Data-Driven Analysis of Optimal Repositioning in Dockless Bike-Sharing Systems

by

Emre Berk Unsal

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Sciences
in
Management Sciences

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Bike-sharing systems provide sustainable and convenient mobility services for short-distance transportation in urban areas. The dockless or free-floating bike-sharing systems allow users to leave vehicles at any location in the service zones which leads to an imbalance of inventory between different areas across a city. Hence, vehicles in such dockless bike-sharing systems need to be repositioned throughout the day to be able to capture and serve more demand. In this study, we analyze the impact of optimal repositioning on the efficiency of dockless bike-sharing systems under several performance measures. We first develop a multi-period network flow model to find the optimal repositioning decisions which consist of the origin, destination, and the time of the repositioning that maximize the total profit of the bike-sharing system. The proposed model is then implemented on the real-world bike-sharing data of New York, Toronto, and Vancouver. After finding the optimal repositioning actions, we analyze the effect of repositioning on the fulfilled demand, the number of required vehicles, and the utilization rates of the vehicles. Through computational experiments, we show that repositioning significantly increases the efficiency of bike-sharing systems under these performance measures. In particular, our analyses show that up to 41% more demand can be satisfied with repositioning. Moreover, it is possible to reduce the required fleet size up to 61% and increase the average utilization rate of the vehicles up to 21% by employing repositioning. We also demonstrate that the effect of optimal repositioning is robust against the uncertainty of demand.

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my thesis supervisor Dr. Sibel Alumur Alev for her leadership, advices, criticism and deep knowledge throughout this thesis. Without her guidance and help this research would not have been possible.

I would like to thank Dr. Ada Hurst and Dr. Fatih Safa Erenay for their contributions and valuable comments.

I would also like to thank to Zeynep and my family for their unconditional support. Their tremendous encouragement kept my spirits and motivation high during this research.

Lastly, I am also grateful to all people who made this research possible.

Dedication

"Science is the most reliable guide for civilization, for life, for success in the world. Searching a guide other than science is meaning carelessness, ignorance, and heresy."

Mustafa Kemal Ataturk

Table of Contents

List of Figures

List of Tables

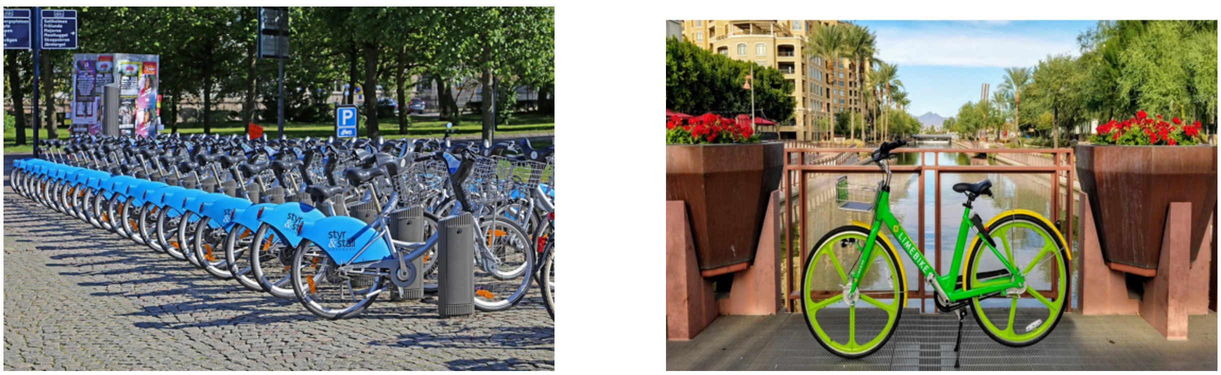
Chapter 1

Introduction

As the motor vehicle population ascends in cities, people are struggling with short-distance transportation at desired times due to traffic congestion. As a substitute for public transportation, the bike- and scooter-sharing systems, providing short-term rentals of these vehicles, have emerged in the last years. Although the first bike-sharing system was implemented in 1965 in the Netherlands (Kabak et al., 2018), the new generation systems are revolutionizing the transportation in last decades through the use of Global Positioning Systems (GPS) and electric vehicles (Chen et al., 2020).

The demand for bike-sharing systems steadily increases in time (DeMaio, 2009). Li et al. (2019) expect this increasing trend to continue since transportation with bikes and scooters has numerous advantages such as not being stuck in traffic. Moreover, the authors assert that people tend to use sharing systems rather than private bikes and scooters due to the possibility of theft, which is also a boosting factor in the ascending demand.

The bike-sharing systems are classified into two groups; *station-based* and *dockless* bike-sharing systems (Lazarus et al., 2020). Station-based bike-share systems allow people to rent and pick up bikes only from a pre-built bike station and drop them off at another station. The dockless bike-sharing systems, which are also known as free-floating bike-sharing systems, on the other hand, allow people to drop off the vehicle at any point in the city. The pick-up locations of the vehicles, which actually correspond to the drop-off locations of the previous trips, are generally provided to customers by a mobile application or website. Figure 1.1 illustrates these two types of bike-sharing systems.



(a) Station-based (Styr&Ställ, 2022)

(b) Dockless (Lime, 2022)

Figure 1.1: Bike-sharing systems

The dockless bike-sharing systems trended in recent years as they provide flexible and sustainable transportation (Younes et al., 2020). Xu et al. (2019) underlines the cost-effectiveness of the dockless sharing systems and therefore indicates that this new generation of dockless sharing systems is likely to outperform the station-based systems.

The dockless bike-sharing systems allow customers to leave the rented vehicles at any location they want; accordingly, the customers do not have to end their trips at a station. These sharing systems use GPS that attains the location and time information on each vehicle from the satellites. The GPS is incorporated into mobile applications in which customers can see the available vehicles and their current locations on a map. After signing up in the mobile application, customers can rent the vehicles by arriving at the vehicles' existing locations. At the end of their trip, which refers to the travel from an origin to a destination location, the customers are allowed to leave the bikes at any location. GPS system detects the location where the customer is dropping off the vehicle and prepares the bill. Although there are alternative systems that may allow customers to pay flat monthly fees, the bill is usually calculated based on the duration of the trip. The mobile application requests the payment information of customers while the customers are signing up and the fee is then automatically charged.

In dockless bike-sharing systems that utilize electric vehicles, the operating companies tend to implement an *overnight charging* process. In this strategy, all vehicles in the system are collected at the end of a day for recharge and redistributed to service zones at the beginning of each day (Chen et al., 2018). Recharging is usually done during off-peak hours, such as before 7 am (Bird, 2022). Alternatively, the empty batteries of the vehicles can be swapped during the day.

In bike-sharing systems, as the customers do not have to return the vehicles to the origin location, it is possible to have an imbalance of inventories between stations in station-based systems, and on streets for dockless systems. This imbalance results in unfulfilled demand during the day because potential customers may not find any vehicles in some desired locations. Accordingly, to be able to satisfy more demand and increase their revenue, the companies that operate these systems might need to change the locations of their vehicles throughout the day, which is referred to as *repositioning* or *rebalancing*. In particular, one of the operating companies named "Lime", deploy a team which is composed of approximately 250 employees in San Francisco who are responsible for the rebalancing actions throughout the day as well as being responsible for overnight charging (Vox, 2018). The vehicles can also be recharged or their batteries can be swapped during the repositioning actions (Osorio et al., 2021).

In this study, we develop a mixed-integer programming model to create an optimal repositioning strategy that maximizes the total profit of the bike-sharing system. Given a service area, planning horizon, and demand, the repositioning strategy consists of determining the origin, destination, and the time of the repositioning. In other words, the model aims to determine the number of vehicles to be repositioned between locations and when to do it on the planning horizon. After determining the optimal repositioning decisions, our aim is to analyze these decisions on real-world data sets to measure the effect of repositioning on different performance indicators and also to derive insights from repositioning.

Our contributions in this study are fourfold: (*i*) development of a mathematical model to optimize for the repositioning decisions throughout a multi-period planning horizon, (*ii*) implementation of the model on real-world data sets from three different cities, (*iii*) presenting a data-driven analysis of the optimal repositioning decisions and their effects on several performance measures, and (*iv*) verifying the effects of repositioning with scenarios under demand uncertainty.

This thesis is organized as follows. Chapter 2 presents an overview of the literature related to the problem. In Chapter 3, we describe the repositioning problem in dockless bike-sharing systems and propose a model to optimize these decisions. We evaluate the performance of the proposed model using real-world data from three different cities in Chapter 4. In the last chapter, Chapter 5, we present some concluding remarks and future research directions.

Chapter 2

Literature Review

The repositioning, also known as rebalancing, strategies in bike-sharing systems have been drawing attention in recent years with the increase in interest in sharing systems. The repositioning strategies in bike-sharing systems are grouped into two main categories: operator-based repositioning and user-based repositioning (Jin and Tong, 2020). In an operator-based repositioning strategy, the repositioning processes are conducted by the company which operates the bike-sharing system, whereas, in a user-based repositioning strategy, the company offers incentives to their customers to reposition the vehicles to the desired location. The research on user-based repositioning strategies mainly focuses on dynamic pricing and bidding strategies to convince the customers to reposition the vehicles throughout the planning horizon. The research on operator-based repositioning strategies, on the other hand, focuses on optimizing the vehicle routes that are used by the operating companies to reposition the vehicles by owner-operated trucks or vans.

Among the user-based repositioning research, Cheng et al. (2021) proposes a dynamic bidding-model-based incentive mechanism to encourage the customers to participate in the rebalancing process. In particular, they offer the customers a price incentive if they agree to take the vehicles to the areas that lack inventory. Similarly, Neijmeijer et al. (2020) introduce a dynamic pricing model for user-based rebalancing in dockless bike-sharing systems. In this model, the authors test their algorithm in a real-world case and find out that the incentives are able to overcome the supply-demand asymmetry in bike-sharing systems. Reiss and Bogenberger (2017) build up a hybrid approach by combining the user-based rebalancing strategies with operator-based rebalancing strategies. Accordingly, they define an urgency index for each rebalancing action in the system. This urgency index is calculated based on the magnitude of the imbalance in the inventories. If the urgency index is above a threshold level, Reiss and Bogenberger (2017) propose to use operator-based rebalancing actions, otherwise, they suggest using price incentives to attract customers to perform rebalancing.

The operator-based repositioning strategies are further categorized into two sub-strategies: static repositioning and dynamic repositioning (Médard de Chardon et al., 2016). In static repositioning, the user intervention in the bike-sharing system throughout the planning horizon is negligible. In dynamic repositioning, on the other hand, the system is still in use while the vehicles are being repositioned. Another difference between the static and dynamic repositioning strategies is that in static repositioning, the demand data is revealed at the beginning of the planning horizon, whereas in dynamic repositioning, the demand information changes throughout the planning horizon.

Contardo et al. (2012) present a mathematical formulation for dynamic rebalancing problems in bike-sharing systems. To be able to solve large-scale instances they implement Dantzig-Wolfe and Benders decomposition methods. As a result, they find the lower and upper bounds to this problem: however, they are not able to reduce the gap between these lower and upper bounds. Chiariotti et al. (2018) analyze the bike-sharing data of New York and estimate the inventory levels of the stations dynamically. Based on their forecast, they determine the time of the rebalancing actions and afterward implement a heuristic to create the rebalancing routes. Similarly, Cipriano et al. (2021) implement frequent pattern mining on the Barcelona bike-sharing system and propose a dynamic rebalancing method based on this pattern mining. Using the data, they analyze the critical stations that are likely to cause customer dissatisfaction due to inventory shortages. Afterward, they dynamically plan the rebalancing actions to overcome the possible shortages.

The research related to static repositioning strategies in bike-sharing systems approach the problem from a vehicle routing perspective. For example, Erdoğan et al. (2014) identifies the problem as a variant of the One Commodity Pickup and Delivery Traveling Salesman Problem (1-PDTSP) which is introduced by Hernández-Pérez and Salazar-González (2007). In the 1-PDTSP, there is a given list of customers and a single vehicle. This single vehicle picks up certain amount of commodities from some customers and delivers these commodities to another customers. Erdoğan et al. (2014) finds the minimum cost route of the single capacitated vehicle which redistributes the bicycles to stations while considering demand constraints. Erdoğan et al. (2014) proposes exact solution algorithms for this problem and tests the proposed algorithms with the instances that are proposed by Hernández-Pérez and Salazar-González (2007). The maximum number of the nodes (customers) in these test instances is 50.

To evaluate different static repositioning strategies, Dell'Amico et al. (2014) tests four mixed-integer linear programming models on the data obtained for the city of Reggio Emilia, Italy. The first two models are based on the well-known Multiple Traveling Salesman Problem (m-TSP), presented in Bektas (2006). The third approach is similar to 1-PDTSP studied in Erdoğan et al. (2014). The last model is inspired by the two-commodity flow model proposed by Baldacci et al. (2004). All four models aim to determine the repositioning routes with minimum total transportation cost. Dell'Amico et al. (2014) implements a branch-and-cut algorithm to solve the models and tests the performance of the algorithm on real-world data from different cities. The number of nodes in these instances varies between 13 and 116.

Pal and Zhang (2017) presents a novel mixed-integer model to find the optimum repositioning routes. This model is similar to 1-PDTSP but allows for multiple visits to a node with the same vehicle as well as the use of multiple vehicles. Bruck et al. (2019) formulates the problem similar to Pal and Zhang (2017) but extends the concept by considering the forbidden temporary operations. The forbidden operations are the cases that prevent some of the vehicles in the fleet to be transported.

In this study, we focus on static repositioning decisions in dockless bike-sharing systems, however, in contrast with the previous studies, we do not optimize the routing decisions. Instead, we focus on determining the optimal repositioning decisions which include the origin and destination of the repositioning as well as its time. To incorporate the time aspect, we consider a multi-period planning horizon, where the repositioning decisions are to be given for each period. We do assume that the user intervention at the time of repositioning is negligible as in the other static studies. On the other hand, we consider a dynamic planning horizon but assume that the demand information does not change throughout the planning horizon as opposed to the literature on dynamic repositioning studies.

In this study, we also present a detailed analysis of the impact of repositioning on key performance indicators in dockless bike-sharing systems. To the best of our knowledge, this research is the first that analyzes the outcomes of optimal repositioning processes in sharing systems for operating companies.

Chapter 3

Problem Setting and Mathematical Formulation

The decision-maker in our problem setting is the operator of a dockless bike-sharing system. Given the demand, the decision-maker needs to determine the best repositioning strategy that would result in maximizing their profit during the planning horizon.

To model this problem, the service area needs to be divided into several geographical zones. Similarly, the planning horizon must be divided into several time periods (e.g., hours in a day) to determine the time of the repositioning actions. There is a given demand for the number of vehicles (bikes) required to travel from one zone to another within each period. The demand between each zone pair in each period is assumed to be known. To be able to meet a period's demand from a zone to another, there should be at least one available vehicle in the origin zone at the beginning of that period. Instead of keeping inventory in each location at the beginning of each period, the availability can be ensured by repositioning the vehicles during the planning horizon.

We propose a network flow model to find the optimal repositioning strategy in the given service area during the planning horizon. The objective of the proposed model is to maximize the total profit for the sharing system. The aim of the model is to decide on which demand to meet and how to best reposition the vehicles to meet this demand.

There is a known revenue from satisfying a unit demand from one zone to another. The bike-sharing companies usually charge their customers based on the time spend on the vehicles during a trip. To estimate the revenue we assume that the customers use the vehicles of the bike-sharing system only for direct transportation where they use the shortest paths to arrive at their destination without any intermediary stops. The revenue can then be estimated by multiplying the network distance from the centroid of the origin zone to the destination zone by a coefficient that accounts for the total revenue per distance.

There is a cost incurred for repositioning the vehicles. In practice, the repositioning of bikes is done through vans or trucks that can carry multiple bikes at a time. Hence, the cost of repositioning is assumed to be dependent on the distance between the zones and the average fuel consumption per distance, but it is not dependent on the number of vehicles to be repositioned. Moreover, there is a fixed cost of using or deploying a vehicle to the system. This fixed cost is assumed to be dependent on the purchasing cost of the vehicles and prorated for the duration of the planning horizon.

We assume that the vans or trucks that are used to reposition the vehicles have sufficient capacity and all of the vehicles in the system are distributed from a hypothetical *depot* to the service zones at the beginning of the planning horizon to account for the distribution of the vehicles to the system after overnight charging.

The following notation is used for the parameters that are required to model the problem:

S	Set of zones (0 denotes the depot).
T	Set of periods (0 denotes the beginning of the planning horizon).
d_{ijt}	Demand from zone $i \in S$ to zone $j \in S$ in time period $t \in T$.
p_{ij}	Revenue of satisfying a demand from zone $i \in S$ to zone $j \in S$.
c_{ij}	Cost of repositioning from zone $i \in S$ to $j \in S$.
f	Fixed cost of deploying a vehicle.

We define the following decision variables for the mathematical formulation:

$$
\begin{aligned}
x_{ijt} &= \text{Number of vehicles assigned to the demand from zone } i \in S \text{ to zone } j \in S \\
&\quad \text{in time period } t \in T. \\
y_{ijt} &= \text{Number of vehicles repositioned from zone } i \in S \text{ to zone } j \in S \\
&\quad \text{in time period } t \in T. \\
z_{ijt} &= \begin{cases} 1, & \text{If any repositioning is done from zone } i \in S \text{ to zone } j \in S \text{ in period } t \in T, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

The problem is formulated as follows:

$$
\max \quad \sum_{i \in S} \sum_{j \in S} \sum_{t \in T} p_{ij} x_{ijt} - \sum_{i \in S} \sum_{j \in S} \sum_{t \in T} c_{ij} z_{ijt} - \sum_{j \in S} f y_{0j1} \tag{3.1}
$$

$$
\begin{aligned}
\text{s.t.} \quad & x_{ijt} \leq d_{ijt} && i,j \in S, t \in T && (3.2)\\
& \sum_{j\in S} x_{ijt} + \sum_{j\in S} y_{ijt} = \sum_{s\in S} x_{si(t-1)} + \sum_{s\in S} y_{si(t-1)} && i \in S, t \in T && (3.3)\\
& y_{ijt} \leq M z_{ijt} && i,j \in S, t \in T && (3.4)\\
& x_{ijt}, y_{ijt} \geq 0 && i,j \in S, t \in T && (3.5)\\
& z_{ijt} \in \{0,1\} && i,j \in S, t \in T && (3.6)
\end{aligned}
$$

The objective function (3.1) maximizes the total profit of the bike-sharing system. The total profit is calculated by subtracting the repositioning and fixed costs of vehicles from the total revenue which is gained from satisfying the customer trips. The first term in the objective function calculates the total revenue obtained from satisfying the demand, the second term calculates the cost of repositioning, and the last term calculates the total cost of deploying vehicles in the system at the beginning of the planning horizon.

Constraint (3.2) links the demand data with the decision variables x_{ijt}. This constraint ensures that the number of vehicles assigned to trips between the zones can not exceed the potential demand. Constraint (3.3) is the flow balance constraint that matches the flow between periods and zones. For each period and zone, the left-hand side of this constraint calculates the total number of vehicles available at that zone to either satisfy the demand of that period or reposition and equals it to the total number of vehicles that arrived at that zone from the previous period either through satisfying the previous period's demand or by relocation. This constraint allows a vehicle to stay idle in a zone during the planning horizon through the use of y_{iit} variables which may or may not have an associated cost in the objective function depending on the decision maker's preferences.

If there is at least one vehicle that needs to be repositioned from zone i to zone j at period t, Constraint (3.4) ensures that the corresponding binary variable z takes on the value of 1. Here, M denotes a large-enough number, for example, the number of vehicles in the system. The cost of repositioning is calculated in the objective function through the use of the binary z variables, where the cost denotes the dedication of one high-capacity vehicle to consolidate the relocation of all vehicles from zone i to zone j at period t. Lastly, Constraint (3.5) and (3.6) are the domain constraints of the model.

If the total number of vehicles in the system is desired to be limited, Constraint (3.7) below can be added to the model. This constraint sets the total number of vehicles to deploy at the beginning of the planning horizon to a predetermined parameter, denoted by V.

$$\sum_{j \in S} y_{0j1} \leq V \tag{3.7}$$

For any reason, if the decision-maker wants to avoid relocation, the repositioning decision variables y in the model can be set to zero. By setting these variables to zero, one can analyze the effects of repositioning in the system. We conduct such analyses that are detailed in the next section.

Chapter 4

Computational Analyses

This section presents the computational experiments with the proposed mathematical model developed to optimize the repositioning decisions for bike-sharing systems. The model was run on a computer that has AMD Ryzen 5 5600X CPU, with 4.6 GHz and 16.00 GB of RAM, running Windows 11 operating system. CPLEX 20.1 was used as a mixed-integer linear programming solver.

The proposed mathematical model is analyzed under real-world bike-sharing data collected from three major cities: New York, Toronto, and Vancouver. The next section details information on these three data sets.

4.1 Data Sets

The annual ride data of station-based bike-sharing systems in New York, Toronto, and Vancouver are collected from the websites of the operating companies which are Citi Bike (2022), Bike Share Toronto (2022), and Mobi Bikes (2022), respectively. All of these data sets are open source and contain each trip's origin, destination, date, and duration. A screenshot from each of these data sets is presented in Appendix A. As a way of comparison, Table 4.1 below lists the number of stations and the annual number of trips in the bike-sharing systems of each of these cities during the year 2021. As can be observed from this table, the New York bike-sharing system is the largest and the most established among the three cities, whereas Vancouver is the smallest.

Table 4.1: Comparison of the data sets.

City	**Number of Stations**	**Number of Annual Trips in 2021**
New York	1522	29,247,005
Toronto	615	3,612,588
Vancouver	197	736,914

Although the focus is on the dockless bike-sharing systems in this thesis, the performance of the model is tested on station-based sharing system data. The main reason is the availability of the data only for station-based systems. Currently, none of these cities have operating dockless sharing systems. In particular, the launch of dockless bike-sharing systems is delayed in New York because of the pandemic (NYC 311, 2022).

To scale the problem size to be able to implement and solve the optimization model, we clustered stations in each city into 100 zones. In particular, we implemented a k-means clustering algorithm, where $k = 100$ (Hartigan and Wong, 1979). In our preliminary analysis, we observed that having a higher number of zones makes it difficult to solve the optimization model whereas by using fewer zones we might be unnecessarily agrregating the data. The k-means clustering algorithm is coded in Python and provided in Appendix B. After implementing the k-means clustering algorithm, the shortest network distances between the centroids of the clusters are calculated by using GraphHopper Routing API (GraphHopper, 2022). The demand data for our experimentation is generated by considering the demand between these clustered zones and it is assumed that distances between the centroids of the zones is a reasonable estimate for the travel distances of the demand from one zone to another. Figure 4.1, Figure 4.2, and Figure 4.3 depict the coordinates of the stations and centroids of the 100 clusters in New York, Toronto, and Vancouver, respectively. The blue points represent the locations of the existing stations, whereas the red points show the centroids of the clusters.

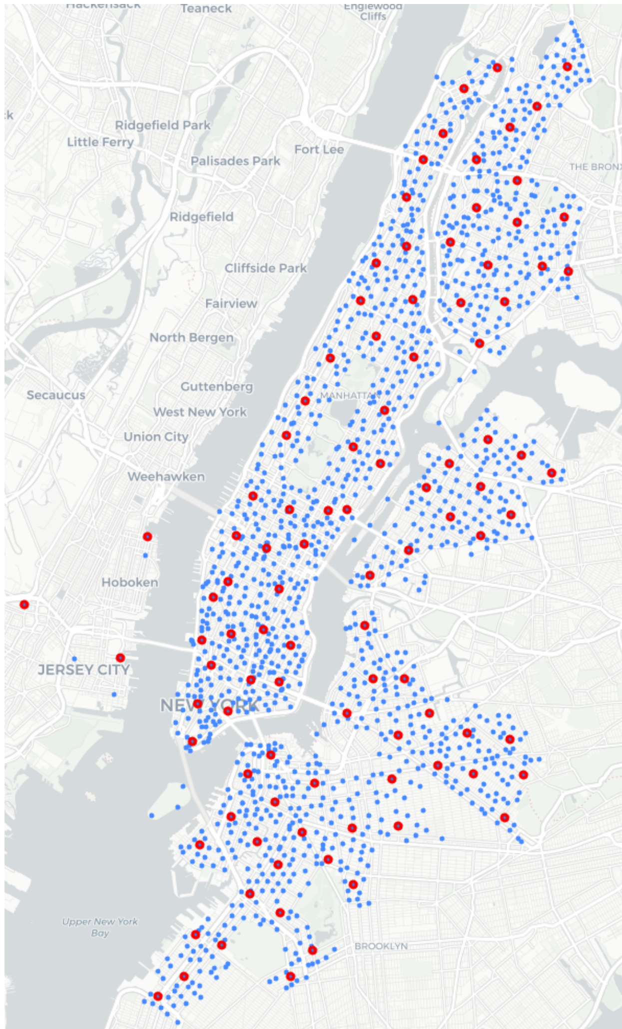


Figure 4.1: New York data - 1522 stations (blue), 100 zone centres (red).

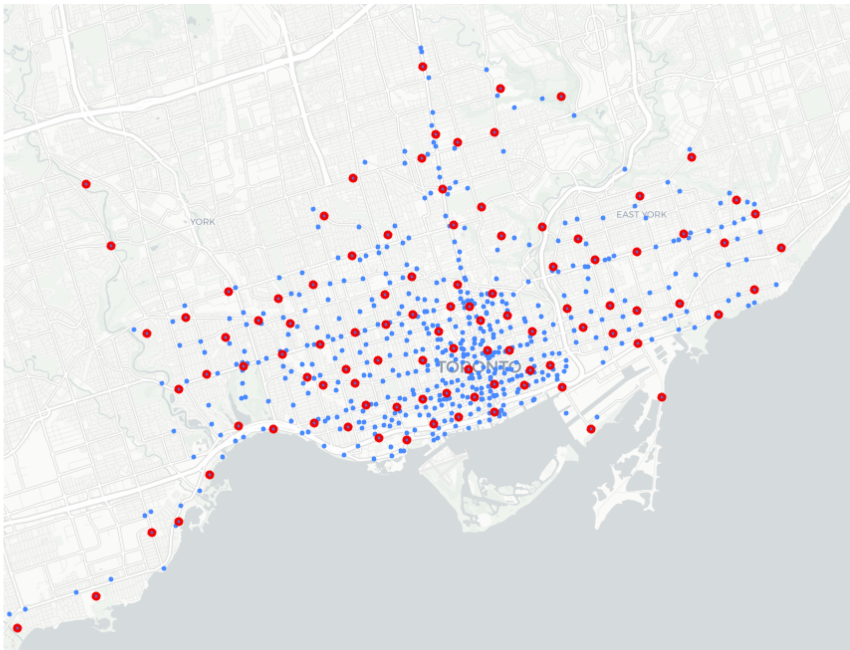


Figure 4.2: Toronto data - 615 stations (blue), 100 zone centers (red).

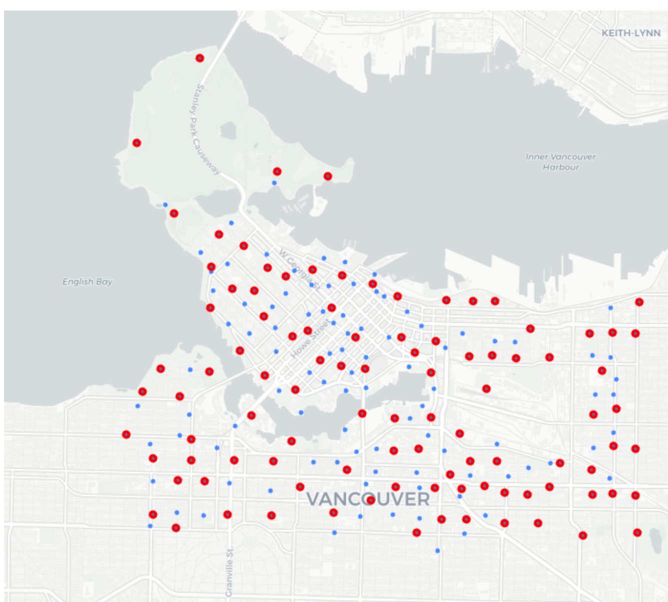


Figure 4.3: Vancouver data - 197 stations (blue), 100 zone centers(red)

For our computational experiments, we consider the planning horizon as a day since recharging of the vehicles is usually done overnight. From the data sets, we calculated the average trip durations in the year 2021 as 18 minutes 2 seconds for New York, 17 minutes 15 seconds for Toronto, and 22 minutes 3 seconds for Vancouver. Moreover, 98% of all trips were less than an hour in all data sets. Accordingly, we assumed that the demand for each trip length is for a period of less than one hour and that every trip starts and ends in the same period. Consequently, the day is divided into an hour-long 24 equal periods, where each hour corresponds to a period. In particular, the first period starts at 00:00 AM and ends at 00:59 AM, the second period starts at 01:00 AM and ends at 01:59 AM, and so on and so forth until the last period, which starts at 11:00 PM and ends at 11:59 PM.

We do not consider the demand from one period to another, however, note that this depends on the start time of each period. If the periods start, say, 5 minutes past the hour, the demand set might be different. We conducted a sensitivity analysis with our data sets to observe the effect of changes in the starting time of the periods in a day on demand. In particular, we generated different demand data by changing the start times of the 1-hour periods for 60 possibilities, corresponding to each minute of the hour, for each day in the month of August 2021 with New York City data. We then calculated the standard deviation of the total demand as 0.003 among all of these options. Since the demand data does not change significantly when we change the start time of the 1-hour periods, we continued our analysis by starting the periods at full hours.

In dockless sharing systems, the operating companies tend to implement an overnight charging process for electric vehicles. In this strategy, all vehicles in the system are collected at the end of a day for recharge and redistributed to service zones from charging centers at the beginning of the day. Accordingly, it is assumed that all of the vehicles in the existing bike-sharing systems in New York, Toronto, and Vancouver are distributed to the service zones at the beginning of each day. The cost of distributing the vehicles from a depot (node 0) to service zones is assumed to be negligible as this cost is incurred by the overnight chargers who are already compensated.

Table 4.2 summarizes the values of the parameters used in our computational experiments. For each city, the demand between the zones is obtained from the data that is provided by the operator companies. To estimate the revenue that is gained from the trips of the customers, Citi Bike Pricing (2022) is used, where the company is charging approximately \$11 per hour. The average speed of a bike is taken between 10-15 km/hr in the urban areas (Jensen et al., 2010). As a result, the revenue (p_{ij}) is estimated to be \$1 per km. The average fuel consumption is gathered from the official United States Department of Energy's website (U.S. Department of Energy, 2022). The fixed cost per day (f), on the other hand, is calculated by dividing the average bike price by 365 days where we assume that the life cycle of a bicycle is one year in bike-sharing systems. Based on Bicycle Universe (2022), the average price of a road bike varies between \$350-700. The operator companies tend to use basic bikes so it is assumed that the price of a bike is close to the lower bound of this range.

Table 4.2: Values of the parameters.

Parameter	**Value and Source**
$\|S\|$	100
$\|T\|$	24
d_{ijt}	From Citi Bike (2022), Bike Share Toronto (2022), Mobi Bikes (2022)
p_{ij}	\$1/km (Citi Bike Pricing, 2022)
c_{ij}	\$0.1/km (U.S. Department of Energy, 2022)
f	\$1/day (Bicycle Universe, 2022)

4.2 Results and Insights

We considered the data between February 2021 and February 2022 for a period of one year and a month (13 months). Within this period, we solved the model with data from the 10^{th} and 20^{th} day of each month for each of the three cities to have random representative days. All instances were solved to optimality by using CPLEX 20.1. The average run time was 3 minutes 17 seconds per instance. All results are provided in the table presented in Appendix C.

For each instance, the optimal solution obtained from the model provides us with the total demand to be served and the total number of vehicles that need to be repositioned between each pair of zones in each period that maximize the profit. After finding the optimal repositioning actions, we analyze the effect of repositioning on the fulfilled demand, number of required vehicles, and utilization rates of the vehicles which is illustrated in Figure 4.4. The results of these analyses are presented in the respective subsections below.

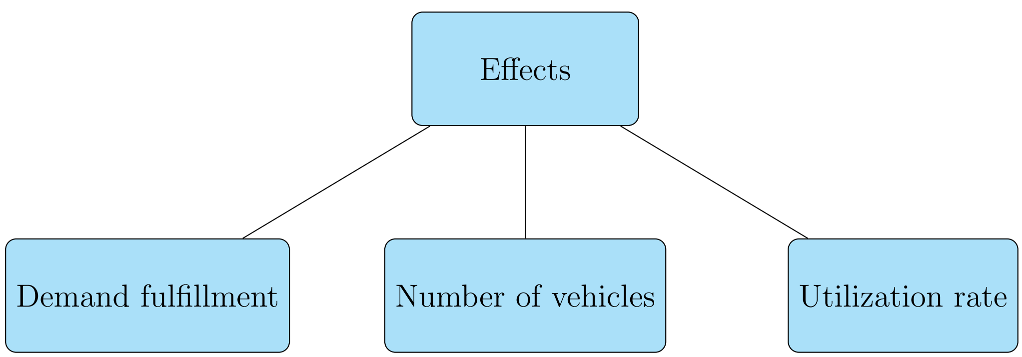


Figure 4.4: Effects of repositioning

4.2.1 Demand fulfillment

We first analyze the effect of repositioning on fulfilled demand. For this analysis, as a first step, we solve the optimization model by forcing to meet all demands, i.e., we set Constraint (3.2) to equality and solve our model. From these optimal solutions, we can calculate the minimum number of vehicles required to meet all demand in the system. This value is equal to $\sum_{j \in S} y_{0j1}$, as this summation gives exactly the number of vehicles that need to be dispatched in the system from the depot at the beginning of the planning horizon.

In the second step, we solve the optimization model by setting the repositioning decision variables to zero and limiting the number of vehicles in the system to the value that we found in the previous step by setting this value to the right-hand side of Constraint (3.7). For each instance, we then calculate the amount of demand that can be satisfied using the same number of vehicles, without any repositioning, to maximize profit. Note that since we do not allow for repositioning in this second step, we can cover less demand with the same number of available vehicles. Figure 4.5 presents the daily percentages of potential lost demand with each of the data sets when repositioning is not allowed in the system.

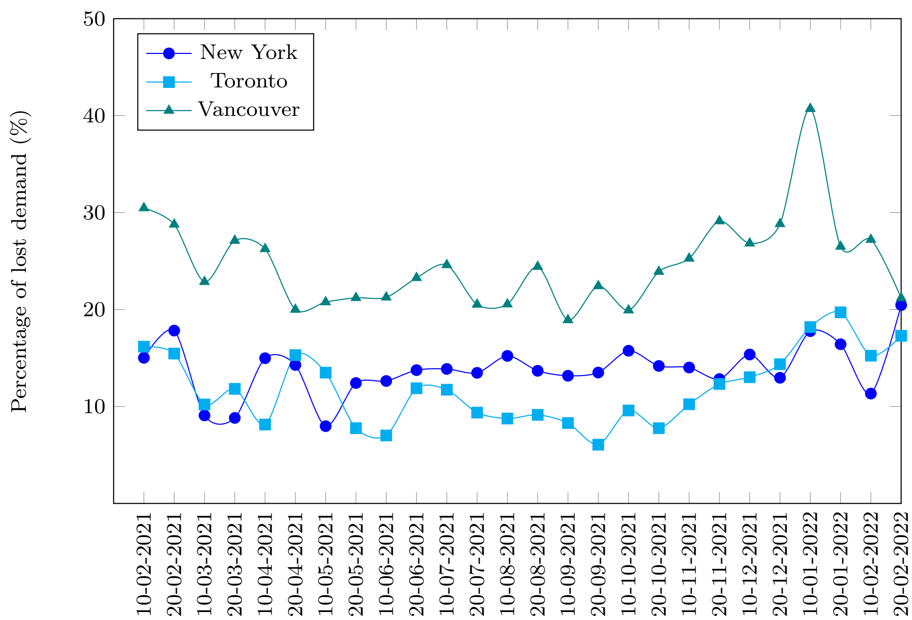


Figure 4.5: Effect of repositioning on the percentage of lost demand.

Among these three cities, the highest impact of repositioning on the lost demand is observed in Vancouver. If repositioning is not allowed in Vancouver, the system is able to meet 24.73% less demand on average. The average percentages of lost demand in New York and Toronto are 13.86% and 11.86%, respectively.

Among the days that are presented in Figure 4.5, the highest percentage of lost demand is observed as 40.71% (January 20, 2022, Vancouver) and the lowest is 6.06% (September 20, 2021, Toronto).

4.2.2 Number of vehicles

We next investigate the number of vehicles needed to fulfill the entire demand with and without repositioning. We solve the optimization model by setting the demand Constraint (3.2) to equality and find the minimum number of vehicles required to meet all demand in the system as detailed in Section 4.2.1. Subsequently, we set the repositioning variables to zero and resolve the model to calculate the number of vehicles needed to fulfill the entire demand when repositioning is not allowed in the system. The difference in the numbers of vehicles to satisfy all demand with and without repositioning in New York, Toronto, and Vancouver are shown, respectively, in Figure 4.6, Figure 4.7, and Figure 4.8.

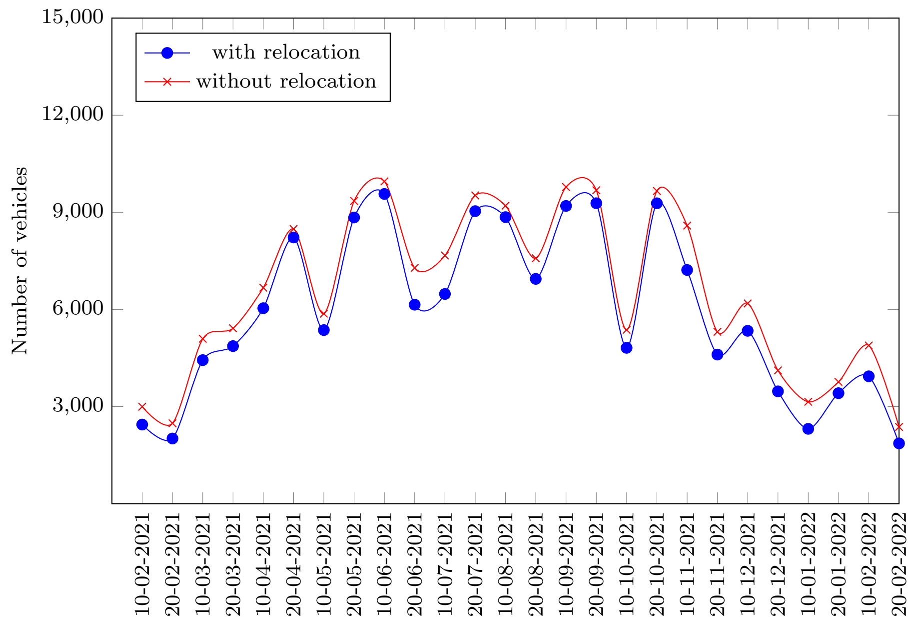


Figure 4.6: Number of vehicles to meet all demand in New York.

In New York, it is possible to fulfill the same amount of demand with 638 fewer vehicles per day on average with repositioning. The highest impact of repositioning is observed on November 10, 2021 (Figure 4.6). To fulfill the same amount of demand on this day, the bike-sharing system needs 1372 more vehicles when no repositioning is allowed.

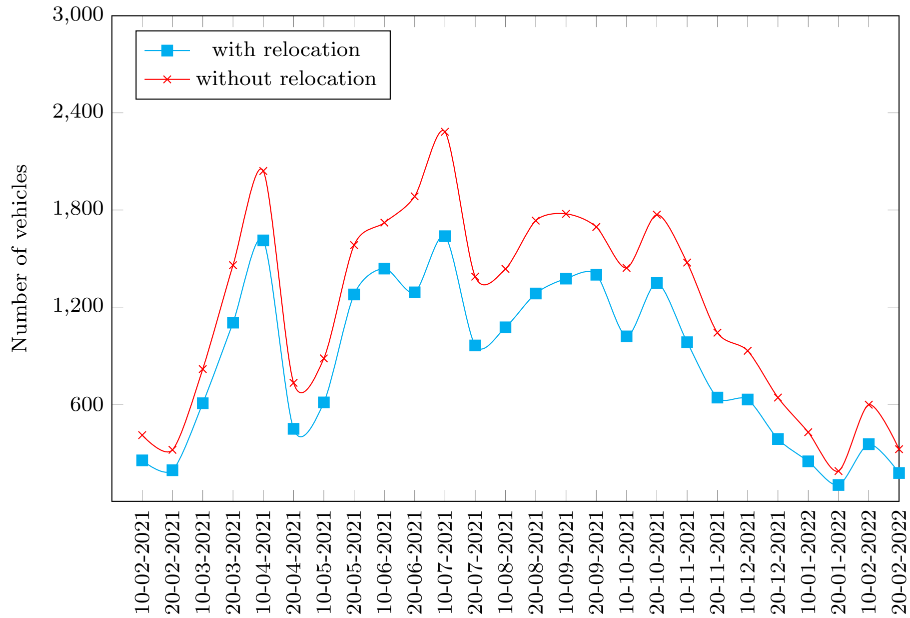


Figure 4.7: Number of vehicles to meet all demand in Toronto.

The bike-sharing system in Toronto requires 317 fewer vehicles per day on average to satisfy the whole demand if repositioning is allowed. On July 10, 2021, the effect of repositioning on the number of required vehicles is observed to be the highest, where the entire demand could be fulfilled by utilizing 645 fewer vehicles (Figure 4.7).

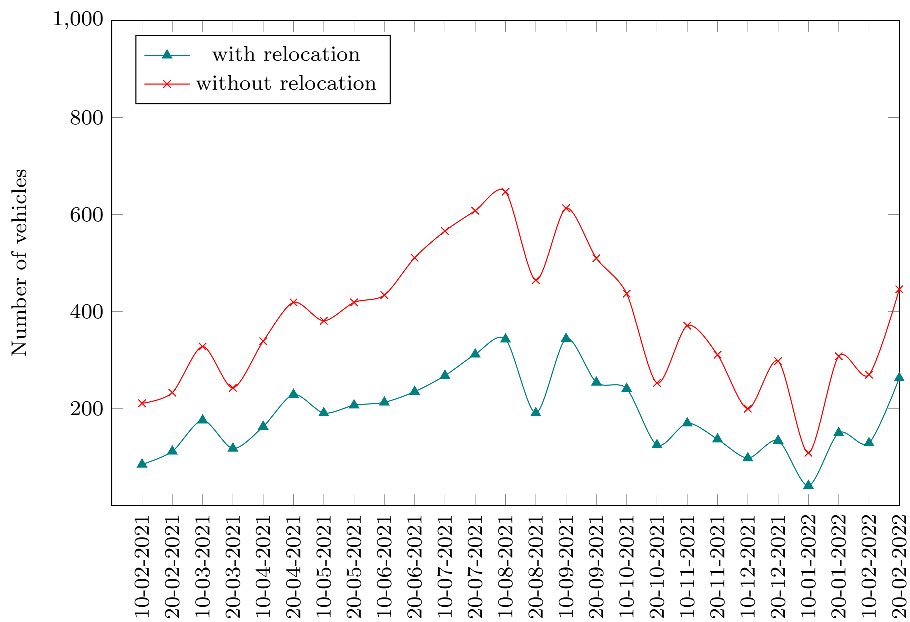


Figure 4.8: Number of vehicles to meet all demand in Vancouver.

In Vancouver, when repositioning is allowed, the bike-sharing system is able to fulfill the same amount of demand with 192 fewer vehicles per day on average. The difference in the number of vehicles needed to fulfill the demand with and without repositioning varies between 68 and 304. The smallest difference is observed on January 10, 2022, whereas the highest is observed on August 10, 2022 (Figure 4.8).

As a result of the analyses using the data sets from three different cities, it can be concluded that repositioning is able to reduce the fleet size significantly. The highest impact is observed in Vancouver; as demonstrated in Figure 4.8, it is possible to cover the same daily demand with approximately 50% fewer vehicles with repositioning. Although the difference in the required number of vehicles to fulfill the demand is highest in New York, it is observed that the percentage decrease in the required number of vehicles decreases with the increase in demand.

4.2.3 Utilization rate

In this section, we analyze the effect of repositioning on the utilization rates of the vehicles used in the three bike-sharing systems. To be able to calculate the utilization percentages of the vehicles, each vehicle that is assigned to a trip in a period is considered to be busy or fully utilized throughout the whole period. The average utilization rate of the vehicles is then calculated as follows:

$$\text{Utilization Rate} = \frac{\sum\limits_{i \in S} \sum\limits_{j \in S} \sum\limits_{t \in T} x_{ijt}}{hv} \times 100 \tag{4.1}$$

In this equation, h denotes the total number of periods, which is 24 in our setting, and v is the total number of vehicles needed to fulfill the entire demand. The nominator calculates the total demand that is covered by the vehicles in a day, which also represents the total number of vehicle assignments to the trips. The denominator, on the other hand, is the total available vehicle-periods (or vehicle-hours) in a day.

The average daily utilization rates of the vehicles are calculated for all of the daily instances in each cities. After that, the repositioning actions are disabled and the instances are solved again to find the utilization rates in case of no repositioning. The differences between the utilization rates of the vehicles in each of the cities are depicted in Figure 4.9, Figure 4.10, and Figure 4.11.

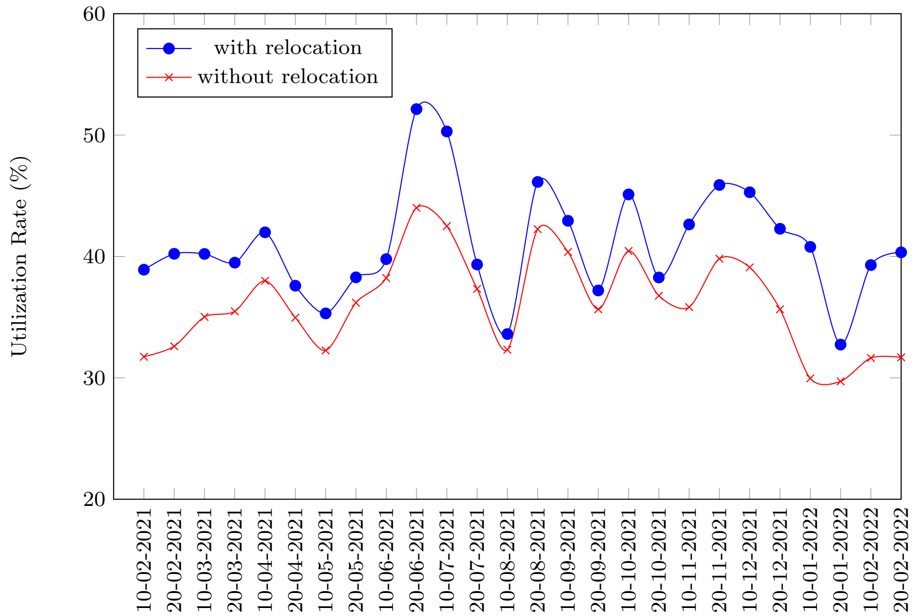


Figure 4.9: Effect of repositioning on vehicle utilization rates in New York.

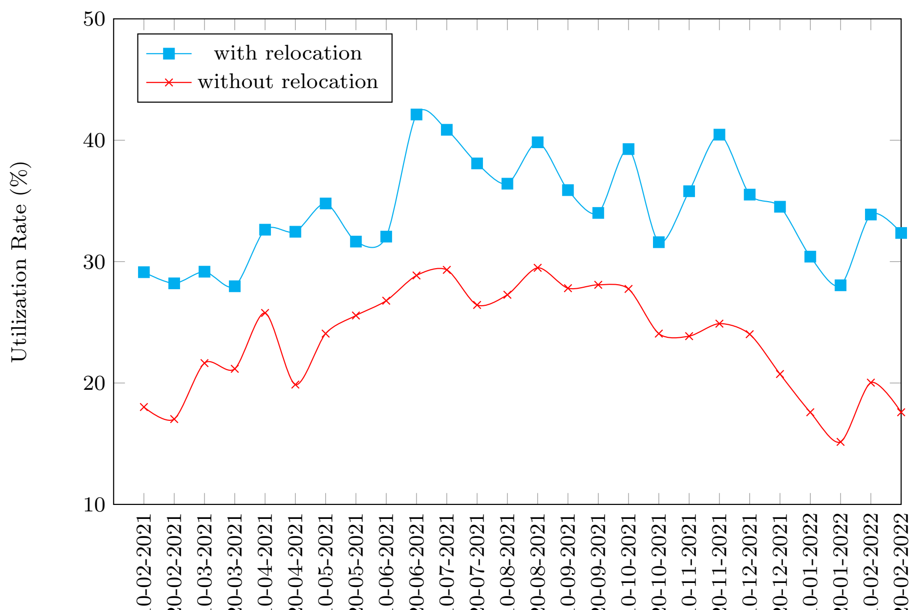


Figure 4.10: Effect of repositioning on vehicle utilization rates in Toronto.

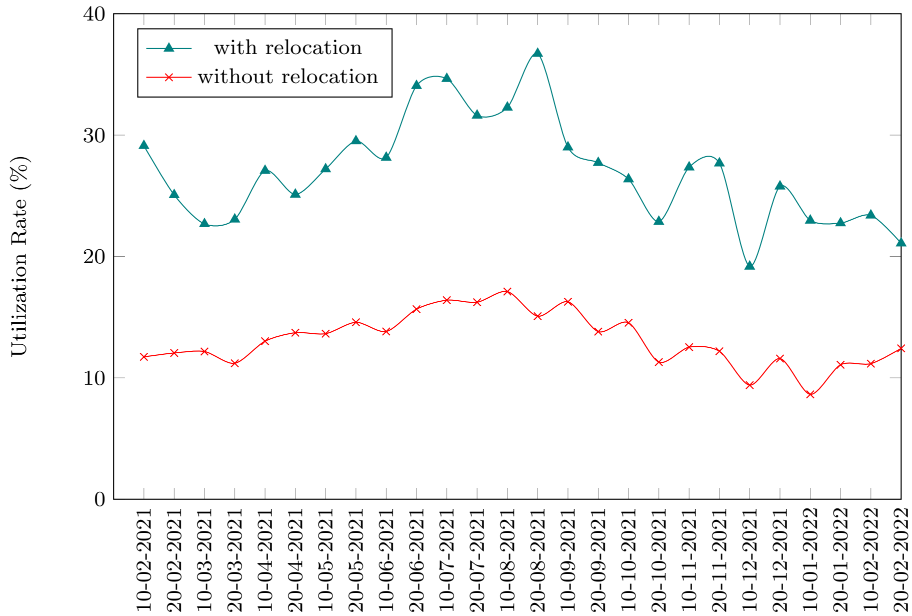


Figure 4.11: Effect of repositioning on vehicle utilization rates in Vancouver.

In New York, the repositioning process increase vehicle utilization from 36.31% to 41.10% on average. The highest increase is on January 10, 2022 (Figure 4.9). On this day the average utilization rate of the vehicles is increased from 29.96% to 40.79% with repositioning.

The average daily utilization rates of the vehicles in Toronto are calculated as 34.12% with repositioning and 23.57% without repositioning. Hence, repositioning increases the average utilization rate by 10.55% in the instances presented in Figure 4.10. The highest increase is observed on November 20, 2021, when the average utilization rate is increased from 24.89% to 40.46%.

In Vancouver, the average daily vehicle utilization rate is more than doubled with repositioning. The average daily utilization rate is 27.01% with repositioning, whereas it is 13.13% without repositioning. The highest improvement is seen on August 20, 2021, when the utilization rate is increased from 15.08% to 36.71% (Figure 4.11). Although the bikes in New York are utilized the most, the highest impact of repositioning on the utilization rates is observed in Vancouver.

In addition to the daily utilization rate comparisons, the effect of repositioning is also analyzed for all days throughout the month of August 2021, which is one of the busiest months of the year. The proposed model is solved with the demand data for every day of this month and average daily utilization rates are calculated for each of the cities. Figure 4.12 illustrates the average daily utilization rates in New York, Toronto, and Vancouver with repositioning.

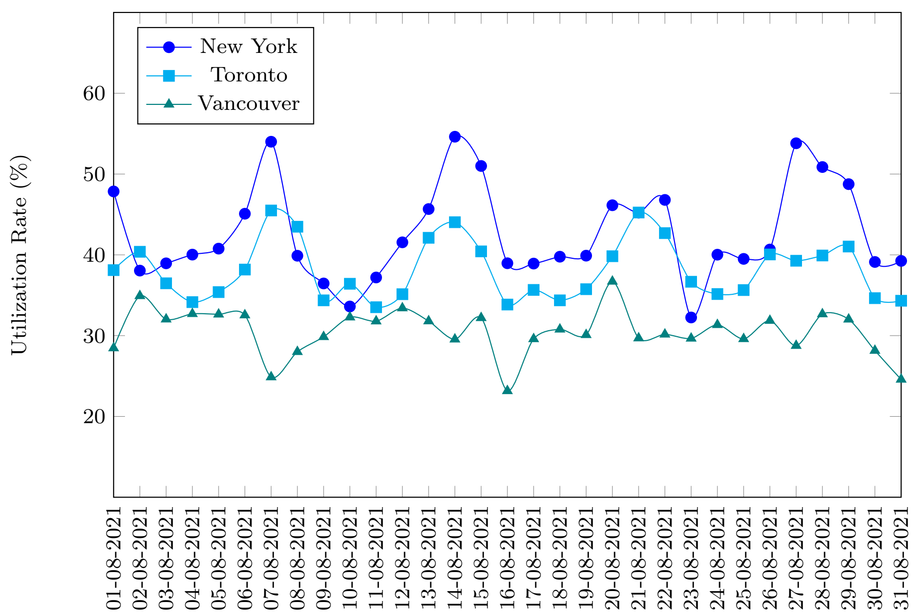


Figure 4.12: Utilization rates in August 2021 with repositioning.

It can be observed from Figure 4.12 that New York and Toronto have similar trends in the sense that the vehicles employed in the bike-sharing systems of these two cities are utilized more during the weekends as compared to weekdays. On the other hand, although vehicles in Vancouver also tend to be utilized more during the weekends, the difference between the utilization rates during weekdays and weekends is not as significant compared with New York and Toronto.

In addition to the average daily utilization rates, we also calculated the hourly utilization rates of the vehicles. Since each period corresponds to an hour in our analysis, we used the following equation to calculate the hourly utilization rates of the vehicles in the system:

$$\text{Hourly Utilization Rate} = \frac{\sum\limits_{i \in S} \sum\limits_{j \in S} x_{ijt}}{v} \times 100 \tag{4.2}$$

For the hourly utilization rate analysis, we selected the days that have the highest demand among all the previously tested daily instances from February 2021 to February 2022. The days that have the highest demand are September 10, July 10, and August 10 for New York, Toronto, and Vancouver, respectively. The results are depicted in Figure 4.13.

Observe from Figure 4.13 that the utilization rates in all cities peak during rush hour, in particular between 4 PM and 8 PM, and the vehicles are barely utilized just before sunrise. These less busy times of the day will be used by overnight chargers to collect and redistribute the vehicles in the system for the following day's demand or for regular maintenance.

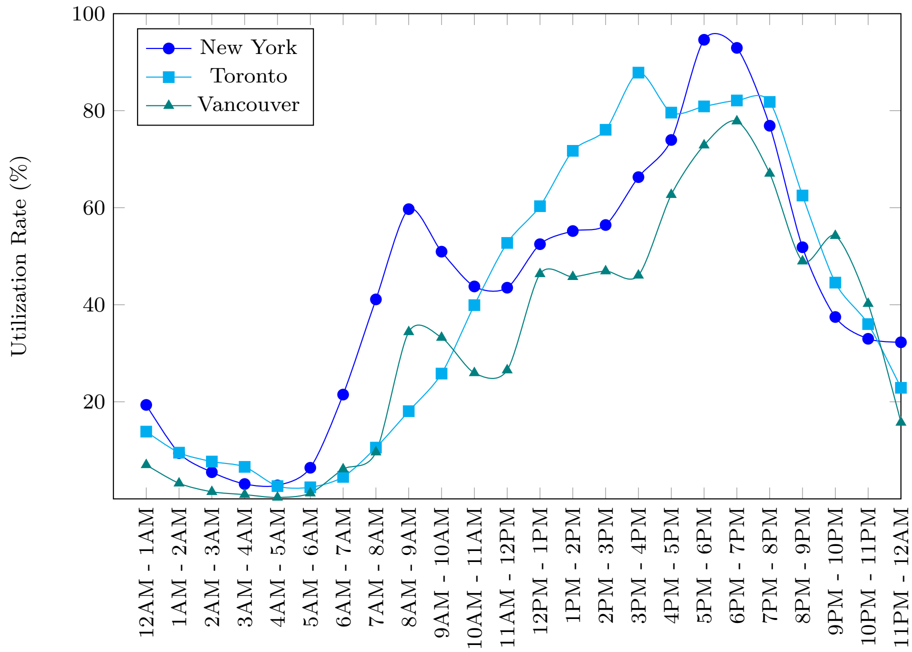


Figure 4.13: Utilization rates on the busiest days in 2021 with repositioning.

4.2.4 Visualization of repositioning

In this section, we visually represent the repositioning actions on a map and analyze the consequences of these actions on the busiest days of each of the cities in 2021.

In each of the figures (Figure 4.14, Figure 4.15, Figure 4.16), we represent the optimal repositioning actions with lines. In particular, there is a line between the origin and destination of carrier trucks that are conducting repositioning of the vehicles in each of the three bike-sharing systems. For New York, in Figure 4.14, the color of the line is red if the destination point is located relatively on the south of the origin (the longitude of the destination is smaller than the longitude of the origin), and the color of the line is blue if the destination is located relatively in the north of the origin (the longitude of the destination is larger than the longitude of the origin). For Toronto and Vancouver, in Figure 4.15 and Figure 4.16, on the other hand, the color of the line is red if the destination is located on the west of the origin (the latitude of the destination is smaller than the latitude of the origin), and the color of the line is blue if the destination is located on the east of the origin (the latitude of the destination is bigger than the latitude of the origin).

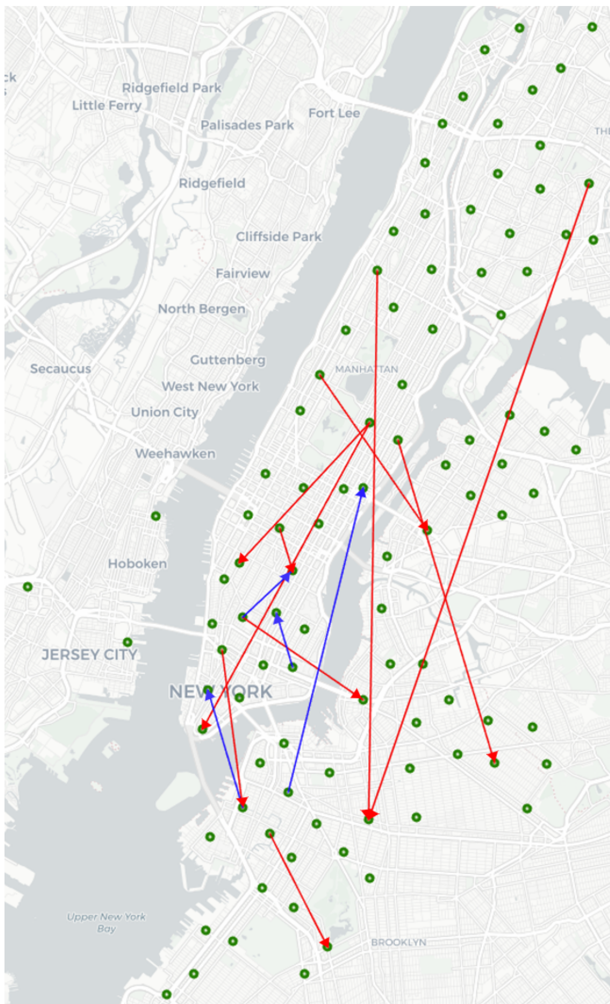


Figure 4.14: Visualization of relocations in New York on September 10, 2021.

[Figure 4.14](#) presents the optimal repositioning decisions in New York on September 10, 2021, which has the highest demand among all the tested daily instances. Throughout this day, the vehicles in the system are repositioned 14 times. These repositioning actions increased the average vehicle utilization rate from 40.37% to 42.94%. Along with the increase in the utilization rates, the number of vehicles required to fulfill the demand on September 10, 2021, decreased from 9783 to 9198 as a result of repositioning. In other words, it is possible to fulfill the demand on this day with 585 fewer vehicles if these optimal repositioning decisions are implemented. On the other hand, the system is able to meet 15.75% less demand if no repositioning is allowed.

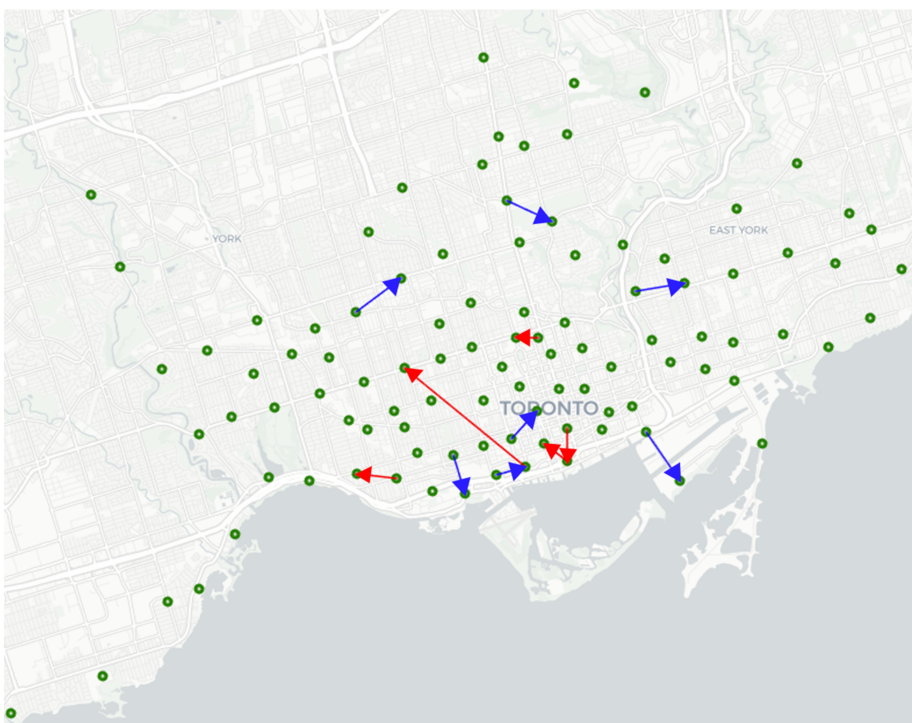


Figure 4.15: Visualization of relocations in Toronto on July 10, 2021.

As illustrated in Figure 4.15, there are 12 repositioning actions in Toronto on July 10, 2021 which is the busiest day among the tested daily instances. The average vehicle utilization on this day increased from 29.32% to 40.86% as a result of the proposed repositioning actions. Moreover, as in New York, repositioning reduced the number of vehicles required to serve the demand by 645 vehicles. If repositioning is not implemented, the system is able to meet 11.73% less demand.

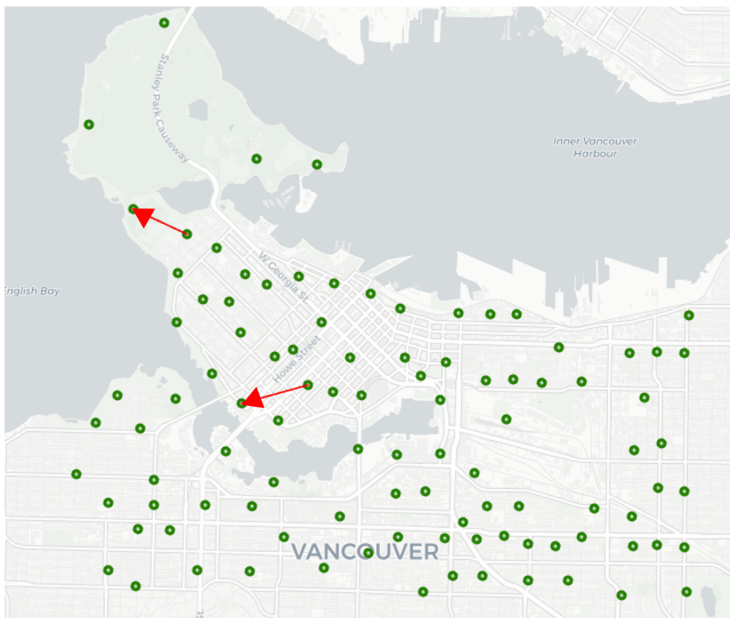


Figure 4.16: Visualization of relocations in Vancouver on August 10, 2021.

There are only two repositioning actions in Vancouver on August 10, 2021, in the optimal solution. Figure 4.16 illustrates the optimal repositioning actions on this day. Although the vehicles are repositioned only two times, these repositioning actions increased the average vehicle utilization rate from 17.11% to 32.28%. Another key contribution of the repositioning processes on this day is the decrease in the required number of vehicles in order to fulfill the demand. The total number of vehicles required to meet the entire demand decreases from 647 to 343 if the optimal repositioning strategy is implemented.

On the other hand, without repositioning, the bike-sharing system can fulfill 20.55% less demand with the same number of vehicles.

4.2.5 Demand uncertainty

In the previous sections, the effect of repositioning is analyzed by implementing the proposed mathematical model using the real-world bike-sharing data of New York, Toronto, and Vancouver on 26 different daily instances that correspond to the 10th and 20th days of each month between February 2021 and 2022. In this section, we investigate the effect of repositioning under uncertainty by generating random demand scenarios.

To generate these random demand scenarios, we first find the minimum and maximum demand between each pair of zones in each period from the 26 daily instances. For each zone pair and period, we then generate a random demand value from the interval between these minimum and maximum values using a uniform distribution. We generate ten different demand matrices in this manner, referred to as scenarios, and solve the model under each of these random data scenarios.

Table 4.3 compares the effect of repositioning on real and random data based on the three performance indicators defined in Section 4.2.1, Section 4.2.2, and Section 4.2.3. The "Min" and "Max" columns presented under real data represent the minimum and maximum effect of repositioning among the 26 daily instances that correspond to the values depicted in Figure 4.6 – Figure 4.11. The same performance indicators are calculated with the random demand scenarios and listed under the respective columns.

Table 4.3: Comparison of the effect of repositioning on real vs. random data.

		Real Data		Random Data Scenarios									
		Min	Max	1	2	3	4	5	6	7	8	9	10
% of lost demand	New York	7.97	20.46	11.06	11.12	10.44	11.23	11.28	11.54	10.15	13.22	11.78	12.31
	Toronto	6.06	19.71	7.68	7.49	7.69	7.83	7.11	7.24	7.88	8.14	7.92	8.01
	Vancouver	19.93	40.71	22.40	21.54	22.15	21.20	22.95	21.81	22.04	22.14	23.42	22.97
% decrease in the number of vehicles	New York	3.09	26.56	13.62	17.25	12.64	12.95	15.06	12.15	13.77	15.64	14.52	13.43
	Toronto	16.49	45.99	24.92	24.18	25.75	24.18	22.73	23.49	24.04	25.45	24.70	25.19
	Vancouver	41.03	62.39	51.14	48.61	49.98	47.12	50.55	50.49	51.82	49.18	54.33	52.26
% increase in the utilization rate	New York	1.28	10.83	6.80	8.42	6.32	6.43	7.44	7.15	7.86	8.29	8.01	7.91
	Toronto	5.29	15.57	11.80	11.37	11.91	11.52	10.46	10.99	11.54	12.03	11.81	11.99
	Vancouver	8.65	21.63	16.51	15.58	15.83	14.88	16.53	15.96	16.12	15.43	16.45	16.02

Observe from Table 4.3 that all values obtained under all random scenarios fall within the interval between the minimum and maximum values obtained using the real data for each city. Accordingly, we expect repositioning to have a similar positive effect on each of these performance measures under demand uncertainty.

Chapter 5

Conclusions

This research analyzes the effect of optimal repositioning on dockless bike-sharing systems. As a first step, we formulate a multi-period mixed-integer linear model to find the optimal repositioning actions in bike-sharing systems. This model maximizes the total profit by finding the optimal number of vehicles to be repositioned between each location in the service area and its time period.

The developed model is implemented using real-world bike-sharing data from New York, Toronto, and Vancouver on 26 different daily instances that correspond to the 10th and 20th days of each month between February 2021 and 2022. The optimal repositioning schedules are obtained by solving the proposed model with data from these days for each city. Considering the obtained optimal repositioning actions, the effects of repositioning are analyzed on the fulfilled demand, number of required vehicles, and utilization rates of the vehicles in each city.

Our results show that a significant loss of demand can occur if repositioning is not allowed in bike-sharing systems. The percentage of this potential demand loss varies between

6% to 41% in our test instances. Moreover, our results demonstrate that repositioning can considerably reduce the required fleet size. The average number of vehicles to fulfill the same amount of demand with and without repositioning can be up to 63% fewer. Repositioning also has a substantial effect on the daily utilization rates of the vehicles employed in bike-sharing systems. We observe that repositioning can increase the daily utilization rates of the vehicles up to 21%.

Bike-sharing is a capital-intensive industry and operator companies need to bear significant expenditures to start up a business (Tian et al., 2021). Our results demonstrate that an effective repositioning strategy can decrease the capital investment requirements as it allows operating companies to serve their customers with fewer vehicles. This can in turn facilitate more companies to enter the sector and, subsequently, customers can benefit from the competition in the market that will be reflected in prices.

We assume in this research that the carrier trucks that are responsible for the repositioning of the vehicles have sufficient capacity. For future research, the capacities of these carrier trucks can be incorporated into the mathematical model. To address this modification, the binary variables in our model need to be replaced by integer decision variables. As a consequence, the problem sizes will get bigger and there might be a need for customized solution methodologies, such as decomposition algorithms, to be able to solve realistic-sized instances.

Currently, the output of the proposed model can readily be used as an input for a vehicle routing model to optimize the routes of the carrier trucks. Instead of a sequential approach, another future research direction would be to simultaneously find the optimal routes of these carrier trucks that would allow them to deliver multiple origin-destination points and do more than one repositioning in a single period within a pick-up and delivery type of a vehicle routing problem. Such an extension can provide a better evaluation of repositioning costs, whereas our analysis uses an upper bound for calculating these costs.

References

Baldacci, R., Hadjiconstantinou, E., and Mingozzi, A. (2004). An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation. *Operations Research*, 52(5):723–738.

Bektas, T. (2006). The multiple traveling salesman problem: An overview of formulations and solution procedures. *Omega*, 34(3):209–219.

Bicycle Universe (2022). How much does a bike cost? https://bicycleuniverse.com/how-much-does-a-bike-cost/. Accessed: March 2022.

Bike Share Toronto (2022). Toronto system data. https://ckan0.cf.opendata.inter.prod-toronto.ca/tr/dataset/bike-share-toronto-ridership-data. Accessed: March 2022.

Bird (2022). Bird flyer users' manual. https://help.bird.co/hc/en-us/articles/360015872951-Getting-Started-FAQ-s. Accessed: July 20222.

Bruck, B. P., Cruz, F., Iori, M., and Subramanian, A. (2019). The static bike sharing rebalancing problem with forbidden temporary operations. *Transportation Science*, 53(3):882–896.

Chen, Y. W., Cheng, C. Y., Li, S. F., and Yu, C. H. (2018). Location optimization for multiple types of charging stations for electric scooters. *Applied Soft Computing Journal*, 67:519–528.

Chen, Z., van Lierop, D., and Ettema, D. (2020). Dockless bike-sharing systems: what are the implications? *Transport Reviews*, 40(3):333–353.

Cheng, Y., Wang, J., and Wang, Y. (2021). A user-based bike rebalancing strategy for free-floating bike sharing systems: A bidding model. *Transportation Research Part E: Logistics and Transportation Review*, 154:102438.

Chiariotti, F., Pielli, C., Zanella, A., and Zorzi, M. (2018). A dynamic approach to rebalancing bike-sharing systems. *Sensors*, 18(2):1–22.

Cipriano, M., Colomba, L., and Garza, P. (2021). A data-driven based dynamic rebalancing methodology for bike sharing systems. *Applied Sciences*, 11(15):6967.

Citi Bike (2022). New York system data. https://ride.citibikenyc.com/system-data. Accessed: March 2022.

Citi Bike Pricing (2022). Citi Bike single ride pricing. https://citibikenyc.com/pricing/single-ride. Accessed: March 2022.

Contardo, C., Morency, C., and Rousseau, L.-M. (2012). Balancing a dynamic public bike-sharing system. *Technical Report*, CIRRELT-2012-09, Montreal, Canada.

Dell'Amico, M., Hadjicostantinou, E., Iori, M., and Novellani, S. (2014). The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, 45:7–19.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4):41–56.

Erdoğan, G., Laporte, G., and Wolfler Calvo, R. (2014). The static bicycle relocation problem with demand intervals. *European Journal of Operational Research*, 238(2):451–457.

Garcia-Palomares, J., G. J. L. M. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography*, 35:235–246.

GraphHopper (2022). The GraphHopper Directions API: Route planning for your application. https://www.graphhopper.com. Accessed: September 2018.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 28(1):100–108.

Hernández-Pérez, H. and Salazar-González, J.-J. (2007). The one-commodity pickup-and-delivery traveling salesman problem: Inequalities and algorithms. *Networks*, 50(4):258–272.

Jensen, P., Rouquier, J. B., Ovtracht, N., and Robardet, C. (2010). Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation Research Part D: Transport and Environment*, 15(8):522–524.

Ji, S., C.-C. H. L. and Jordan, D. (2014). Electric bike sharing: simulation of user demand and system availability. *Journal of Clean Production*, 85:250–257.

Jin, X. and Tong, D. (2020). Station-free bike rebalancing analysis: Scale, modeling, and computational challenges. *ISPRS International Journal of Geo-Information*, 9(11).

Kabak, M., Erbaş, M., Çetinkaya, C., and Özceylan, E. (2018). A GIS-based MCDM approach for the evaluation of bike-share stations. *Journal of Cleaner Production*, 201:49–60.

Lazarus, J., Pourquier, J. C., Feng, F., Hammel, H., and Shaheen, S. (2020). Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete – A case study of San Francisco. *Journal of Transport Geography*, 84:102620.

Li, X., Zhang, Y., Du, M., and Yang, J. (2019). Social factors influencing the choice of bicycle: Difference analysis among private Bike, public bike sharing and free-floating bike sharing in Kunming, China. *KSCE Journal of Civil Engineering*, 23:2339–2348.

Lime (2022). Limebike announces plans for dockless bike sharing launch in scottsdale. https://www.li.me/en-ca/blog/limebike-plans-dockless-bike-sharing-launch-scottsdale. Accessed: July 2022.

Lin, J. and Yang, T. (2011). Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E*, 47:284–294.

Liu, J., Li, Q., Qu, M., Chen, W., Yang, J., Xiong, H., Zhong, H., and Fu, Y. (2016). Station site optimization in bike sharing systems. *Proceedings - IEEE International Conference on Data Mining, ICDM*, January:883–888.

Liu, Y., Szeto, W. Y., and Ho, S. C. (2018). A static free-floating bike repositioning problem with multiple heterogeneous vehicles, multiple depots, and multiple visits. *Transportation Research Part C: Emerging Technologies*, 92:208–242.

Médard de Chardon, C., Caruso, G., and Thomas, I. (2016). Bike-share rebalancing strategies, patterns, and purpose. *Journal of Transport Geography*, 55:22–39.

Mete, S., C.-Z. and Ozceylan, E. (2018). Location and coverage analysis of bike-sharing stations in university campus. *Business Systems Research*, 9(2):80–95.

Mobi Bikes (2022). Vancouver system data. https://www.mobibikes.ca/en/system-data. Accessed: March 2022.

Muren, Li, H.-M. S. W. J. Z. L. and Zhiping, D. (2020). Balanced maximal covering location problem and its application in bike-sharing. *International Journal of Production Economics*, 223.

Neijmeijer, N., Schulte, F., Tierney, K., Polinder, H., and Negenborn, R. R. (2020). Dynamic pricing for user-based rebalancing in free-floating vehicle sharing: A real-world case. 12433 LNCS:443–456.

NYC 311 (2022). Dockless bike share. https://portal.311.nyc.gov/article/?kanumber=KA-02688. Accessed: September 2021.

Osorio, J., Lei, C., and Ouyang, Y. (2021). Optimal rebalancing and on-board charging of shared electric scooters. *Transportation Research Part B: Methodological*, 147:197–219.

Pal, A. and Zhang, Y. (2017). Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies*, 80:92–116.

Reiss, S. and Bogenberger, K. (2017). A relocation strategy for Munich's bike sharing system: Combining an operator-based and a user-based scheme. *Transportation Research Procedia*, 22:105–114.

Shen, Y., Z.-X. and Zhao, J. (2018). Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation*, 12(9).

Styr&Ställ (2022). Bike sharing in göteborg, mölndal. https://styrochstall.se/en/. Accessed: July 2022.

Tian, X., Wang, Y., and Shi, Y. (2021). Research on risk management and profitability of bike sharing business model. *Proceedings of the 2021 International Conference on Financial Management and Economic Transition (FMET 2021)*, 190:110–116.

U.S. Department of Energy (2022). The new fuel economy label. https://www.fueleconomy.gov/feg/label/learn-more-gasoline-label.shtml. Accessed: March 2022.

Vox (2018). Inside how a scooter-sharing startup navigates san francisco. https://www.vox.com/2018/5/20/17328110/lime-limebike-dockless-scooters-san-francisco-charge-drop-off. Accessed: July 20222.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., and Ratti, C. (2019). Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems*, 75:184–203.

Younes, H., Zou, Z., Wu, J., and Baiocchi, G. (2020). Comparing the temporal determinants of dockless scooter-share and station-based bike-share in Washington, D.C. *Transportation Research Part A: Policy and Practice*, 134:308–320.

APPENDICES

Appendix A

Data Sets

	A	B	C	D	E	F	G	H	I	J	K
1	ride_id	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng
2	FB6B89D05B67EBED	2021-08-24 15:59	2021-08-24 16:42	Broadway & E 21 St	6098.1	Central Park North & Adam Clayton Powell Blvd	7617.07	40.7398884	-73.9895859	40.799484	-73.955613
3	E13DA3E30CEF8DFC	2021-08-18 13:12	2021-08-18 13:21	E 13 St & 2 Ave	5820.08	Henry St & Grand St	5294.04	40.7315394	-73.9853024	40.714211	-73.981095
4	56617490AB8AE69C	2021-08-17 14:31	2021-08-17 14:35	E 95 St & 3 Ave	7365.13	E 84 St & Park Ave	7243.04	40.7849032	-73.950503	40.7786269	-73.9577207
5	CA908B271C7D6663	2021-08-11 10:00	2021-08-11 10:31	Madison Ave & E 82 St	7188.13	E 84 St & Park Ave	7243.04	40.7781314	-73.960694	40.7786269	-73.9577207
6	3E170CE1F4FE179D	2021-08-12 19:28	2021-08-12 19:48	E 74 St & 1 Ave	6953.08	E 84 St & Park Ave	7243.04	40.7689738	-73.9548227	40.7786269	-73.9577207
7	09A4D1869E7817D2	2021-08-15 1:44	2021-08-15 1:56	Newtown Ave & 23 St	7026.08	35 St & 21 Ave	7170.04	40.7713615	-73.9246145	40.776745	-73.906558
8	3D505A8B988E4A9B	2021-08-31 0:02	2021-08-31 0:24	Fulton St & Irving Pl	4263.12	Willoughby Ave & Tompkins Ave	4665.02	40.68186	-73.959432	40.694254	-73.9462692
9	74D40040E5BD33D4	2021-08-29 18:44	2021-08-29 18:48	28 St & 41 Ave	6462.19	38 Ave & 29 St	6538.11	40.751047	-73.93797	40.75473	-73.93367
10	0CF01DC4DC8E746C	2021-08-09 18:31	2021-08-09 18:58	Reade St & Broadway	5247.1	Cadman Plaza E & Tillary St	4677.01	40.7145045	-74.0056279	40.6959768	-73.9901489
11	9F8A0D1BBEEA4CD4	2021-08-01 14:10	2021-08-01 14:31	West St & Chambers St	5329.03	E 20 St & Park Ave	6055.08	40.7175483	-74.0132207	40.7382743	-73.9875197
12	E80BAD050F4ACAB8	2021-08-25 18:22	2021-08-25 18:40	West St & Chambers St	5329.03	E 20 St & Park Ave	6055.08	40.7175483	-74.0132207	40.7382743	-73.9875197
13	BA51F4FB155C73C4	2021-08-04 16:00	2021-08-04 16:10	Cleveland Pl & Spring St	5492.05	E 20 St & Park Ave	6055.08	40.7221038	-73.997249	40.7382743	-73.9875197
14	17048C616EE83002	2021-08-05 19:01	2021-08-05 19:35	Cleveland Pl & Spring St	5492.05	Willoughby Ave & Tompkins Ave	4665.02	40.7221038	-73.997249	40.694254	-73.9462692
15	04E2287CD0DC9B91	2021-08-05 8:00	2021-08-05 8:10	Pershing Square South	6432.08	E 20 St & Park Ave	6055.08	40.751581	-73.97791	40.7382743	-73.9875197
16	5A77B437FCCFAC15	2021-08-05 19:24	2021-08-05 20:05	Cleveland Pl & Spring St	5492.05	Cadman Plaza E & Tillary St	4677.01	40.7221038	-73.997249	40.6959768	-73.9901489
17	F6C4B204DF82AD8D	2021-08-27 17:31	2021-08-27 17:39	W 4 St & 7 Ave S	5880.02	E 20 St & Park Ave	6055.08	40.7340114	-74.0029388	40.7382743	-73.9875197
18	DB450381ADB381E7	2021-08-27 9:03	2021-08-27 9:11	McKibbin St & Manhattan Ave	4996.08	Willoughby Ave & Tompkins Ave	4665.02	40.7051092	-73.9440728	40.694254	-73.9462692
19	4838C1A37B042371	2021-08-12 18:52	2021-08-12 19:19	E 25 St & 2 Ave	6046.02	E 20 St & Park Ave	6055.08	40.739126	-73.9797378	40.7382743	-73.9875197
20	9AC41DCCB5BB3BDC	2021-08-15 17:34	2021-08-15 18:04	Forsyth St & Canal St	5270.07	Cadman Plaza E & Tillary St	4677.01	40.7158155	-73.9942237	40.6959768	-73.9901489
21	846431DD3E4F85A2	2021-08-28 21:29	2021-08-28 21:44	Riverside Dr & W 153 St	8108.02	W 145 St & Amsterdam Ave	7997.08	40.832164	-73.949702	40.825244	-73.947257
22	7500DC08A95A9546	2021-08-29 23:17	2021-08-29 23:38	4 Ave & E 12 St	5788.15	Graham Ave & Withers St	5403.04	40.732647	-73.99011	40.7169811	-73.9448592

Figure A.1: A screen shot from the Citi Bike (2022) data set.

	A	B	C	D	E	F	G	H	I
1	Trip Id	Trip Duration	Start Station Id	Start Time	Start Station Name	End Station Id	End Time	End Station Name	Bike Id
2	11015571	195	7032	04/01/2021 00:01	Augusta Ave / Dundas St W	7049	04/01/2021 00:04	Queen St W / Portland St	656
3	11015572	938	7168	04/01/2021 00:01	Queens Quay / Yonge St	7508	04/01/2021 00:17	Berkeley St / Dundas St E - SMART	5272
4	11015573	1145	7012	04/01/2021 00:03	Elizabeth St / Edward St (Bus Terminal)	7012	04/01/2021 00:23	Elizabeth St / Edward St (Bus Terminal)	3253
5	11015574	1061	7037	04/01/2021 00:04	Bathurst St / Dundas St W	7079	04/01/2021 00:22	McGill St / Church St	3233
6	11015575	460	7198	04/01/2021 00:07	Queen St W / Cowan Ave	7662	04/01/2021 00:15	Beaty Ave / Queen St W	1381
7	11015576	643	7311	04/01/2021 00:08	Sherbourne St / Isabella St	7551	04/01/2021 00:18	The Esplanade / Hahn Pl	3450
8	11015577	1486	7075	04/01/2021 00:08	Queens Quay W / Dan Leckie Way	7344	04/01/2021 00:33	Cherry Beach	5991
9	11015579	1078	7248	04/01/2021 00:11	Baldwin Ave / Spadina Ave - SMART	7264	04/01/2021 00:29	Bloor St E / Huntley St - SMART	1498
10	11015580	829	7042	04/01/2021 00:18	Sherbourne St / Wellesley St E	7118	04/01/2021 00:32	King St W / Bay St (East Side)	3050
11	11015581	492	7402	04/01/2021 00:19	Wellington St W / Bathurst St	7052	04/01/2021 00:27	Wellington St W / Bay St	3848
12	11015582	3191	7468	04/01/2021 00:22	Front St / Simcoe St	7468	04/01/2021 01:15	Front St / Simcoe St	4072
13	11015583	3161	7468	04/01/2021 00:22	Front St / Simcoe St	7468	04/01/2021 01:15	Front St / Simcoe St	5596
14	11015584	335	7025	04/01/2021 00:28	Ted Rogers Way / Bloor St E	7121	04/01/2021 00:33	Jarvis St / Dundas St E	6596
15	11015585	677	7599	04/01/2021 00:29	Richmond St W / York St	7255	04/01/2021 00:40	Stewart St / Bathurst St - SMART	4236
16	11015586	191	7264	04/01/2021 00:30	Bloor St E / Huntley St - SMART	7530	04/01/2021 00:33	Sherbourne St N / Elm Ave	6378

Figure A.2: A screen shot from the Bike Share Toronto (2022) data set.

	A	B	C	D	E	F
1	Departure	Return	Departure station	Return station	Covered distance (m)	Duration (sec.)
2	2021-08-01 0:00	2021-08-01 0:00	0026 Beatty & Robson	0191 7th & Laurel	2178	760
3	2021-08-01 0:00	2021-08-01 0:00	0137 Beach & Seymour	0028 Davie & Beach	1984	549
4	2021-08-01 0:00	2021-08-01 0:00	0026 Beatty & Robson	0026 Beatty & Robson	0	57
5	2021-08-01 0:00	2021-08-01 0:00	0026 Beatty & Robson	0191 7th & Laurel	2203	856
6	2021-08-01 0:00	2021-08-01 0:00	0222 Adanac & McLean	0215 Princess & Union	1413	369
7	2021-08-01 0:00	2021-08-01 0:00	0229 Keefer & Princess	0053 Keefer & Abbott	1937	497
8	2021-08-01 0:00	2021-08-01 0:00	0230 Alexander & Railway	0230 Alexander & Railway	0	17
9	2021-08-01 0:00	2021-08-01 0:00	0273 Victoria & 4th	0206 8th & Scotia	3365	766
10	2021-08-01 0:00	2021-08-01 0:00	0002 Burrard Station (Melville & Dunsmuir)	0266 St Catherines & 7th	4688	1484
11	2021-08-01 0:00	2021-08-01 0:00	0258 13th & St George	0212 Union & Dunlevy	2846	560
12	2021-08-01 0:00	2021-08-01 1:00	0030 Abbott & Cordova	0105 Stanley Park - Totem Poles	4698	1243
13	2021-08-01 0:00	2021-08-01 1:00	0208 Arbutus Greenway & Broadway	0137 Beach & Seymour	3083	730
14	2021-08-01 0:00	2021-08-01 1:00	0148 Creekside Park North	0129 Richards & Robson	2176	1681
15	2021-08-01 0:00	2021-08-01 1:00	0036 Bute & Robson	0138 Richards & Helmcken	2031	885
16	2021-08-01 0:00	2021-08-01 1:00	0148 Creekside Park North	0129 Richards & Robson	2042	1551

Figure A.3: A screen shot from the Mobi Bikes (2022) data set.

Appendix B

Python Code for k-means Clustering Algorithm

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import geopy
import folium
from sklearn import cluster
import scipy
import webbrowser
data = pd.read_excel(data file)
city = "cityname"
locator = geopy.geocoders.Nominatim(user_agent="MyCoder")
location = locator.geocode(city)
print(location)
location = [location.latitude, location.longitude]
print("[lang, long]:", location)
map_ = folium.Map(location=location, tiles="cartodbpositron",
                  zoom_start=12)
```

```
x, y = "Long", "Lang"
data.apply(lambda row: folium.CircleMarker(location=[row[y],row[x]],
fill=True, color =
'red', radius=0.001).add_to(map_), axis=1)
map_.save('mymap.html')
webbrowser.open_new_tab('mymap.html')
X = data[["Long","Lang"]]
max_k = 40
distortions = []
for i in range(1, max_k+1):
    if len(X) >= i:
       model = cluster.KMeans(n_clusters=i, init='k-means++',
       max_iter=300,
       n_init=10, random_state=0,tol=0.0005)
       model.fit(X)
       distortions.append(model.inertia_)

k = [i*100 for i in np.diff(distortions,2)].index(min(
[i*100 for i in np.diff(distortions,2)]))
fig, ax = plt.subplots()
ax.plot(range(1, len(distortions)+1), distortions)
ax.axvline(k, ls='--', color="red", label="k = "+str(k))
ax.set(title='The Elbow Method', xlabel='Number of clusters',
ylabel="Distortion")
ax.legend()
ax.grid(True)
plt.show()
k = 100
model = cluster.KMeans(n_clusters=k, init='k-means++')
X = data[["Long","Lang"]]
dtf_X = X.copy()
dtf_X["cluster"] = model.fit_predict(X)
## find real centroids
closest, distances = scipy.cluster.vq.vq(model.cluster_centers_,
                    dtf_X.drop("cluster", axis=1).values)
dtf_X["centroids"] = 0
for i in closest:
    dtf_X["centroids"].iloc[i] = 1
```

```
data[["cluster","centroids"]] = dtf_X[["cluster","centroids"]]
plt.show
fig, ax = plt.subplots()
sns.scatterplot(x="Long", y="Lang", data=data,
                palette=sns.color_palette("bright",k),hue='cluster',
                 size="centroids", size_order=[1,0],
                legend="brief", ax=ax).set_title('Clustering 20')
th_centroids = model.cluster_centers_
ax.scatter(th_centroids[:,0], th_centroids[:,1], s=50, c='black',
           marker="x")
plt.show()
data.to_excel(result file)
```

Appendix C

Solutions

Table C.1: Solutions of New York instances.

Date	Total demand	Demand met without repositioning	Number of vehicles to meet demand with repositioning	Number of vehicles to meet demand without repositioning	Average utilization rate with repositioning	Average utilization rate without repositioning
2021-02-10	22823	19396	2444	2996	38.91%	31.74%
2021-02-20	19432	15968	2013	2484	40.22%	32.60%
2021-03-10	42804	38916	4435	5093	40.21%	35.02%
2021-03-20	46139	42069	4868	5419	39.49%	35.48%
2021-04-10	60856	51741	6039	6673	41.99%	38.00%
2021-04-20	71199	61031	8224	8486	37.59%	34.96%
2021-05-10	45444	41820	5363	5870	35.31%	32.26%
2021-05-20	81227	71137	8841	9352	38.28%	36.19%
2021-06-10	91338	79800	9568	9956	39.78%	38.23%
2021-06-20	76926	66345	6147	7283	52.14%	44.01%
2021-07-10	78208	67359	6478	7667	50.30%	42.50%
2021-07-20	85315	73824	9037	9523	39.34%	37.33%
2021-08-10	71397	60533	8852	9204	33.61%	32.32%
2021-08-20	76890	66368	6944	7582	46.14%	42.25%
2021-09-10	94789	82303	9198	9783	42.94%	40.37%
2021-09-20	82858	71675	9281	9683	37.20%	35.65%
2021-10-10	52119	43909	4814	5369	45.11%	40.45%
2021-10-20	85236	73153	9280	9660	38.27%	36.77%
2021-11-10	73881	63525	7219	8591	42.64%	35.83%
2021-11-20	50733	44223	4606	5308	45.89%	39.82%
2021-12-10	58044	49128	5340	6186	45.29%	39.10%
2021-12-20	35214	30651	3470	4116	42.28%	35.65%
2022-01-10	22636	18613	2312	3148	40.79%	29.96%
2022-01-20	26829	22423	3414	3762	32.74%	29.71%
2022-02-10	37106	32901	3935	4887	39.29%	31.64%
2022-02-20	18026	14338	1862	2370	40.34%	31.69%
Random Scenario 1	107573	95680	8982	10398	49.90%	43.11%
Random Scenario 2	107853	95862	9211	11131	48.79%	40.37%
Random Scenario 3	108688	97339	9061	10372	49.98%	43.66%
Random Scenario 4	107668	95578	9039	10384	49.63%	43.20%
Random Scenario 5	108114	95919	9123	10741	49.38%	41.94%

Table C.2: Solutions of Toronto instances.

Date	Total demand	Demand met without repositioning	Number of vehicles to meet demand with repositioning	Number of vehicles to meet demand without repositioning	Average utilization rate with repositioning	Average utilization rate without repositioning
2021-02-10	1769	1483	253	409	29.13%	18.02%
2021-02-20	1300	1099	192	318	28.21%	17.03%
2021-03-10	4243	3810	606	817	29.17%	21.64%
2021-03-20	7412	6536	1104	1459	27.97%	21.17%
2021-04-10	12623	11595	1612	2041	32.63%	25.77%
2021-04-20	3490	2956	448	732	32.46%	19.87%
2021-05-10	5101	4413	611	883	34.79%	24.07%
2021-05-20	9709	8956	1278	1583	31.65%	25.56%
2021-06-10	11066	10290	1438	1722	32.06%	26.78%
2021-06-20	13049	11499	1291	1884	42.12%	28.86%
2021-07-10	16064	14179	1638	2283	40.86%	29.32%
2021-07-20	8804	7979	963	1388	38.09%	26.43%
2021-08-10	9397	8574	1075	1436	36.42%	27.27%
2021-08-20	12274	11152	1284	1734	39.83%	29.49%
2021-09-10	11852	10870	1376	1776	35.89%	27.81%
2021-09-20	11427	10734	1400	1695	34.01%	28.09%
2021-10-10	9603	8682	1019	1442	39.27%	27.75%
2021-10-20	10231	9437	1349	1771	31.60%	24.07%
2021-11-10	8445	7581	983	1475	35.80%	23.86%
2021-11-20	6225	5458	641	1042	40.46%	24.89%
2021-12-10	5362	4664	629	930	35.52%	24.02%
2021-12-20	3190	2732	385	641	34.52%	20.74%
2022-01-10	1803	1475	247	427	30.41%	17.59%
2022-01-20	680	546	101	187	28.05%	15.15%
2022-02-10	2870	2424	353	597	33.88%	20.03%
2022-02-20	1359	1124	175	322	32.36%	17.59%
Random Scenario 1	32926	30397	2896	3857	47.37%	35.57%
Random Scenario 2	32587	30147	2888	3809	47.01%	35.65%
Random Scenario 3	33103	30559	2982	4016	46.25%	34.34%
Random Scenario 4	32968	30386	2882	3801	47.66%	36.14%
Random Scenario 5	32790	30458	2968	3841	46.03%	35.57%

Table C.3: Solutions of Vancouver instances.

Date	Total demand	Demand met without repositioning	Number of vehicles to meet demand with repositioning	Number of vehicles to meet demand without repositioning	Average utilization rate with repositioning	Average utilization rate without repositioning
2021-02-10	594	413	85	211	29.12%	11.73%
2021-02-20	674	480	112	233	25.07%	12.05%
2021-03-10	958	739	176	328	22.68%	12.17%
2021-03-20	653	476	118	243	23.06%	11.20%
2021-04-10	1059	781	163	339	27.07%	13.02%
2021-04-20	1380	1104	229	419	25.11%	13.72%
2021-05-10	1247	988	191	381	27.20%	13.64%
2021-05-20	1466	1155	207	419	29.51%	14.58%
2021-06-10	1439	1133	213	434	28.15%	13.82%
2021-06-20	1921	1474	235	511	34.06%	15.66%
2021-07-10	2228	1680	268	566	34.64%	16.40%
2021-07-20	2368	1882	312	608	31.62%	16.23%
2021-08-10	2657	2111	343	647	32.28%	17.11%
2021-08-20	1683	1272	191	465	36.71%	15.08%
2021-09-10	2394	1941	344	613	29.00%	16.27%
2021-09-20	1690	1311	254	510	27.72%	13.81%
2021-10-10	1525	1221	241	437	26.37%	14.54%
2021-10-20	686	522	125	253	22.87%	11.30%
2021-11-10	1116	834	170	371	27.35%	12.53%
2021-11-20	910	645	137	311	27.68%	12.19%
2021-12-10	451	330	98	200	19.18%	9.40%
2021-12-20	829	590	134	298	25.78%	11.59%
2022-01-10	226	134	41	109	22.97%	8.64%
2022-01-20	819	602	150	308	22.75%	11.08%
2022-02-10	724	527	129	270	23.39%	11.17%
2022-02-20	1331	1049	263	446	21.09%	12.43%
Random Scenario 1	10591	8219	1099	2249	40.15%	23.64%
Random Scenario 2	10769	8449	1113	2166	40.32%	24.74%
Random Scenario 3	10614	8263	1117	2233	39.59%	23.76%
Random Scenario 4	10494	8269	1091	2064	40.08%	25.20%
Random Scenario 5	10643	8200	1088	2200	40.76%	24.23%