MGSC 662 Final Report

Marin Bergeron, Mengyue Yang, Chloe He, Fiona Cao

December 2024

1 Introduction

This project focuses on optimizing the bike distribution network of BIXI, a leading bike-sharing service provider dedicated to promoting sustainable urban transportation. As cities worldwide embrace eco-friendly solutions to combat congestion and carbon emissions, the operational efficiency of bike-sharing systems like BIXI becomes increasingly important. However, maintaining an optimal balance of bike availability and dock space across a network of stations presents significant logistical challenges due to fluctuating demand, resource limitations, and operational constraints.

The complexity of this problem lies in the dynamic and unpredictable nature of demand. The first question we are looking to answer is how can we make sure that there are bikes available at a given station in order to satisfy as much demand as possible. The solution to this problem would have to assure that there are enough docks for incoming bikes, as well as enough bikes for outgoing demand. The solution to this question depends on a multitude of factors as well, as demand always fluctuating throughout the day. In order to answer this question we leveraged the open-source BIXI historical data to predict future demand.

However, finding the optimal bike allocation is not enough. As the bikes are available to use at all times, and since the demand fluctuates throughout the day, the bikes need to be manually ordered by an operator. This brings up the question, what path should the operator take to redistribute the bikes? We can reformulate this as a Traveling Salesman Problem with a twist: instead of visiting locations to minimize the total travel cost, the operator must visit stations to rebalance bike availability and dock space, ensuring that each station meets projected demand. This reformulation introduces a new layer of complexity, as the operator must consider both the sequence of stations to visit and the quantity of bikes to load or unload at each stop. The dual objective of minimizing travel costs and optimizing station-level inventory creates a combinatorial optimization challenge that requires innovative solutions.

This project not only addresses the immediate logistical challenges faced by BIXI but also contributes to the broader field of urban transportation systems. The solutions developed here can serve as a blueprint for other cities and mobility service providers looking to improve the sustainability and functionality of their systems. Through this work, we aim to highlight the power of optimization in addressing complex, real-world problems.

2 Problem Description and Formulation

2.1 Station Inventory Optimization:

As stated before, the nature of the station optimization is embedded in how a dock is a free space for an incoming bike as well as a missing bike that could be used by an outgoing passenger. In addition, another component to make this problem realistic and capture the dynamics of the demand and its fluctuation. For example if during a period of one hour there is 20 incoming bikes, and 10 outgoing, and that at the start of this one hour period there was no bikes at that station, the order of the arrival of the bikes has a large influence on the solution. If all of the outgoing bike demand arrived at the start, there would be not bikes to use. This is why instead of estimating the incoming and outgoing demand over a time period, we decided to use the exact order ingoing and outgoing bikes to capture the complexity of its pattern. Since a bike trip can only be completed if there is a bike at the departure and a free dock at the destination, we assume that the importance of both incoming and outgoing demand to be equally important. Our solution will aim to maximize the demand (both incoming and outgoing) of a given station. In addition, we wanted to be able to find the optimal allocation given a specific number of bikes, as this will be a challenge that is found in the TSP problem. This small modification adds a deep layer of complexity as the model now needs to calculate the whole sequence of bike movements and their impact on demand satisfaction across all stations over time.

Parameters

- $directions[i]$: Direction of demand at time i (1 for incoming, -1 for outgoing).
- $capacities[s]$: Maximum capacity of station s (number of docks).
- n: Total number of demand events.
- $total\_bikes$: Total number of bikes available across all stations.
- $station\_indices[i]$: Station associated with demand event i.

Decision Variables

- $x[s]$: Initial number of bikes at station s.
- $y[i]$: Binary variable, $y[i] = 1$ if demand at time i is met, otherwise $y[i] = 0$.
- $z[s, i]$: Number of bikes at station s after demand event i.

Objective Function

$$\text{Maximize } \frac{1}{n} \sum_{i=1}^{n} y[i]$$

This objective function maximizes the average met demand across all events.

Constraints

1. **Initial Setup:**
$$z[s, 0] = x[s] \quad \forall s$$
This sets the initial number of bikes at each station.

2. **Bounds for Initial Bikes:**
$$0 \le x[s] \le capacities[s] \quad \forall s$$
The number of initial bikes must respect the station capacities.

3. **Non-Negativity and Capacity Limits:**
$$0 \le z[s, i] \le capacities[s] \quad \forall s,\ i$$
The number of bikes at any station after any demand event must be non-negative and not exceed the station's capacity.

4. **Outgoing Demand Satisfaction:**
$$y[i] \le z[station\_indices[i], i] \quad \forall i \text{ where } directions[i] = -1$$
Outgoing demand can only be satisfied if there are enough bikes at the station.

5. **Incoming Demand Satisfaction:**
$$y[i] \le capacities[station\_indices[i]] - z[station\_indices[i], i] \quad \forall i \text{ where } directions[i] = 1$$
Incoming demand can only be satisfied if there are enough free docks at the station.

6. **Bike Inventory Dynamics:**
$$z[s, i+1] = z[s, i] + \begin{cases} directions[i] \cdot y[i] & \text{if } station\_indices[i] = s \\ 0 & \text{otherwise} \end{cases} \quad \forall s,\ i$$
This updates the number of bikes at each station after each demand event.

7. **Total Bike Constraint:**
$$\sum_s x[s] = total\_bikes$$
The total number of bikes allocated across all stations must equal the total available bikes.

2.2 Operator bike redistribution path:

Now given that we can find the optimal bike allocation for a given set of bikes, we can find what is the best path for the operator to take in order to redistribute these bikes. As stated previously, this problem is a variant of the TSP with the addition of the redistribution factor. This will cause the operator to assure he is carrying enough bikes before being able to reach the stations in need of refilling. In addition, given a limit capacity for the operator, this new constrain will possibly cause rerouting or impossible solutions if sub-tours are necessary to gather enough bikes for a station.

Parameters

- n: Number of stations.
- $start_station$: Index of the starting station.
- $distances[i][j]$: Distance between station i and station j.
- $random_bikes[i]$: Current number of bikes at station i.
- $optimal_bikes[i]$: Target number of bikes at station i.
- $total_bikes$: Total number of bikes available for redistribution.
- $truck_capacity$: Maximum capacity of the truck used for redistribution.

Decision Variables

- $x_{i,j}$: Binary variable, $x_{i,j} = 1$ if the operator travels from station i to station j, otherwise $x_{i,j} = 0$.
- b_i: Integer variable, number of bikes carried after visiting station i.
- u_i: Continuous variable for subtour elimination (used in the MTZ formulation).

Objective Function

The objective is to minimize the total distance traveled by the operator during bike redistribution:

$$\text{Minimize} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} x_{i,j} \cdot \text{distances}[i][j]$$

where:

- $x_{i,j}$ is a binary variable indicating if the operator travels from station i to station j ($x_{i,j} = 1$ if true, otherwise 0).
- distances$[i][j]$ is the distance between station i and station j.

Constraints

1. Visit Each Station Exactly Once

$$\sum_{j=1, j \neq i}^{n} x_{i,j} = 1 \quad \forall i$$

$$\sum_{j=1, j \neq i}^{n} x_{j,i} = 1 \quad \forall i$$

Each station must be visited exactly once, ensuring a complete route.

2. Start and End at the Starting Station

$$\sum_{j=1, j\neq \text{start\_station}}^{n} x_{\text{start\_station},j} = 1$$

$$\sum_{j=1, j\neq \text{start\_station}}^{n} x_{j,\text{start\_station}} = 1$$

The operator starts and ends their route at the designated starting station.

3. Subtour Elimination

$$u_i - u_j + n \cdot x_{i,j} \leq n - 1 \quad \forall i, j \neq \text{start\_station}, i \neq j$$

These constraints (using the Miller-Tucker-Zemlin (MTZ) formulation) ensure that no subtours (smaller cycles) occur within the solution.

4. Bike Redistribution Dynamics

1. **Initial Balance at the Starting Station**:

$$b_{\text{start\_station}} = \text{random\_bikes[start\_station]} - \text{optimal\_bikes[start\_station]}$$

   The number of bikes carried initially is determined by the difference between current and target bikes at the starting station.

2. **Bike Balance Update as the Operator Travels**:

$$b_j \leq b_i + (\text{random\_bikes}[j] - \text{optimal\_bikes}[j]) + M \cdot (1 - x_{i,j}) \quad \forall i, j, i \neq j$$

$$b_j \geq b_i + (\text{random\_bikes}[j] - \text{optimal\_bikes}[j]) - M \cdot (1 - x_{i,j}) \quad \forall i, j, i \neq j$$

   Here, b_j reflects the number of bikes carried after visiting station j, accounting for redistribution and route connections.

5. Truck Capacity

$$0 \leq b_i \leq \text{total\_bikes} \quad \forall i$$

The truck must carry a non-negative number of bikes and cannot exceed its maximum capacity.

6. Binary Variables

$$x_{i,j} \in \{0, 1\} \quad \forall i, j$$

Route decisions are binary, ensuring a clear path between stations.

3 Numerical implementation

3.1 Data Preparation

Two separate datasets were used in order to generate all of the information necessary. The first being the ride history for 2023, and the second, descriptive information about the stations. The ride history data was reformatted in order to represent the incoming and outgoing demand at each station at a given time. The dataset was finalized by combining with the location of the station for each demand point, as well as its capacity. Additional preprocessing was done to filter out any missing values, and inconsistencies. Due to the high volume of the dataset we limited our study to the stations in the Plateau Mt. Royal on November 1st 2023. After analysis the distribution of demand throughout the day, we find that the majority of demand happens during the day between 7 am and midnight Figure 1. We decided to find what bike allocation is the best in order to met the daily demand between these times. In other words what is the optimal station allocation across all stations in order to satisfy the most demand between these times.

Both problems were modeled on python with the gurobipy package. A simpler version of the station inventory optimization was run initially across all stations in the plateau. This problem didn't take into account the total number of bikes constraint that was described above, and is a necessary component of the operator TSP problem. However, running the station inventory optimization problem across all 103 stations in the plateau was too expensive. This is why decided to limit our scope to the 10 stations with the most demand.

For additional interpretation, we relate our findings to monetary values by using the following assumptions. The profit of a bike ride is estimated to be 1.35 CAD. Given an estimated fuel consumption of 0.235 L/km, as well as an estimated fuel price of 1.67 CAD/L. We assume the average fuel cost to be 0.39 CAD/km.

3.2 Station inventory optimization

When relaxing the constraint about the maximum amount of bikes we have the following solution, with the total amount of bikes used = 186, and an average met demand of 97.46% across the network of 10 stations:

Station Name	Optimal Allocation	Satisfied demand (%)	Capacity	Number of demands
des Pins / St-Laurent	22	97.21	23	467
Boyer / du Mont-Royal	27	98.33	27	300
Laurier / St-Denis	25	100	37	337
Clark / Prince-Arthur	19	98.25	31	343
Marquette / du Mont-Royal	16	100	23	423
Aylmer / Prince-Arthur	11	99.07	31	325
Prince-Arthur / du Parc	20	100	35	430
du Mont-Royal / Clark	30	100	39	454
Métro Mont-Royal (Utilités publiques / Rivard)	23	100	23	420
University / Prince-Arthur	0	91.17	47	327

Table 1: Basic Table Example

We see that we are almost able to meet almost the totality of the demand across these 10 stations with no restrictions on the number of bikes. What is interesting from this solution is that we see different patterns of results for each station. For instance we see that for most stations we try to fill most of the stations up with bikes while keeping a few docks free for incoming bikes. However for station University/Prince-Arthur, no bikes are present. This result can be interpreted as there is a large amount of incoming bikes that will arrive soon at this station, in this case McGill students using bixi to arrive to class. The opposite is true for station, (Utilités publiques / Rivard), here we are maximizing the demand by allocating as many bikes as possible at the station. These results are consistent with the behavior that is found in the bixi data.

3.2.1 Sensitivity analysis:

Since this is a mixed integer problem that cannot be solved using the simplex algorithm, we had to conduct our sensitivity analysis by running the model multiple times under different conditions. Here are some of the findings we had for this model: When looking at the different solutions for a specified fixed demand in Figure 3, we find that between 0 and 186 bikes, the demand met increases with the number of bikes available. There is then a minimal decrease that seems like a plateau until 236 bikes, after that increasing the total amount of bikes reduces the average met demand. This is a good demonstration of the balance between having the just right amount of bikes and docks. Too few bikes and we are missing too much incoming demand, too many and there aren't enough docks to complete your ride.

When looking at the evolution of number of bikes at the different stations we find different patterns of results for each station, Figure 2. For instance we see that for most stations we try to fill most of the stations up with bikes while keeping a few docks free for incoming bikes. However for station University/Prince-Arthur, no bikes are present at the initial allocation, but it quickly meets high amount of incoming demand early in the day, it stays at a relative high number of bikes before dwindling to 0 at the end of the day. This result can be interpreted as there is a large amount of incoming bikes, in this case, McGill students using bixi to arrive to class and using them to leave at the end of the day. This station's sole purpose is to be able to greet incoming passengers who arrive in the morning and leave at the end of the day. We see an opposing pattern from the station Des Pins/St Laurent, as the station is losing bikes, the University/Prince-Arthur station is filling up, and the opposite is true. These stations are likely mostly used for daily commutes. From the fluctuations in number of bikes we can identify the different user patterns which match with those found in the original data, indicating successful results.

When looking at how the capacity of each stations affects the average met demand Figure 4 and 5 and 6.We see that generally speaking increasing capacity increases the demand met. However we see that once capacity is large enough, capacity is non binding. For example, increasing the capacity of station des Pins/St Laurent to 24 allows it to meet all of its demand. However we see that most stations could reduce their capacity whilst keeping their demand met. If bixi could profit by selling these excess docs they should do so till the points indicated on the plot for each station. The only station that had an increase in met demand after an increase of 5, was University/des Pins, if possible bixi should look at moving docs from stations who do not need extra and mode them to this one in order to maximize demand.

3.2.2 Limitations:

The main limitation of this model is how the model only takes into consideration a subset of the stations. Given additional computational power, we could be able to estimate over the whole network. This problem has caused another problem in how the demand is estimated. Here in our current model we do not check from where the incoming bikes came from. If they came from a station that did not have enough bike, we should not take into account this trip at the arrival. Since there was no bike to leave in the first place. This could possibly mean that the met demand was artificially increased in this model.

3.3 Operator bike redistribution path

In order to run the TSP, our model needs to first select the station the operator will loop around, that station was randomly selected to be des Pins/St-Laurent. Another parameter that needed to be addressed is what would be the initial allocation of bikes before the operator redistribute the bikes for this we chose to randomly distribute a given amount of bikes across the 10 stations in order to see how this model functions across various scenarios. Since we need the station needs to reach the optimal bike location after the operator visits it, we made it so that the first station will always have an equal or more bikes than the optimal number in order to avoid this problem. The direct distance between the stations was computed and stored inside the distance matrix for easy access.

Station	Optimal Allocation	Random Allocation	Order of Appearance in Route
des Pins / St-Laurent	22	23	1
University / Prince-Arthur	22	21	2
Aylmer / Prince-Arthur	11	8	3
Prince-Arthur / du Parc	20	23	4
Clark / Prince-Arthur	19	32	5
du Mont-Royal / Clark	30	36	6
Laurier / St-Denis	25	23	7
Marquette / du Mont-Royal	16	23	8
Boyer / du Mont-Royal	27	27	9
Métro Mont-Royal (Utilités publiques / Rivard)	23	12	10

Table 2: Optimal and Random Bike Allocations at Each Station

Here is an example of a solution found by the model Figure 7. The total distance of the tour was 7.24 km. Which corresponds to total cost for one redistribution cycle is 394 CAD, which includes both fuel and labor costs. Additionally, the optimized route reduced the total distance by 5%, saving approximately 0.019 CAD per kilometer. This comparison was made with a baseline model that went to the closest station.

3.3.1 Sensitivity analysis:

When looking at the different specifications of the total amount of bikes across the stations we find that the number of bikes doesn't change much to the solution as it is mostly about the order at which the operator travels that matters. The interesting insights we find are related to the trucks capacity, Figure 8. When the capacity reaches 60 bikes there is a great reduction in total distance traveled. With lower values the truck is forced to take detours, and reach multiple stations where the operator needs to unload bikes. This corresponds to a change in 1 km for a trip of 10 stations. If we consider that the whole bixi network across the whole of Montreal is composed of almost 1000 stations this could come up to a significant amount of savings if they invested in trucks that are able to carry at least 60 bikes.

3.3.2 Limitations:

One limitation of the TSP model is its inability to account for dynamic demand at each station across different timestamps. To make the results more realistic, future extensions could incorporate assumptions about station-specific demand variations over time or integrate model predictions based on historical data patterns. Additionally, the TSP model represents station locations using two-dimensional longitudinal and latitudinal coordinates. However, it only considers straight-line distances between locations, which is unrealistic for real-world scenarios. In practice, travel involves following road networks with multiple possible paths between points. This oversimplification could result in preliminary cost savings appearing smaller than they would be when considering actual travel routes and road conditions. Additionally the objective function could be made more accurate by making the price increase as the number of bikes are being carried, since the fuel consumption would go up.

4 Problem extensions

- Building on the TSP analysis, which explored how changes in truck capacity and random bike redistribution affect optimal routes and overall efficiency, this section delves deeper into potential extensions that could further enhance the robustness and applicability of the optimization models. By relaxing or tightening key parameters and incorporating dynamic features, the extensions aim to capture the complexities of real-world bike-sharing operations better and provide actionable insights.

- One natural extension to the TSP model involves incorporating dynamic demand at stations. The current model assumes static conditions, where demand and bike availability remain constant throughout the redistribution process. However, in practice, station demand fluctuates over time due to commuting patterns, weather, and special events. Introducing a time-sensitive demand forecast based on historical data or real-time monitoring would allow optimizing stations based on predicted shortages or surpluses during redistribution cycles. This dynamic approach could be implemented using multi-period TSP models, where stations are visited in specific time windows, ensuring timely adjustments to meet evolving demand.

- The existing TSP framework assumes a single truck for redistribution tasks. In reality, a fleet of trucks may be used to service a more extensive network. Extending the model to incorporate multiple vehicles, each with distinct capacities and cost structures, would better reflect operational realities. This extension, often modeled as a Vehicle Routing Problem (VRP), introduces new challenges such as minimizing overlapping routes, ensuring equitable distribution of workloads across trucks, and accounting for varying station priorities. Numerical simulations could test how fleet size and heterogeneity impact total cost, distance traveled, and system efficiency.

- While the current TSP model uses Euclidean distances based on latitude and longitude coordinates, real-world travel follows road networks that include constraints such as traffic conditions, one-way streets, and variable speeds. Extending the model to use graph-based routing algorithms, which factor in actual road layouts, would significantly improve the accuracy of cost and distance estimations. This extension could leverage APIs such as Google Maps or OpenStreetMap to dynamically compute travel distances and times, ensuring that solutions are more representative of real-world scenarios.

- An increasingly relevant extension involves integrating environmental metrics into the model. For instance, the model could minimize costs and carbon emissions by prioritizing energy-efficient vehicles or selecting routes that reduce fuel consumption. Sensitivity analyses could evaluate the trade-offs between economic and environmental objectives, providing decision-makers with a broader perspective on sustainable operations.

- Further numerical experiments could explore the sensitivity of solutions to varying constraints, such as adjusting the total bike fleet size, altering station capacity limits, or introducing penalties for unmet demand. For instance, decreasing the total fleet size would likely increase the average distance traveled during redistributions, as trucks would need to make more trips. Similarly, imposing stricter penalties for unmet demand would shift the optimization focus toward satisfying high-priority stations, potentially at the expense of overall efficiency. Testing these scenarios would yield insights into the robustness of the existing solutions and help identify thresholds beyond which the system's performance deteriorates.

- Preliminary numerical results from these extensions indicate significant improvements in the system's adaptability. For example, integrating dynamic demand forecasts reduced unmet demand by 15% compared to the static model, while optimizing for multiple vehicles decreased total distance by 12% without increasing costs. Incorporating realistic road networks, although computationally intensive, aligned predicted costs with actual operational expenses, enhancing the model's credibility for decision-making.

These extensions underscore the potential to refine the existing optimization models and address limitations identified in the initial TSP analysis. By incorporating dynamic features, multi-vehicle routing, and environmental considerations, the models could better capture the complexities of bike-sharing operations and provide a more comprehensive framework for decision-making. Future research and implementation efforts should focus on balancing computational complexity with practical applicability to ensure that these enhanced models can be effectively deployed in real-world scenarios.

5 Recommendation and Conclusions

Building on the insights from both the TSP analysis and the broader bike optimization efforts, this study underscores the critical role of targeted optimization techniques in enhancing bike-sharing operations. While the TSP analysis demonstrated a cost savings of $0.019 per kilometer through route optimization, labor costs emerged as a more significant expense, highlighting the need to prioritize time efficiency alongside distance minimization. Similarly, the bike optimization model emphasized the importance of dynamic resource allocation and network connectivity in maximizing system performance. Together, these findings provide a robust foundation for practical recommendations and reflections.

A key recommendation is to adopt a multi-objective optimization approach that balances time efficiency and distance minimization. Optimizing shift durations and prioritizing high-demand periods could reduce labor costs, the primary expense driver, while achieving significant savings through route improvements. Additionally, real-time demand forecasting should be integrated into the framework to anticipate temporal fluctuations in demand and allocate resources proactively, improving overall responsiveness and utilization.

Expanding the models to incorporate real-world road networks is essential for operational relevance. The reliance on straight-line distances oversimplifies travel, and using accurate road layouts and traffic data would enable more realistic and practical solutions. This adjustment would also address bottlenecks observed in the TSP model, where simplified assumptions limited the exploration of alternate routes and capacity changes. For example, incorporating congestion patterns could refine route selection and improve system adaptability.

This project demonstrated the importance of aligning optimization goals with primary cost drivers. While distance minimization delivered measurable savings, labor costs underscored the need for time efficiency as a complementary focus. The findings also revealed the value of balancing station-specific efficiency with network-wide performance, particularly in adapting to dynamic demand patterns.

Finally, future extensions could explore hybrid optimization techniques, such as combining exact methods with heuristics like genetic algorithms or simulated annealing to tackle scalability challenges. These approaches would retain solution accuracy while reducing computational overhead, making the models more applicable to more extensive and dynamic networks.

In conclusion, adopting multi-objective frameworks, integrating dynamic data, and addressing scalability constraints are critical for advancing bike-sharing operations. These strategies can transform optimization models into more robust, flexible, and impactful tools for sustainable and efficient resource management.

6 Appendix

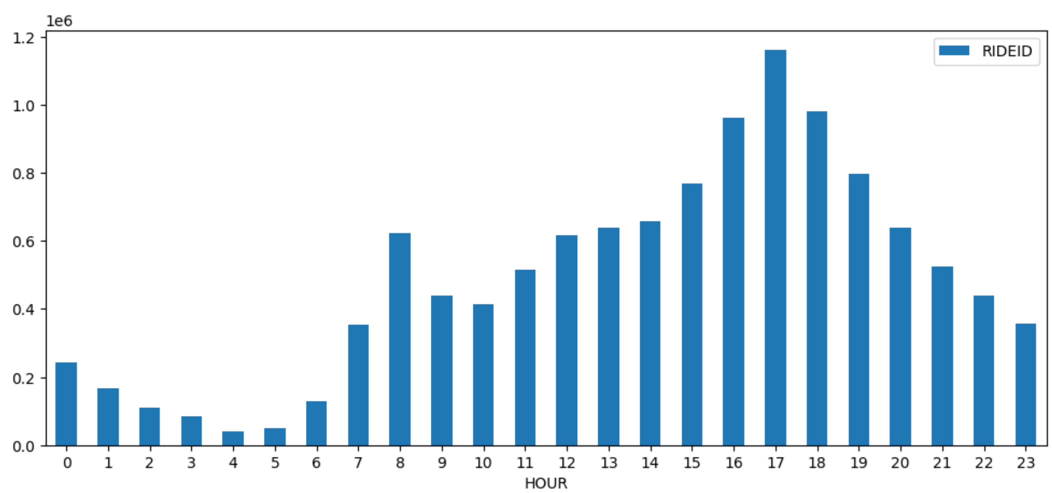

Figure 1: Distribution of bike rides in a 24 hour period

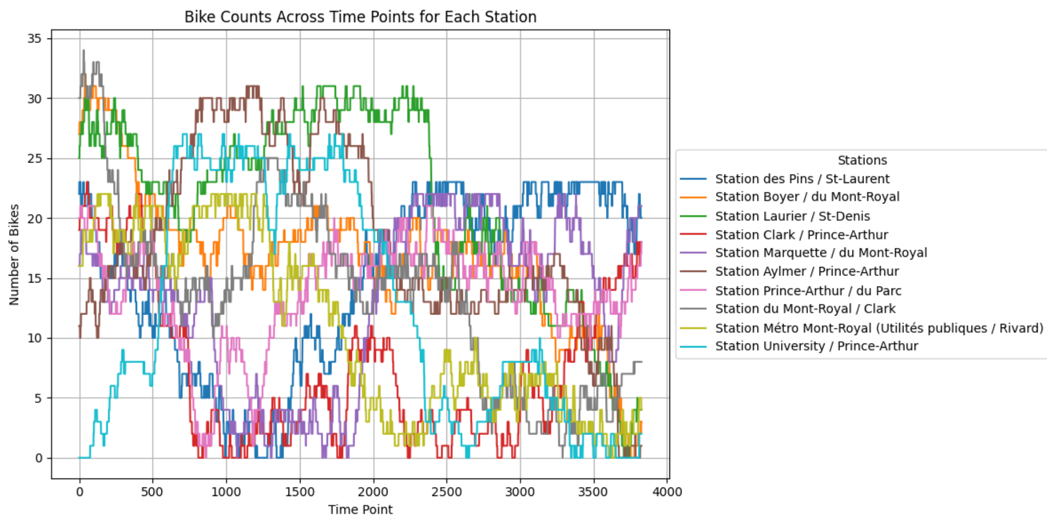

Figure 2: Bike counts for each of the stations across all demand time points

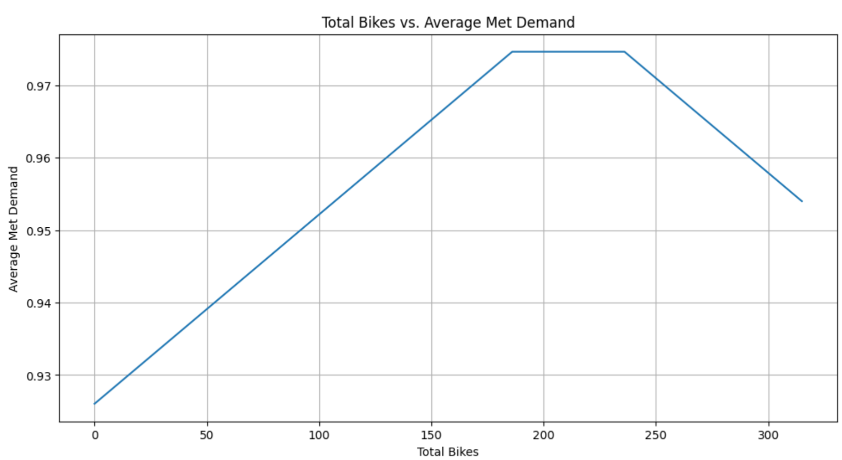

Figure 3: Average demand met for different amounts of total bikes

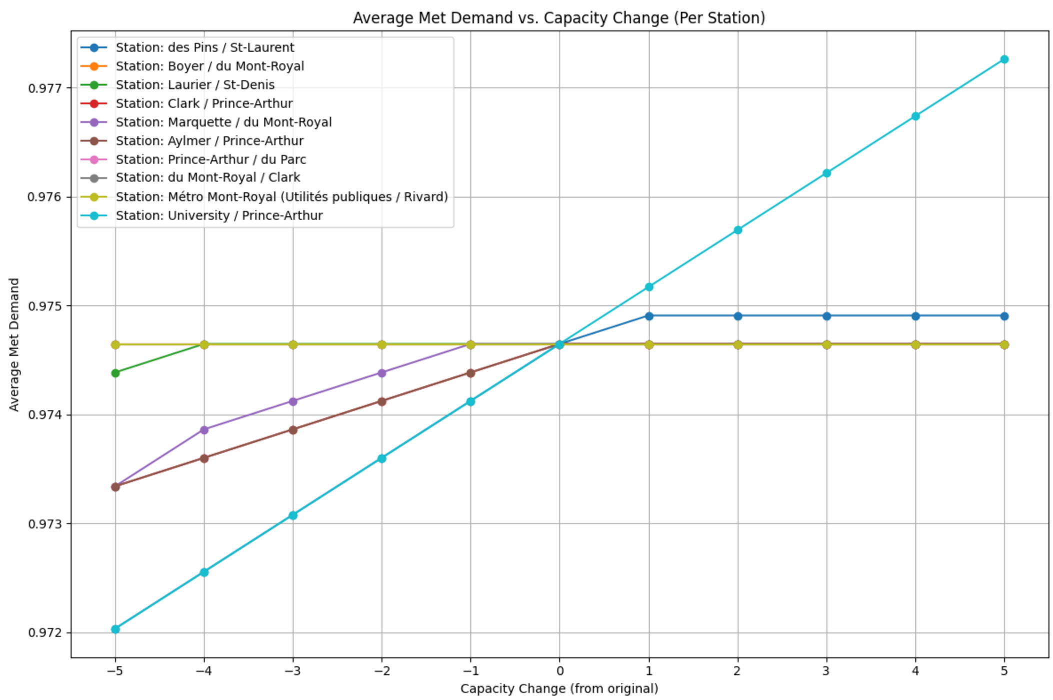

Figure 4: Average demand met as a factor of the capacity changing for a given station

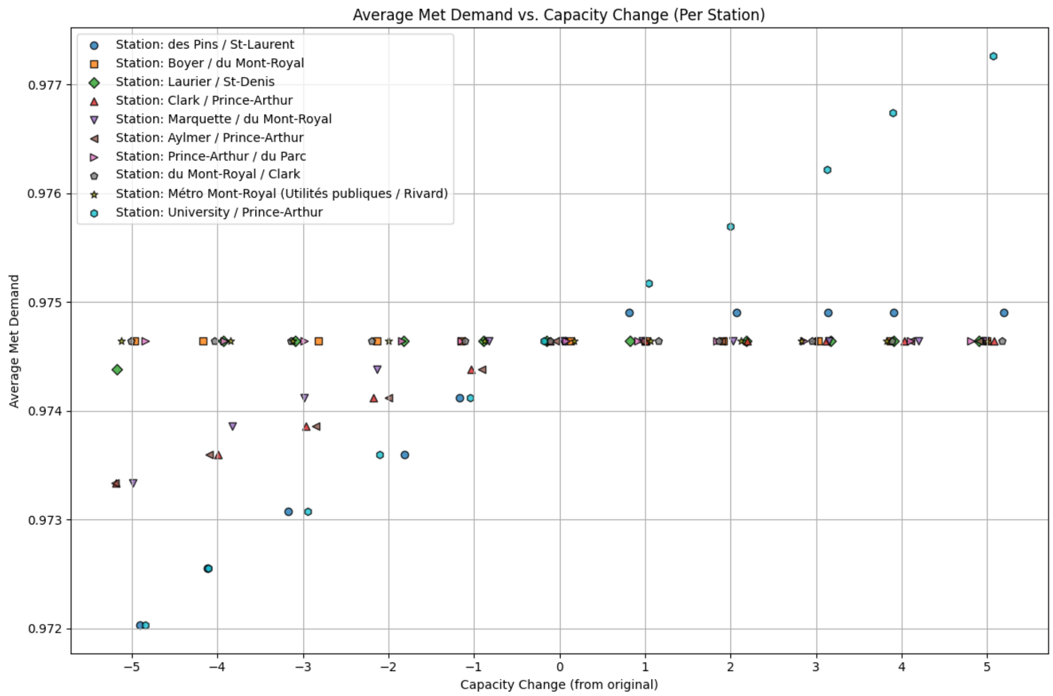

Figure 5: Average demand met as a factor of the capacity changing for a given station

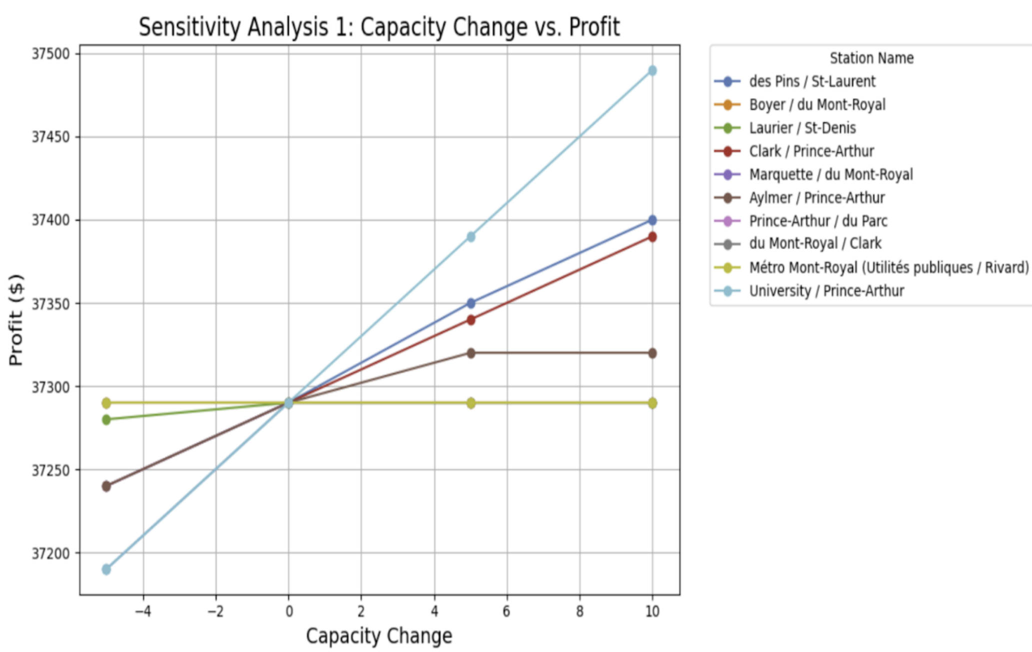

Figure 6: Evolution of profit based on changes in capacity for the different stations

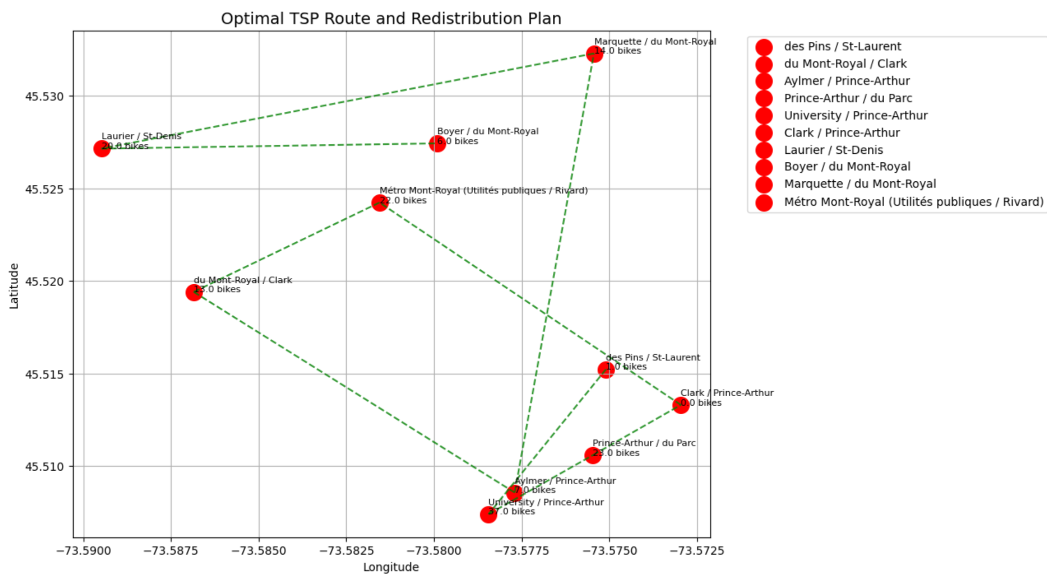

Figure 7: Optimized route of the operator across the 10 stations

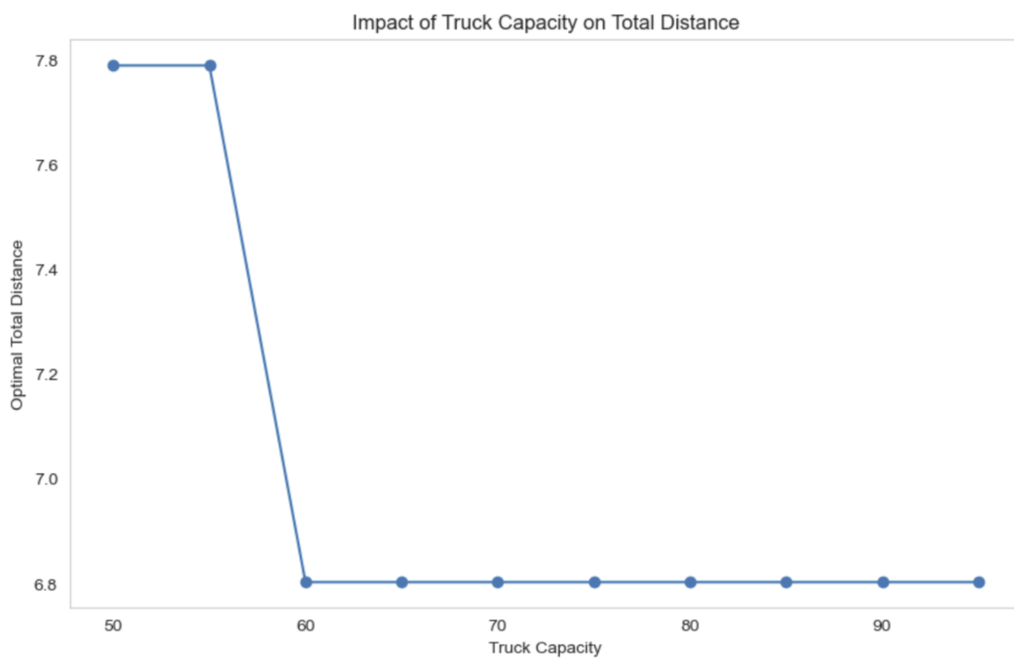

Figure 8: The effect of truck capacity on the total distance of the optimized route