OPTIMIZATION OF BLUEBIKES IN BOSTON

OR6205 PROJECT REPORT 2023

Group Number: 8

Group Members:

Anuj Pawar, Anand Burkul, Ameya Patil, Sony Kumar, Shreya Pande, Naresh Rajendiran.

ACKNOWLEDGEMENTS

We would like to thank our course instructor, Professor Mehdi Behroozi, and Rakhsith Linkappan Anand, our teaching assistant for the course, for providing valuable insight and advice for the duration of this project and guiding us and helping us throughout the semester. Their input has helped us get a strong footing in the fundamentals of this subject, for which we would like to express our gratitude.

ABSTRACT

Amidst the COVID-19 pandemic, Bluebikes experienced a significant decline in overall revenue. Operated jointly by 10 municipalities, pinpointing the specific areas facing more significant losses proves challenging. Complicating matters, the primary customer demographic for Bluebikes consists of students, a number expected to rise annually. In response, a project was initiated to enhance revenue across Bluebike stations throughout Boston. The focus was on optimizing costs based on the activity levels of bike stations. Data, including bike and station counts, membership costs (monthly, annually, and per ride), was gathered from the official Bluebikes website. Using this information, an algebraic Integer Linear Programming model was formulated and implemented in Python to identify optimal station locations and student pathways for maximizing revenue.

Table of Contents

1. Introduction 5

2. Problem Description 8

3. Preliminary Mathematical Model 9

   3.1 Decision Variables and Objective Function 9

   3.2 Dataset Description 10

   3.3 Constraint Equations 10

   3.4 Assumptions and justifications on obtaining model inputs (data) 11

   3.5 Approach 11

   3.6 Preliminary Result Analysis 13

   3.7 Limitation for iteration 1 13

4. Iteration 2 Results Analysis: 14

   4.1 Assumptions and justifications on obtaining model inputs (data) for Iteration 214

   4.2 Limitation for Iteration 2:21

5. Sensitivity Analysis 21

6. Conclusion 23

7. Reference 24

1. INTRODUCTION

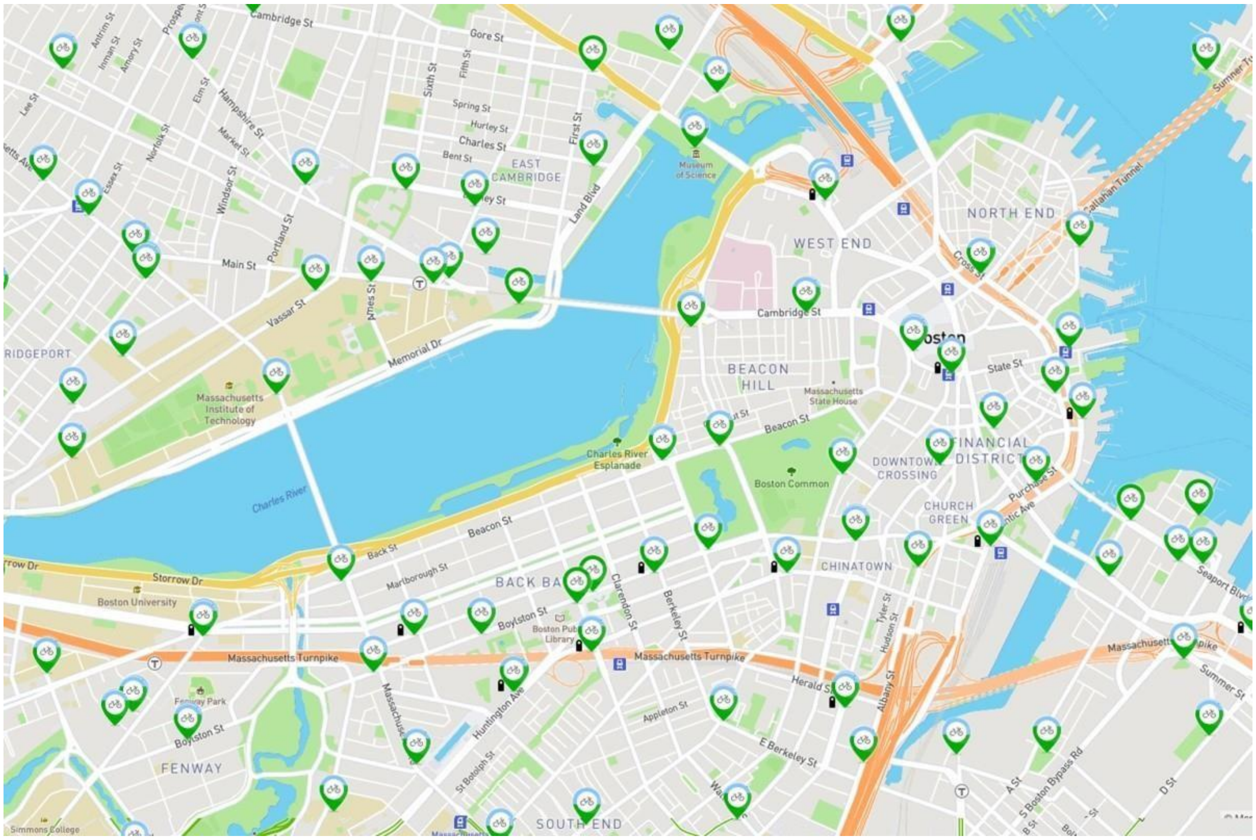
Bluebikes is a public transportation system by bike. The town of Brookline and the cities of Boston, Cambridge, Everett, Salem, and Somerville jointly own and operate the system. Blue Bikes is made up of a fleet of bicycles that are dispersed throughout the city via a system of docking stations. 400 stations and 4,000 bikes are available to riders in Metro Boston's 11 municipalities.
More than 15 million journeys have been made using bike sharing since 2011. This service is advantageous for one-way excursions because users can unlock bikes at one station and return them at any other station. This expedites the additional cumbersome task of returning the bike to where it was originally taken from.

Since its introduction in 2011, the bike-sharing program in Boston has experienced successful expansion, offering residents a convenient means
of city travel without the challenges of public transportation or car ownership
in densely populated areas. Notably, the program's positive impact extends beyond transportation, promoting both mental and physical well-being through regular bike riding. Recognized as an affordable option, Bluebikes has become
a popular choice for short-distance travel, offering diverse payment options to cater to citizens' needs.

A key strength lies in its environmental friendliness, contributing to reduced carbon emissions and improved air quality as more individuals embrace this eco-friendly mode of transportation. The accompanying mobile app enhances user experience by providing real-time updates on dock station statuses across Boston, enabling users to optimize their routes effectively.

However, the COVID-19 outbreak significantly impacted Bluebikes' usage, with a noticeable decline in journeys during April and May 2020, following record-high levels in the first quarter. This pandemic-induced decrease in bike usage resulted in a substantial drop in Bluebikes' overall revenue.

In response to these challenges, our initiative aims to boost Bluebikes' revenue while maintaining service quality. Our goal is to encourage continued usage of Bluebikes as a reliable and eco-friendly mode of transportation, despite the setbacks caused by the pandemic.



System of bluebike

2.PROBLEM STATEMENT

The aim of this project is to enhance Bluebikes' profitability while preserving the high standards of customer service they provide. Given the notable decline in Bluebikes users following the onset of the COVID-19 outbreak, our objective is to use an optimization function to maximize Bluebikes' profits. The overarching goal is to promote this public transportation mode, which carries various social, health, and environmental benefits

In our problem, we divide the city of Boston into four separate regions i.e., East, West, North and South. Each of these four regions is further divided into two distinct regions based on the data of the total number of trips recorded by Bluebikes. For example, East Boston is divided into a less active region (with minimum number of trips) and an active region(with maximum number of trips). Similarly, we consider the data of the total number of dock stations in each region and the fixed maintenance cost required for it. According to the data provided by Bluebikes, there are around 400 dock stations available across the city ofBoston, which we have divided according to the number of stations available in each region. Hence, the objective function defines the cost for each trip of a Blue Bike ride, which is $2.95, times the total number of trips and the reduction of the overall maintenance cost for each dock station in different regions.

This objective function is subject to a set of constraints that we have taken. The constraint limits have been partially assumed from data that was taken from Bluebikes regarding the number of monthly users in specific and distinct regions of the city.

2. PRELIMINARY MATHEMATICAL MODEL

An initial preliminary linear programming model was formulated. Atotal of 16 decision variables were considered the following way:

X_1: Total number of trips in East Boston (active)

X_2: Total number of trips in East Boston (less active)

X_3: Total number of trips in West Boston (active)

X_4: Total number of trips in West Boston (less active)

X_5: Total number of dock stations in East Boston(active)

X_6: Total number of dock stations in West Boston (less active)Maintenance

Costs:
$500 - East Boston
$1200 - West Boston

CONSTRAINTS AND DATASET DESCRIPTION:

Now that the objective function and variables have been established, we can proceed with the formulation of the constraints for the problem. Dueto lack of data on Bluebikes’ website, we must make several assumptions to move on with the bluebikes profit maximization strategy. Here, we have assumed that the number of docks at each station will always be the same, and that the cost of maintenance will be fixed as well (since the maintenance cost is based on the number of docks and number of docks are assumed to be constant).
word
The initial set of constraints will be based on how many trips are madein each of the four regions being evaluated (East, West, North, and South). The total number of trips in the East area should be greater than or equal to the number of trips that are currently made in the region, bothin the active and less active regions. As a result, the constraints shouldbe as follows when applying the same to all regions:

$X_1 + X_2 \leq 521000$

$X_3 + X_4 \leq 1070000$

Next set of constraints will be based on the number of docks available inboth regions being evaluated (East, West). The total number of docks in the East and West regions should be greater than or equal to the number of docks that are currently available in the region, both in the active and less active stations. As a result, the constraints should be as follows when applying the same to all regions:

$X_5 \leq 100$

$X_6 \leq 150$

LINEAR PROGRAMMING MODEL

MAXIMIZE Z = 2.95*(X1 + X2 + X3 + X4) - 500*X5 - 1200*X6

All Xi are ≥ 0, where i = 1,2,3,4,5,6

APPROACH

The linear program developed in this analysis attempted to maximize the overall revenue for the upcoming years by considering the data of number of trips from each station in the different region of boston i.e. in East and West direction. Our initial program will attempt to optimize using the current resources available. Further iterations will investigate the impact of adding other sources to solve the problem.

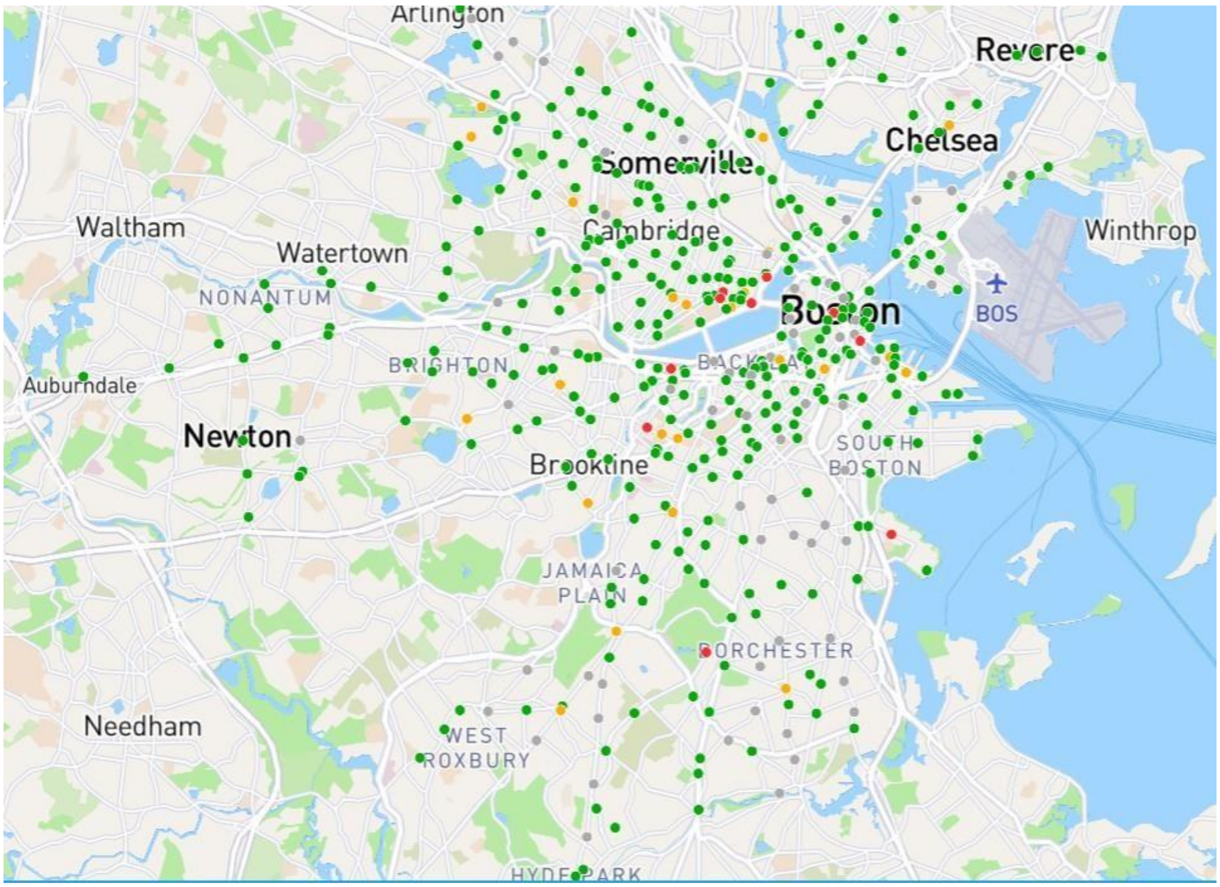
ASSUMPTIONS AND JUSTIFICATIONS ON OBTAININGMODEL INPUTS (DATA)

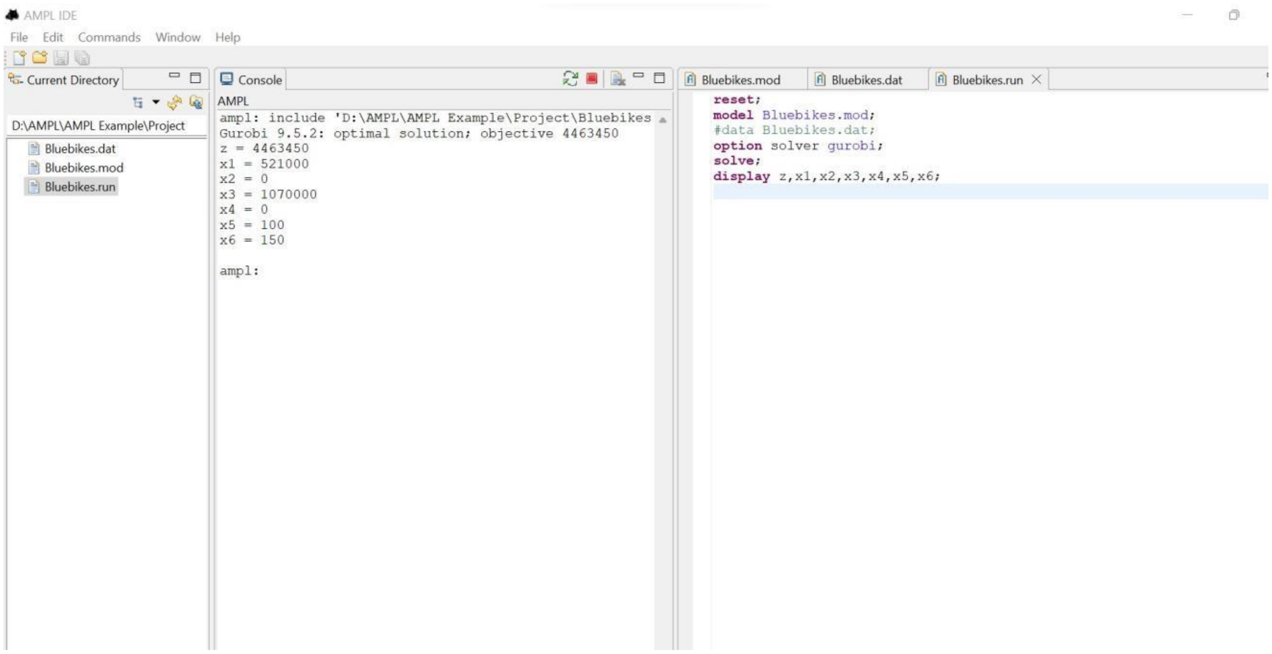
There are a range of membership and payment plans for using Bluebikes. The options are:
$2.95 for one ride
$26.75 for a monthly membership
$119 for an annual membership

The above data was collected from the Bluebikes website. We have also assumed the cost of maintaining a dock station in each of the following four regions: East Boston, West Boston. The corresponding maintenancecosts are listed below:

$500 - East Boston
$1200 - West Boston

Additionally, for the number of dock stations in both regions, we have considered the map of Boston which is further categorized into the various parts of boston so based on number of trips from each station we have considered the active and less active dock stations





LIMITATIONS OF PRELIMINARY MODEL

The drawbacks from the preliminary model observed was that the result for maximization of profit was not accurate enough to make a proper prediction, as we have considered only two regions and divided the data accordingly. Also, we have only considered the data for the year 2022, which does not consider how the number of trips might change yearly. Therefore, we are not able to accurately predict how we should maximize the profit for the future of Bluebikes. We will have to further adjust and take up a more specific approach to further optimize the problem and get the optimal result.

FINAL MATHEMATICAL MODEL

The entire data of the total number of trips for each year from the beginning of Bluebikes to the present (2011 - 2022) is provided on the Bluebikes' website. The next step taken was to consider these values from each year and assign a certain percentage of each of these values toeach of the four regions considered, i.e., East, West, North and South. This is done considering the frequency of trips in each region. These four percentages obtained are further divided into two each, according tosubdivisions based on active regions and less active regions in the four considered regions. This will give us eight percentages for each year from 2011 - 2022. The objective of this approach is that we can get the maximum profit Bluebikes had for each year, and then take the maximum of these values. This will help Bluebikes to better analyze and predict how to distribute dock stations by considering the minimum values above which the revenue generated will result in profit.Doing this will provide a better idea of how to maximize their profit for the coming years.

DECISION VARIABLES

Since we are considering four different regions for this iteration, we will have to define a new set of variables accordingly:

X_1: Total number of trips in East Boston (active)

X_2: Total number of trips in East Boston (less active)

X_3: Total number of trips in West Boston (active)

X_4: Total number of trips in West Boston (less active)

X_5: Total number of trips in North Boston (active)

X_6: Total number of trips in North Boston (less active)

X_7: Total number of trips in South Boston (active)

X_8: Total number of trips in South Boston (less active)

X_9: Total number of dock stations in East Boston X_{10}:

Total number of dock stations in West BostonX_{11}:

Total number of dock stations in North BostonX_{12}:

Total number of dock stations in South Boston

OBJECTIVE FUNCTION

MAXIMIZE Z = 2.95*(X_1 + X_2 + X_3+ X_4+ X_5+ X_6+ X_7+ X_8) - 500*X_9 - 1200*X_{10} - 1500*X_{11}- 800*X_{12}

All $X_i \geq 0$, where i = 1,2,3,4,5,6,7,8,9,10,11,12

CONSTRAINTS

For this iteration, the number of constraints will increase as we have further considered four different regions, as opposed to only two regionsin the first iteration. The constraints will they be as follows:

For total number of trips in four regions in one year:

$X_1 + X_2 \leq 521000$

$X_3 + X_4 \leq$
1070000 $X_5 + X_6$
≤ 1750000

$X_7 + X_8 \leq 632000$

For total number of dock stations in four regions:

$X_9 \geq 100$

$X_{10} \geq 150$

$X_{11} \geq 175$

$X_{12} \geq 225$

ASSUMPTIONS AND JUSTIFICATIONS ON OBTAININGMODEL INPUTS (DATA)

There are a range of membership and payment plans for using bluebikes. The options are:
$2.95 for one ride
$26.75 for a monthly membership
$119 for an annual membership

The above data was collected from the Bluebikes website. We have also assumed the cost of maintaining a dock station in each of the following four regions: East Boston, West Boston, North Boston, and South Boston. The corresponding maintenance costs are listed below:
$500 - East Boston
$1200 - West Boston
$1500 - North Boston

$800 - South Boston

Additionally, for the number of dock stations in each of the four regions,we have considered the map of Boston which is further categorized into the various parts of Boston so based on number of trips from each station we have considered the active and less active dock stations.

STATION	TOTAL TRIPS IN 2021
MIT at Mass Ave / Amherst St	70,300
Central Square at Mass Ave / Essex St	53,015
Charles Circle - Charles St at Cambridge St	42,172
Harvard Square at Mass Ave/ Dunster	40,647
Ames St at Main St	37,241
Christian Science Plaza - Massachusetts Ave at Westland Ave	34,356
MIT Pacific St at Purrington St	31,530
Beacon St at Massachusetts Ave	31,233
MIT Vassar St	28,684
Commonwealth Ave at Agganis Way	28,469
Cross St at Hanover St	28,215

Bluebikes data in .csv file:

BluebikesData

Year	Total Number of trips	Total available dock stations	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12
2011	140974	61	11277	7048	21146	16916	35243	26785	14097	8458	15	57	6	39
2012	531588	105	42527	26579	79738	63790	132897	101001	53158	31895	126	40	38	85
2013	908165	130	72653	45408	136224	108979	227041	172551	90816	54489	68	15	8	61
2014	1184558	140	94764	59227	177683	142146	296139	225066	118455	71073	28	65	24	70
2015	1122544	155	89803	56127	168381	134705	280636	213283	112254	67352	46	79	43	43
2016	1236236	180	98898	61811	185435	148348	309059	234884	123623	74174	9	8	17	78
2017	1313837	190	105106	65691	197075	157660	328459	249629	131383	78830	110	74	39	34
2018	1767806	262	141424	88390	265170	212136	441951	335883	176780	106068	54	13	28	90
2019	2520418	325	201633	126020	378062	302450	630104	478879	252041	151225	144	16	25	49
2020	2065292	365	165223	103264	309793	247835	516323	392405	206529	123917	103	36	33	53
2021	2954624	400	236369	147731	443193	354554	738656	561378	295462	177277	88	80	34	86
2022	3944956	440	315596	197247	591743	473394	986239	749541	394495	236697	53	37	47	49
	Constraints RHS													
EAST	512843	521000												
WEST	1065137	1070000												
NORTH	1735780	1750000												
SOUTH	631192	632000												

RESULTS ANALYSIS

The linear program, consisting of the above- mentioned objective function along with the constraints were written as code and entered in Python. The code entered along with the solution is shown below

In [1]:

```
import numpy as np
import pandas as pd
import pickle as pkl
```

In [2]:

```
FP = "./BluebikesData.csv"
```

In [3]:

```
with open(FP, 'r') as f:
    data_full = pd.read_csv(f)
```

In [4]:

```
data_df = data_full.iloc[:12]
n = len(data_df)
```

In [5]:

```
ans = [0, 0, 0, 0] # first index is (x1+x2), etc. till (x7, x8)
const = [ 521000, 1070000, 1750000, 632000 ]
inds = [ (0, 0), (0, 0), (0, 0), (0, 0) ] # stores indices of rows for each

for upd in range(4):
    for i in range(n):
        for j in range(n):
            v1 = data_df.iloc[i, 3+(2*upd)]
            v2 = data_df.iloc[j, 3+(2*upd)+1]
            if (v1+v2 < const[upd]) and (v1+v2 > ans[upd]):
                ans[upd] = v1+v2
                inds[upd] = (i, j)
```

In [8]:

```
print("Max value rows:")
for i in range(len(ans)):
    print(f"x_{(i*2)+1}: {inds[i][0]}, val: {data_df.iloc[inds[i][0], (2*i)+
```

```
Max value rows:
x_1: 11, val: 315596.0
x_2: 11, val: 197247.0
x_3: 11, val: 591743.0
x_4: 11, val: 473394.0
x_5: 11, val: 986239.0
x_6: 11, val: 749541.0
x_7: 11, val: 394495.0
x_8: 11, val: 236697.0
```

```
In [7]:
min_vals = [1e9, 1e9, 1e9, 1e9]
min_inds = [0, 0, 0, 0]
min_const = [100, 150, 175, 225]

for upd in range(4):
    col = upd+11
    for j in range(n):
        if (data_df.iloc[j, col] < min_vals[upd]) and (data_df.iloc[j, col]
            min_vals[upd] = data_df.iloc[j, col]
            min_inds[upd] = j
for i in range(len(min_vals)):
    print(f"x_{i+9}: {min_inds[i]}, val: {min_vals[i]}")
```

```
x_9: 5, val: 9.0
x_10: 5, val: 8.0
x_11: 0, val: 6.0
x_12: 6, val: 34.0
```

```
In [13]:
z = (2.95 * sum(ans)) - (500*min_vals[0]) - (1200*min_vals[1]) \
    - (1500*min_vals[2]) - (800*min_vals[3])
print(z)
```

```
11587308.4
```

As seen from the code above, we have an optimal solution which isas follows:

Z = 11587308.4

X_1= 315596	X_9= 9
X_2= 197247	X_{10}= 8
X_3= 591743	X_{11}= 6
X_4= 473394	X_{12}= 34
X_5= 986239	
X_6= 749541	
X_7= 394495	
X_8= 236697	

LIMITATIONS FOR FINAL MATHEMATICAL MODEL

On comparing the final model to the initial model, we found that while we had divided Boston into four regions based on the number of trips, we had only considered the cost per ride and had neglected to account for monthly and annual membership, which resulted in some inaccuracy in the final figure. The fact that the annual number of bike trips varies is another obstacle to this undertaking.

Another drawback of our model is that we did not account for the number of employees who support the operation of the Bluebikes system, which has an impact on bluebike's overall revenue**.**

SENSITIVITY ANALYSIS

A sensitivity analysis in operations research is an important tool in the search for an optimal solution to engineering problems in a project where it must be determined how different parameters will affect it. It is an analysis that seeks to determine the effects produced in the optimal solution by a change in any parameter of a linear programming model. Itinvestigates how a modification in the coefficients of the basic and non-basic variables will affect the available resources or if a new restriction is introduced.

For this analysis, we are using the total number of dock stations and the maintenance cost as our parameters. So as a result, we are getting to know that if we change the number of dock stations with respect to change in maintenance cost, our profit value is being changed.

Number of dock stations	400
Maintenance Cost	$1,000.00
Total	$50,300.00

$50,300.00	300	350	400	450	500
$ 1,200.00	40500	60400	80300	100200	120700
$ 1,100.00	22300	45300	65300	85300	100200
$ 1,000.00	20300	35300	50300	65300	80300
$ 900.00	18500	28500	35300	40500	60300
$ 800.00	15300	19300	20300	23300	30100
$ 700.00	12500	12500	5300	20100	22500

CONCLUSION

In the preliminary model, we divided the city of Boston into two regions: East and West. This was done based on the number of trips recorded at each dock station. Since we only considered two regions and divided the data accordingly, one of the shortcomings of the original model was that the result for profit maximization was not precise enough to make an accurate prediction and also we have considered the data only for one year.

Once the appropriate changes were made in the second iteration, by adding more variables and considering data from multiple years, we were able to acquire a solution more suitable for our needs. The code forthis iteration was solved in Python, as we found it to be more accessible to us for acquiring the solution we needed.

After this, through sensitivity analysis, we were able to pinpoint what combination of cost and number of dock stations is generating the most revenue. This also provides valuable insight.

REFERENCES

1) https://www.bluebikes.com/system-data

2) https://www.kaggle.com/datasets/jackdaoud/bluebikes-in-boston

CONTRIBUTION OF MEMBERS

- Anuj Pawar – 16.66%
- Anand Burkul - 16.66%
- Ameya Patil - 16.66%
- Sony Kumar - 16.66%
- Shreya Pande - 16.66%
- Naresh Rajendiran - 16.66%

SUMMARY OF PRESENTATION:

Group 1
Title: The Role of Operations Research at DoorDash and Formula 1: A Comprehensive Study

This project studies the uses of advanced mathematical techniques to improve operational efficiency in two companies: Door Dash and Formula 1. In Door Dash, the project explores a new optimization method namely Mixed Integer Programming that increases efficiency by tenfold. This allows Door Dash to handle complex delivery routes in real-time. The project then explains our interpretation of Mixed Linear Programming using IBM CPLEX to demonstrate the code that the Door Dash used in Gurobi. The project then delves into exploring the field of Formula 1, it analyzes how OR can be applied to race strategy, pit stops, and car design. Three case studies of successful F1 teams are included. Overall, this project demonstrates the value of OR methods in improving performance and decision-making across various industries.

Group 2
Title: Online Retail Inventory Management Case Study

This project studies and suggest a large-scale optimization framework-based approach to the online-retail inventory placement problem. We formulate the issue as a program with mixed integers. Our framework seamlessly integrates item aggregation and column production to provide a full solution. To guide the aggregation process, we use the problem data to construct an a priori theoretical optimality gap bound. We demonstrate our framework's ability to find nearly optimal solutions in a few hours using a numerical example including one million items. Additional computation time reductions can be achieved by utilizing parallel processors. In particular, our method outperforms a sequential placement heuristic that is representative of the present state of affairs. Managing scenarios where sparsity limits are directly placed on fulfillment center count or item count is an extension of our work. Compared to indirect control through established cost parameters, this creates extra challenges and results in more complex and difficult-to-solve optimization models. In light of this, fixed expenses appear to offer a more straightforward approach to tackling. It seems that future development could include the concept of iterative aggregation. We propose to allow numerous layers of aggregation, repeatedly fine-tuning the placement method at each level, instead of restricting it to one level (items to clusters). This approach has been successfully used in various aggregation-based optimization frameworks and may increase scalability without significantly narrowing the optimality gap.

Group 3
Title: Linear Programming to Optimize the LOUISVILLE MSD Sewer System.

This project studies, Tetra Tech used Mixed Integer Linear Programming to optimize the LOUISVILLE MSD sewer system, which successfully captured the non-linear and discontinuous behavior of the system. Gurobi optimization combined with an iterative method simplified the optimization procedure, resulting in an optimization duration of only 5 minutes and impressive efficiency. This methodology not only guaranteed precise portrayal of intricate sewer system dynamics but also showcased the capacity to provide timely and efficient resolutions.

Group 4
Title: Nuts & Fit Profit Maximization

Profit maximization through operational research is the goal of this project, The Nuts&Fit project. The group applied the simplex algorithm to five items while taking infrastructure, demand, and raw material constraints into account. Sensitivity analysis was used to determine how changing certain factors would affect earnings. After creating the problem statement, the operations team had to take into account all of these limitations and execute simplex for five of her goods. It was discovered that each product has a different contribution, with granola bars outperforming nut butters. However, nut butters are more in demand than granola bars. To achieve the best outcomes, however, it is preferable to make the largest variety of nut butters rather than the most lucrative one. Sensitivity analysis is used to calculate the total impact of adding independent variables to dependent variables on profit margin. Two scenarios were tested: managing a 20% reduction in the supply of oats owing to a heatwave, and raising the pricing of items X1 and X4, which resulted in increased profits. It is advised to diversify the sources of oats, maintain strong inventory control, optimize production in a flexible manner, assess price changes, and conduct continuous market research. The CEO uses the analysis to guide well-informed decisions that maximize profits.

Group 5
Title: UPS Route Optimization: A Comprehensive Case Study

Group 5's "UPS Route Optimization: A Comprehensive Case Study" revolves around UPS and its endeavors to improve operational efficiency by means of route optimization. The report highlights UPS's position as a leader in global logistics, its competitive environment, its main business segments, and data for 2022. It explores the problems with UPS's old routing system and how it changed to a more sophisticated, algorithm-driven system dubbed ORION (On-Road Integrated Optimization and Navigation). By streamlining service sequences and routes, ORION seeks to address the Traveling Salesman Problem with Time Windows (TSPTW), which will result in considerable financial and environmental savings. The study has wider applicability than just logistics and is in line with the ideas of operations research.

Group 6

Title: Application of the simplex method on profit maximization in Baker's Cottage

The focal point of the Baker's Cottage case study is to strategically optimize profit through the production of various bread types, utilizing data on five specific varieties. It uses an Operational Research approach to solve the problem, encompassing the establishment of an objective function, the definition of constraints, and the formulation of mathematical models. Subsequentlyx, the linear programming problem is systematically tackled using Simplex Method via the Excel software. Noteworthy findings highlight the substantial profit maximization achieved by focusing on the production of 332 loaves of Chicken Floss and 196 loaves of Frank Cheese, with other bread types demonstrating a limited impact on overall profitability. In conclusion, the significance of integrating linear programming into the manufacturing sector is underscored, emphasizing its potential to elevate profitability, optimize resources, and enhance operational efficiency.

Group 7

Title: OPTIMIATZATION OF FOOD PARCEL ASSEMBLY USING LINEAR PROGRAMMING CONCEPTS

The objective of this project is to create a model for optimizing a Rio de Janeiro company's food package line using the ideas of linear programming. Their offering is a food basket filled with a variety of goods like cheese, ham, and other items. A supply chain department team member's manual surveying of the raw materials was prone to human error, which caused the final basket to be assembled unevenly. Particularly during Christmas, when sales were at their highest, which was having an impact on their delivery and client happiness. Therefore, through our case study, we investigated the issue and created a model that functions flawlessly and is provided in an approachable manner utilizing an Excel solution. By this, the company will be able to keep the stock in check and will be able to assemble the final basket properly without missing out on a single item in the combo. This also helped the company by stopping their investment in professional ERP software like SAP and Talley.