Feasibility Validation of Launching LA Metro Bike Relocation Service using Stochastic Programming

Jae Yul Woo
University of Southern California

Abstract— **The LA Metro Bike is mainly used as a commuting service by many LA citizens. Hence, there exists imbalance between the stations and huge cost of relocating the bikes by LA Metro employers. This study is focused on the unforced bike relocation by launching a new service. With current operation cost in LA Metro and price policy to users, the company can induce users by discounting 25% of the price to 'non-passholders' and rewarding the same amount to 'passholders.' The optimal routes and number of rides for the new service is in the file 'optimal_sol.csv' Furthermore, it is expected that the operation cost after launching the new service will fluctuate. Thus, the LA Metro should keep track of operation cost and keep the cost in the range of $0.10/mile to $3.10/mile to maintain the service at proper discount rate.**

I. Introduction

LA citizens are using metro for their daily basis. Although there are many kind of substitutes such as car-sharing services and electric-scooters prevalent in LA area, the metro bike has its strength in the accessibility and healthy features. LA Metro team publicized that there were total 700k trips and 2 million miles traveled so far with metro-bike. However, the trips are mostly focused on few routes since the bikes are mainly used for commutes. The LA metro team is taking charge of moving the unbalanced demand for their expenses. Hence, this study aims to validate the feasibility of the metro bike relocation service in the LA area.

II. Data Configuration & Preprocessing

A. Configuration

The data is shared in the LA Metro bike-share web page. Each csv file contains data for one quarter of the year. Each file contains the following data columns:

- trip_id: Locally unique integer that identifies the trip
- duration: Length of trip in seconds
- start_time: The date/time when the trip began, presented in ISO 8601 format in local time
- end_time: The date/time when the trip ended, presented in ISO 8601 format in local time
- start_station: The station ID where the trip originated
- start_lat: The latitude of the station where the trip originated
- start_lon: The longitude of the station where the trip originated
- end_station: The station ID where the trip terminated
- end_lat: The latitude of the station where the trip terminated
- end_lon: The longitude of the station where the trip terminated
- bike_id: Locally unique integer that identifies the bike
- plan_duration: The number of days that the plan the passholder is using entitles them to ride; 0 is used for a single ride plan (Walk-up)
- trip_route_category: "Round Trip" for trips starting and ending at the same station or "One Way" for all other trips
- passholder_type: The name of the passholder's plan

B. Data Preprocessing

Most of the data columns are categorical values except trip duration, start and end time. The bike trip data is from July 2016 to June 2017 with about 112 thousand rows.
Rides with NaN values(11177 rides), by 'LA Metro staff(3225 rides) and 'Round Trips(10459 rides) type rides are excluded.(Only the 'One-way' trips types are considered for the relocation service.) The passholder types are also reassigned to 0 and 1. Category 0 as non-passholders ('Walk-up, 24241), and category 1 as passholders. ('Monthly Pass + 'Flex Pass, 64025)
The 'distance' feature is added by calculating the Manhattan distance in miles between starting stations and ending stations. Station ID 4108 is classified as an outlier and all the trips that starts or ends at station 4108 are removed. 'Duration_min' column is also added to calculate the price for each trip. The current price policy is in the table. 1 below. Lastly, 9 features that are not relevant to the relocation service are removed. Thus, there are 'Duration'(in seconds), 'Start Time', 'Starting Station ID', 'Ending Station ID', 'Passholder Type', 'Distance', 'Duration_min'(in minutes) as columns in the processed dataset.

Pass Type	Price	Charges	Category
1 Ride	$1.75/30min.	every 30 min	0
Monthly	$1.75/30min.	first 30 min free	1
Annual	$1.75/30min.	first 30 min free	1

TABLE I: Price Policy

III. Exploratory Data Analysis

A. Duration Time by Passholder Types

By observing distances between stations and the mean, median duration time with respect to the passholder types, most passholders are using bikes for no more than 30 minutes. Non-passholders are also using bikes less than 30 minutes but few riders use a lot more than 30 minutes that the mean duration time exceeds 30 minutes.

Pass Type	Duration	Duration_min	Distance
0	2166.9427	36.1157	1.4315
1	771.7251	12.8621	5.3473

TABLE II: Mean Value

Pass Type	Duration	Duration_min	Distance
0	960	16.0	0.7308
1	480	8.0	0.6129

TABLE III: Median Value

B. Supply and Demand Imbalance in stations

The visualization of the top 15 popular starting and ending stations are in fig. 1. There are imbalances between starting and ending stations even though the popularity trends in both starting and ending stations are similar.

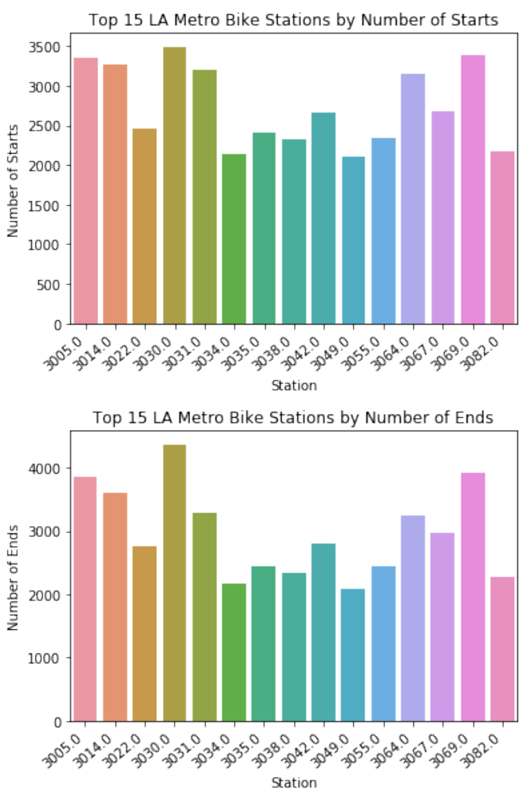

Fig. 1: Imbalance in Stations
(Upper: Starting Stations, Lower: Ending Stations)

C. Popular Trips

Popular trips in the dataset is derived as in fig.2. The most popular route is between station 3030 and station 3014, both ways.

D. Feature Correlations

The correlation between features are not significant as in fig.3. Thus, every feature is independent to each other.

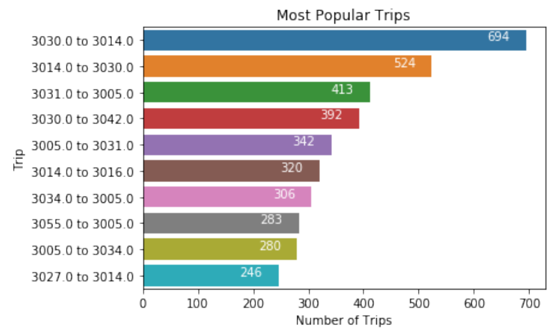

Fig. 2: Most Popular Trips

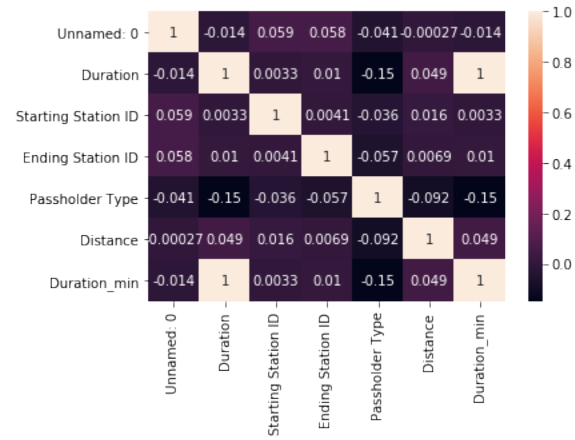

Fig. 3: Correlation Matrix

IV. Network

This relocation problem is inspired by the network problem and the problem consists of supplies, demands and the cost for each route. The bike-supplying stations and bike-demanding stations are defined as the supplies and demands in this relocation problem. The supplying stations are the stations with more incoming bikes than outgoing bikes, which means in trip data, the stations with more ending trips than the starting trips. Likewise, demanding stations are the stations with more outgoing bikes than incoming bikes, which means in trip data, the stations with more starting trips than ending trips. After grouping the data, there are 36 supplying stations and 27 demanding stations with their IDs and net supplies/demands, respectively.

Now in this network problem, the costs for each station routes are defined as the discounted dollar amount from the regular price without any discounts. LA Metro will receive $1.75 per every 30 minutes from non-passholders if there is no discount for relocation service. Passholders will also pay $1.75 per 30 minutes after the first exemption period. This price is calculated solely by the trip duration minutes. Since the trips from supplying stations to demanding stations do not occur naturally, LA metro needs to support people by discounting the prices, and this generates cost. This study is based on the assumption that discounted trip demands are always met.

There are 972 pairs of stations (36×27) for this problem but 154 pairs are missing by the fact that they were not present in the original dataset. The **big M method** is used for those pairs in the mathematical optimization.

$$
\begin{array}{|l|}
\hline
Supply\_dict. = \{0\text{: } 1012,\ 1\text{: } 143,\ 3\text{: } 107,\ \cdots\} \\
Demand\_dict. = \{2:488,\ 4:39,\ 7:21, \cdots\} \\
Cost\_dict. = \{(0,1):1.75, (0,2):1.75, \cdots\} \\
Distance\_dict. = \{(0,2):0.41,\ (0,4):1.17, \cdots\} \\
Missing\_dict. = \{(1,60):\text{M},\ (3,7):\text{M}, \cdots\} \\
\hline
\end{array}
$$

Fig. 4: The data in dictionary formats. (key: Station ID, value: amount)

The supplying station, demanding station, costs are defined for the relocation network problem and Manhattan distances between the stations were calculated during preprocessing procedure. The cost and distance dictionaries are *deep-copied* and updated with 'missing_dict.' to get full pairs of routes also with routes that did not have any trip before. The dictionaries will be used in pyomo with gurobi solver to optimize the relocation service.

V. Optimization

A. Optimization Objective

This optimization aims to find the optimal discount rate for the relocation service which makes the sum of total discount costs and operation costs minimized. With the optimal discount rate, the routes(Starting Station to Ending Station) for the relocation service and the optimal number of discounted trips for each route are specified. (Phase 1) Because LA metro team cannot be sure how the operation cost will fluctuate after the launch of the service, this study addresses the range of operation costs which makes the relocation service remains profitable. (Phase 2)

B. Mathematical Programming

For each pair, parameters $Price[s,d]$ denote the **maximum discount dollar-amount** from the current price policy. $R_{discount} \in [0,1]$ is defined as the **discount rate** for the 'Walk-ups' and $Price[s,d] \times R_{discount}$ is defined as the **reward dollar-amount** for 'Passholders.' Then, this becomes the company's cost for relocation service. The model solver will determine **the amount of bikes to be relocated** over each pair, which will be represented as non-negative integer decision variables $x[s,d]$.

- The discount price for 'walk-ups' and the reward for 'passholders' are $Price[s,d] \times R_{discount}$
- By observing trade-off between discount rates and demands people will use Metro Bike, the realized demand rate is set as a function of the discount rate. Hence, the demand will become $Demand[d] \times f_{realized}(R_{discount})$
- The difference between realized demand and maximum demand (when discount rate is 1) is $(Demand[d] - x[s,d])$. This becomes cost as well.

The problem objective is to minimize the total cost from all supplying stations to all demanding stations.

$$
Minimize \sum_{s \in Supply} \sum_{d \in Demand} Price[s,d] \times R_{discount} \times x[s,d] + (price\_per\_mile) \times Distance[s,d] \times (Demand[d] - x[s,d]) \tag{1}
$$

$$
subject\ to \sum_{d \in Demand} x[s,d] \leq \text{Supply}[s],\ \forall s \in Supply \tag{2}
$$

$$
\sum_{s \in Supply} x[s,d] \geq \text{Demand}[d] \times f_{realized}(R_{discount}),\ \forall d \in Demand \tag{3}
$$

$$
x[s,d]\ is\ integer \tag{4}
$$

Eq. 1 : The first term denotes **Discount Cost** and the second term denotes **Operation Cost**
Eq. 2 : Relocation demands from all sources can not exceed the supplying capacity.
Eq. 3 : Relocations to each station must satisfy their demand with respect to the discount rate.

C. Network Assumptions

Before solving the optimization problem with Pyomo and gurobi solver, this network model needs some assumptions. The discount dollar-amount for non-passholders and reward dollar-amount for passholders are considered the same for model simplicity. Also, the optimal moves from supplying stations to demanding stations will always be met when the price is free to users (when the discount rate is 1). Lastly, the discount rate and the relocation demand percentage for each routes are the same. Which means, when the price is discounted by 10%, 10% of people on each route will use the relocation service.

D. Results

The *y-axis* of cost-graphs in fig. 5, 6 and 7 below represents **total cost**, sum of discount and operation costs, of various operation costs with respect to discount rates. The lowest points on each figure are the optimal discount rate at each cost.

1) Optimal discount rate to minimize the total cost: The average operation cost in LA area when moving each bike by operation team at LA Metro is \$1/mile. The graph presents discount rate of **25%** makes the least cost for moving unbalanced bikes a year. The detailed 7663 optimal rides for the relocation service are stored in 'optimal_sol.csv' as below.

Index	Supply	Demand	Optimal Moves
3	0	14	103.0
5	0	17	117.0
7	0	19	15.0
⋮	⋮	⋮	⋮
941	61	52	58.0
961	62	38	587.0

TABLE IV: Optimal Rides when Discount Rate is 25%

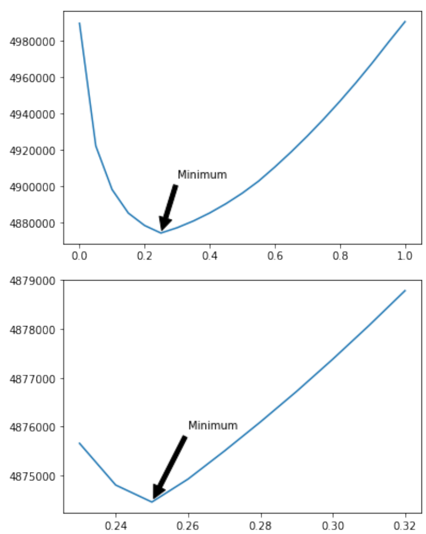

Fig. 5: Optimal Discount Rate when *price_per_mile* is $1/mile.
(Upper: 0% to 100%, LoIr: 23% to 32%)

2) Criteria on no need of the relocation service: The operation costs for the need of relocation service were examined by iterating 10 times from $0.00/mile to $0.45/mile. In fig. 6, there were monotone increase before $0.10/mile which means the least cost occurs when the *x-axis* is at 0 (the discount rate is 0%). Detailed cost-graphs between $0.10/mile and $0.13/mile are in the lower figure. **The minimum operation cost which makes the relocation service profitable is $0.11/mile.**

3) Criteria on the need of providing the relocation service at free price: The operation costs for the free relocation service were examined from $0.50/mile to $5.00/mile by iterating 10 times over each discount rates. In fig. 7 below, monotone decrease was observed after $3.50/mile which means the least cost will be obtained when *x-axis* is at 1 (the discount rate is 100%, free price to users). Detailed cost-graphs between $2.50/mile and $3.40/mile are in the lower figure. **The minimum operation cost at which the relocation service should be free is $3.20/mile.**

VI. Conclusion

From the results, the optimal discount rate was derived to be 25%, with the current price policy and operation cost, $1/mile. The optimal routes from the supplying stations to the demanding stations and optimal number of rides that will get discounts in each route are stored in output file 'optimal_sol.csv'. Although the analysis is based on the current operation cost, it is possible that the operation cost will fluctuate after the launch of the relocation service.
Hence, the range of operation cost needs to be proposed which makes the relocation service profitable. When the operation cost is less than $0.10/mile, the cost-graphs are consistently increasing and the lowest points are placed on 0%. This result is reasonable in that if the operation cost is affordable, there is no need of launching relocation service. On the other hand, when the operation cost is more than $3.10/mile, the cost-graphs are consistently decreasing and the lowest points are placed on 100%. This is also reasonable in that if the operation cost is too high, it is better to encourage people to use relocation service even for free. Finally, if the operation cost after the launch is between $0.10/mile and $3.10/mile, the service remains profitable and worth maintaining the service at proper discount rate.
The assumptions were needed for this problem formulation. With the utilization of more detailed dataset at LA Metro, this model can be more precise to reflect reality. Further studies will include the actual demand variations due to the discount rates and experiment on proper the 'discount rate' for non-passholders and 'reward-rate' for passholders.

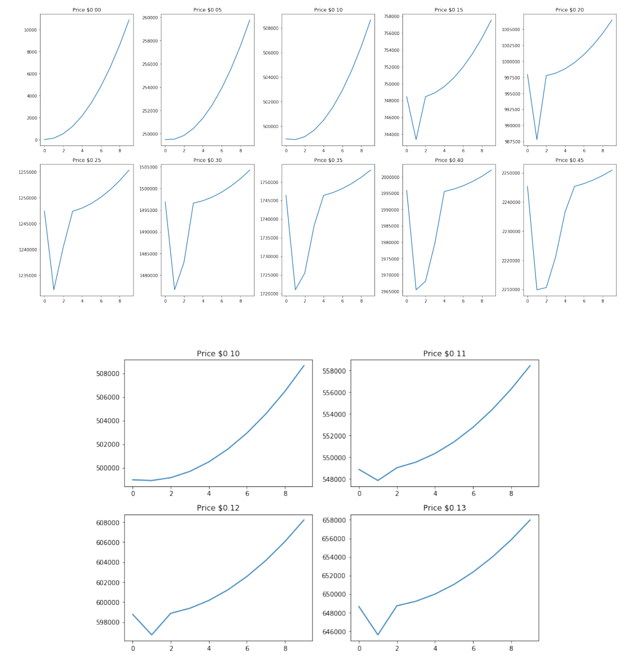

Fig. 6: Cost at which the relocation service is not needed
(Upper: $0.00 to $0.45, Lower: $0.10 to $0.13)

Appendix

The detailed codes and procedure are in the Github page.

References

[1] The data and price policy : https://bikeshare.metro.net/about/data/

[2] Diana Jorge, Gonalo Correia, Cynthia Barnhart, Testing the Validity of the MIP Approach for Locating Carsharing Stations in One-way Systems, Procedia - Social and Behavioral Sciences, Volume 54, 2012, Pages 138-148

[3] Daniel Freund, Ashkan Norouzi-Fard, Alice Paul, Shane G. Henderson, and David B. Shmoys. Data-driven rebalancing methods for bike-share systems. Working Paper, 2017.

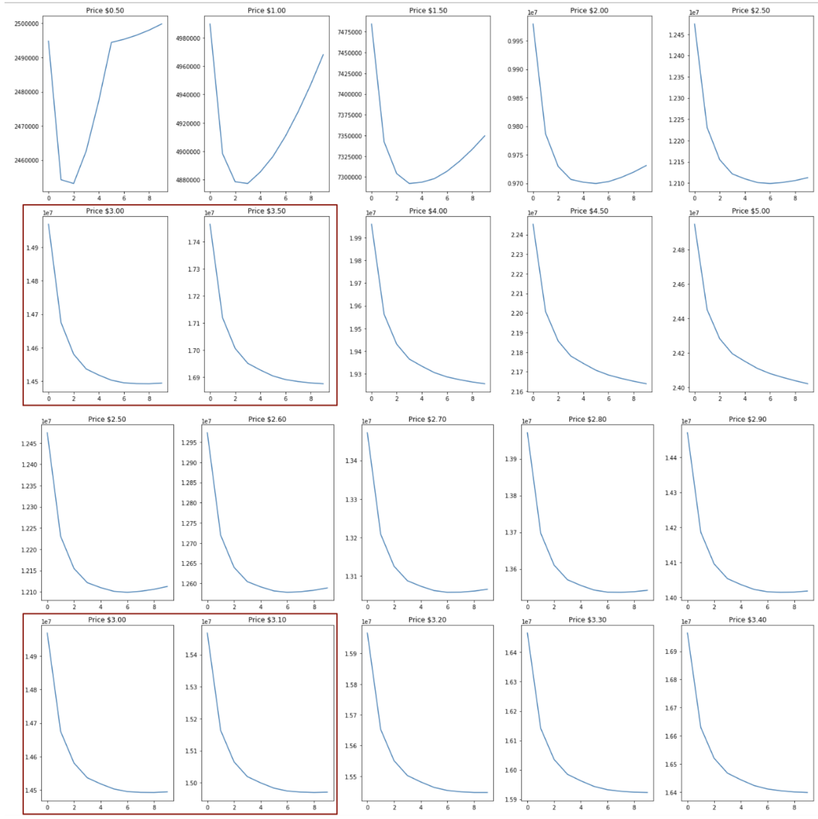

Fig. 7: Cost at which the relocation service is free
(Upper: $0.50 to $5.00, Lower: $2.50 to $3.40)