVar(vtype=GRB.CONTINUOUS, name = "C(%s,%s,%s)" %  (i, j,
k))
model.update()
pC = {}
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range (0, K, 1):
```

```
            pC[i,j,k] = model.addVar(vtype=GRB.CONTINUOUS, name = "pC(%s,%s,%s)" %  (i,
j, k))
model.update()
B = {}
for i in range(0, L, 1):
    for k in range(0, K, 1):
        B[i,k] = model.addVar(vtype=GRB.CONTINUOUS, name = "B(%s,%s)" % (i, k))
model.update()

N = {}
for i in range(0, L, 1):
    for k in range(0, K, 1):
        N[i,k] = model.addVar(vtype=GRB.CONTINUOUS, name = "N(%s,%s)" % (i, k))
model.update()

R = {}
for i in range(0, L, 1):
   for j in range(0, L, 1):
       R[i,j] = model.addVar(vtype=GRB.CONTINUOUS, name = "R(%s,%s)" %  (i, j))
model.update()
Wt = {}
for i in range(0, L, 1):
    Wt[i] = model.addVar(vtype=GRB.CONTINUOUS, name = "Wt(%s)" %  (i))
model.update()

# In[6]:
#Constraints
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range (0, Ki, 1):
            model.addConstr((z[i,j,k] == 1) >> (C[i,j,k] == Mf), name = "D1")
            model.addConstr((z[i,j,k] == 0) >> (C[i,j,k] >= Mf + R[i,j]), name = "D2")
model.update()
```

```
#Equation 1 flow
for j in range(1, (L-1), 1):
    model.addConstr(quicksum(x[i,j,k] for i in range(0, L, 1) for k in range(0, Ki, 1)) == 1, name = "1b")
    model.addConstr(quicksum(x[j,i,k] for i in range(0, L, 1) for k in range(0, Ki, 1)) == 1, name = "1c")
    model.addConstr(quicksum(x[j,j,k] for k in range(0, Ki, 1)) == 0, name = "1d")
    model.addConstr(quicksum(x[j,0,k] for k in range(0, Ki, 1)) == 0, name = "1e")
    model.addConstr(quicksum(x[(L-1),j,k] for k in range(0, Ki, 1)) == 0, name = "1f")
model.update()

for k in range(0, Ki, 1):
    for i in range(1, (L-1), 1):
        model.addConstr(quicksum(x[j,i,k] for j in range(0, L, 1)) - quicksum(x[i,j,k] for j in range(0, L, 1)) == 0, name = "1j")
        model.addConstr(quicksum(x[i,j,k] for j in range(0, L, 1)) - quicksum(x[j,O_D[i],k] for j in range(0, L, 1)) == 0, name = "1k")
model.update()
for k in range(0, Ki, 1):
    model.addConstr(quicksum(x[0,j,k] for j in range(1, (L-1), 1)) == 1, name = "1g")
    model.addConstr(quicksum(x[i,(L-1),k] for i in range(1, (L-1), 1)) == 1, name = "1h")
    model.addConstr(x[0,(L-1),k] + x[(L-1),0,k] == 0, name = "1i")
model.update()
#Equation 2 Time Window
for k in range (0, Ki, 1):
    for i in range (0, L, 1):
        for j in range (0, L, 1):
            model.addConstr((x[i,j,k] == 1) >> (B[i,k] + (s[i] + timestamp[i,j]) * x[i,j,k] <= B[j,k]), name ="2a")
        model.addConstr(B[i,k] >= e[i], name = "2b")
        model.addConstr(B[i,k] <= l[i], name = "2c")
        # model.addConstr(B[i,k] >= 0, name = "2d")
model.update()
```

```
 #Equation 3 Capacity
for k in range (0, Ki, 1):
    for i in range(0, L, 1):
        for j in range(0, L, 1):
            model.addConstr((x[i,j,k] == 1) >> (N[i,k] + (d[i] * x[i,j,k]) <= N[j,k]), name = "3a")
        model.addConstr(N[i,k] >= 0, name = "3b")
        model.addConstr(N[i,k] <= capacity, name = "3c")
    model.addConstr(N[0,k] == 0, name = "3d")
    model.addConstr(N[L-1,k] == 0, name = "3e")
model.update()
#Equation 7 Fare rate
for i in range(0, L, 1):
    for j in range(0, L, 1):
        model.addConstr(R[i,j] == 0*Wt[i] + a2*(quicksum(x[i,j,k]*DistanceMatrix[i, j] for k in
range(0, Ki, 1)) - Id), name = "7")
model.update()
#Equation_Additional 2
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range(0, Ki, 1):
            model.addConstr(N[i,k] * y[i,j,k] == 0, name = "A2a")
            model.addConstr(N[i,k] + y[i,j,k] >= 1, name = "A2b")
model.update()
for k in range(0, Ki, 1):
    for i in range(0, L, 1):
        for j in range(0, L, 1):
            model.addConstr((y[i,j,k] == 1) >> (quicksum(w[i,k,r] for r in range(0, capacity, 1)) ==
0), name = "A2c")
            model.addConstr((y[i,j,k] == 0) >> (quicksum(w[i,k,r] * dN[r] for r in range(0,
capacity, 1)) == 1), name = "A2d")
        #model.addConstr((quicksum(w[i,k,r] for r in range(0, capacity, 1)) == N[i, k]), name =
"A2d")
model.update()
```

```
#Equation_Additional 3
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range (0, Ki, 1):
            model.addConstr(pC[i,j,k] >= quicksum(C[i,j,k] * (w[i,k,r] * dN[r]) for r in range(0,
capacity, 1)), name = "A3")
model.update()
#Equation_Additional 4
for i in range(0, L, 1):
    for j in range(0, L, 1):
        for k in range(0, Ki, 1):
            model.addConstr((quicksum(x[i,j,k]*DistanceMatrix[i, j] for k in range(0, Ki, 1)) - Id)
* z[i,j,k] <= 0, name = "A4a")
            model.addConstr(quicksum(x[i,j,k]*DistanceMatrix[i, j] for k in range(0, Ki, 1)) + (Id
* z[i,j,k]) >= Id, name = "A4b")
model.update()
# In[7]:
AFsum2 = quicksum((pC[i,j,k] * x[i,j,k])
        for i in range(0, L, 1)
        for j in range(0, L, 1)
        for k in range(0, Ki, 1))
model.setParam("NonConvex", 2)
Tsum =  AFsum2
model.setObjective(Tsum, GRB.MINIMIZE)
model.optimize()
print('*******************************')
model.write("FST.lp")
res = np.zeros((L, L, K))
rnp = np.zeros((L, L))
rax = np.zeros((1, 3))
rnp = np.zeros((L, K))
route = np.zeros((Ki, L))
route[:, 1:] = (L - 1)
```

```
for k in range(0, Ki, 1):
    for i in range(0, L, 1):
        for j in range(0, L, 1):
            res[i, j, k] = x[i, j, k].X
            rnp[i, k] = N[i, k].X
rax[0, 1] = a2
rax[0, 2] = Id
 for k in range(0, Ki, 1):
    i = 0
    n = 1
    while i != (L-1):
        for j in range(0, L, 1):
            if res[i, j, k] != 0:
                route[k, n] = j
                i = j
                n = n + 1
                break
wtc = 0.0
dtc = 0.0
for k in range(0, Ki, 1):
    for i in range(0, L, 1):
        for j in range(0, L, 1):
            if (rnp[i, k] >= 1) and (res[i, j, k] == 1):
                # wtc = wtc + (rax[0, 0] * rwt[i, j])
                dtc = dtc + (rax[0, 1] * DistanceMatrix[i, j]) - rax[0, 2]
print(wtc)
print(dtc )
```