Visualization of the Mobility Patterns in the Bike-Sharing Transport Systems in Mexico City

L. A. Moncayo-Martínez[1], A. Ramirez-Nafarrate[1]
[1]Department of Industrial Engineering and Operations, ITAM, Rio Hondo 1 01080, Mexico City, Mexico
(luis.moncayo@itam.mx, adrian.ramirez@itam.mx)

Abstract - **This paper provides an analysis of mobility data of the bike sharing system in Mexico City from the year 2010 to 2015. Based on about 18 million trips, it is possible to compute the mean time of a trip and the number of trips from and to a single station. The 444 stations were classified according to the average number of input and output bikes using a Pareto chart. The patterns of mobility between the stations are shown graphically; thus, some clusters are identified based on the number of arrivals and departures. Finally, the number of input and output for some stations is plotted in ten-minute intervals for weekdays and weekends. Therefore, it is known the flow of bikes in stations per time interval per day. These patterns are applied to predict the behavior of a particular station to define policies to improve the bike system program. An R script, available to the public, is programmed to manage the amount of data and to carry on the analysis.**

Keywords - **bike sharing systems; mobility patterns; clustering; Ecobic**

I. INTRODUCTION

The bike-sharing transport systems are becoming more and more popular in the last few years. Until June 2016, there were about 1000 cities worldwide [1] that have a kind of this system. It has been estimated that there are about 375 bike-sharing schemes. Usually, a person can rent a bicycle at one station, from many around an area, then leave it in another station after a short trip, usually minutes. These systems have a positive impact on both the people who use it and the environment.

People are adopting this mode of transport in densely populated cities because its flexibility, the impact on health, low cost, and benefit of avoiding traffic. A study carried out in 2006 by OYBike in London showed that 40% of users shifted from a motorized transport mode to a bicycle [2]. About sustainability issues, bike-sharing systems allow people to reduce their taxi trips and replace their cars, which reduces the CO2 emissions [3].

In recent surveys, such as [4], [5], and [6], four research areas have been identifying:
a) Strategic decision. The problem is to develop methodologies to implement and expand such systems. Two issues are the location of possible stations given a potential demand and the development of business models to trade-off the interest of users and investors.
b) Service level analysis. The service level is measured as the accessibility to a bike and a spot to drop-off it. So as to compute the service level, authors have modelled the system as a Markov chain or a queue system, see [7] and [8].
c) Rebalancing operations. Due to the imbalance between supply and demand of bikes in stations, there is a need to redistribute them. i.e. they are transferred from a location where there are available bikes to another where there are no enough bikes.
d) Demand Analysis. The purpose is to forecast the demand and to understanding the behavior of the systems to support the making decision process. This stream of research provides insights that are helpful in improving the service levels requirements. Moreover, analysis of demand and its patterns is the first step in the analysis and optimization of this transport system, e.g. Citybike Wien [9], Bicing [10], and NYC (Citi)Bike [11].

Our work is related to the demand analysis of the bike sharing systems (called EcoBici) in Mexico City. Although, the number of users of the EcoBici, it is unknown the mean travel time of a trip, the classification of stations based on the number of departures and arrivals, the interaction between station (i.e. clusters of stations), and the number of output and input bikes per ten-minute interval per day [12].

To answer those questions, we programmed an R script. The script downloads the data from the Data Lab of Mexico City [12] or can read the data included in the source code which can be downloaded from the following link:
https://dl.dropboxusercontent.com/u/6629477/EcobiciCDMX.zip. In this work, we aim to contribute to the understanding of the behavior of a bike-sharing system via the analysis of the mobility patterns.

According to this patterns, the mean time of a trip is 12.68 minutes. Station 27 receives the most number of trips and most trips depart from it. Four clusters are identified based on the number of incoming and outgoing bicycles.
We compute by ten-minute intervals the number of incoming and outgoing bicycles. In Station 27, the mean value of incoming bicycles is 18 and outgoing 13 on Friday. On Sundays, those values are 13 and 7, respectively.

the patterns of the bike-sharing systems of Mexico City. Moreover, an objective of this work is to describe this system to allow the researchers and practitioners to apply complex techniques (data mining, linear programming, or queue theory) to balance the systems based on a given service level such as [6] and [19].

II. THE ECOBICI SYSTEM

In February 2010, the city council established the first bike-sharing system in Mexico City to be part of the intermodal city transport system (subway – microbus – metro bus – car). At the beginning, there were 84 stations and 1200 bicycle. Today, there are 452 stations and 6000 bicycles. Until June 2016, the system has registered about 33.5 millions of trips.

Some of the benefits are the reduction of the travel time from their origin to destination, reduction of CO2 emissions, and effects on users' health. The 62% of the trips begin at user's home, and 45% ends at a workplace. 87% of the users use a bicycle and other transport to reach their destination (35% walk and 29% use subway). The rest of users (13%) use only a bicycle to reach their destination. The main reasons a person uses a bicycle in Mexico City is to exercise (60%), to avoid traffic (30%), and to arrive faster to his/her destination (28%) [13].

In the following sections, we describe the Ecobici systems using about 18 millions of trips from January 2010 to January 2015 (this period is available for public use, see [2]). To do the following plots, we use R (version 3.2.2), RStudio (version 0.99.486), and a MacBook Pro with an Intel Core i7 processor (2.2 GHz) and 16GB of RAM memory.

The data provided by Mexico City Lab [12] is divided by months. Each month has information organized in ten columns with the following information: id (or number of a trip), user gender, user age, bicycle id, departure station, departure date, departure time, arrival station, arrival date, and arrival time, respectively.

A. Travel Time

The travel time is computed without restricting the travel time and is computed restricted the travel time to 200 min, see Table I. In both cases, the median is 10.10 min. However, there is a one-minute difference in the mean of a travel time (13.60 and 12.68, respectively). As shown in Table I, if the travel time is not restricted, there is at least one trip of about 336 days. Thus, the travel time was restricted to 200 minutes. In this case, we can assure that the mean time of a trip is of 12.68 min and the median is 10.10 minutes. A histogram is plotted in Fig. 1 to show the distribution of the travel time of every single trip from January 2014 to January 2015. As shown in Fig. 1 most of the trips ranges from 0 to about 60 min. As shown as well, trips that last more than 60 min are not common.

TABLE I. SUMMARY OF TRAVEL TIME FROM JAN 2014 TO JAN 2015

Measure	*Time (min) No limits in travel time*	*Time (min) Restricted to 200min*
Minimum	0.00	0.00
1st Qu.	6.30	6.30
Median	10.10	10.10
Mean	13.60	12.68
3rd Qu.	16.20	16.10
Maximum	**483900**	**200.00**

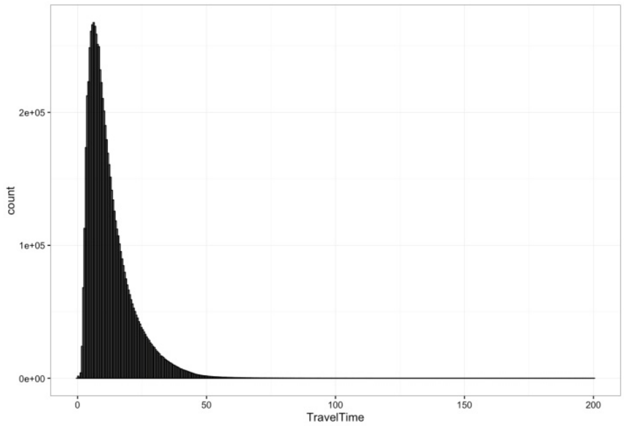

Fig 1. Number of trips based in the travel time

B. Inbound and Outbound Activity per Station

An important characteristic of the system is the activity of every single station. Therefore, the number of inbound and outbound trips is counted separately from January 2014 to January 2015. The number of trips are 7'178'246 and some of the results are shown in Table II.

According to Table II, 1.49% of the trips begin in station 27 and 1.52% of them ends in the same station. On the other hand, 0.04% of the trips departure from station 94 and 0.03% of the trips end in station 94. Four stations (27, 36, 41, and 1) are above the 1% of the departure trips; thus, 4.64% of the trips departure of these four stations. The same number of stations (27, 43, 1, and 36) receive 4.73% of the total number of trips. The percentages of inbound and output trips for stations 27, 36, and 1in Table II are similar but for stations 43 and 41. With the percentage data in Table II, a flow index (ϕ) per stage i that measure the relation of inbound and outbound trips can be computed using Eq. 1.

$$\phi_i = \text{\% inbound trips / \% outbound trips} \quad (1)$$

Clearly, if $\phi_i > 1$, the station i receives more bicycles (assuming one trip per bicycle). Otherwise, when $\phi_i < 1$ more bicycles depart from the station. In case $\phi_i = 1$, the station is in "balance". Those results are plotted in Fig. 2.

TABLE I. % OF INBOUND AND OUTBOUND TRIPS FOR SOME STATIONS

Station (i)	*% inbound trips*	*Station (i)*	*% of outbound trips*
27	1.5215	27	1.4906
43	1.1117	36	1.0857
1	1.0649	41	1.0284
36	1.0387	1	1.0132
64	0.9693	18	0.9748
18	0.9381	43	0.9543
…	…	…	…
275	0.0890	275	0.0842
100	0.0686	220	0.0637
220	0.0521	100	0.0599
94	0.0394	94	0.0321

As shown in Fig. 3, station 193 has the lowest flow index (0.5105); thus, from this station more bicycles depart than those that arrive. Therefore, stations with an index below the black line in Fig. 3 are "supply" stations. On the other hand, stations with index flow value above the black line are "demand" stations, e.g. stations 175 and 267 with the highest index flow, 1.6603 and 1.6807, respectively.

As a result of this section, the stations that receive and depart most bicycles are identified; thus, stations 27, 36, 1, 43, 41, and 18 seem to be those stations, see Table II.

In Fig. 2, the stations are classified as supply or demand stations. The flow index for station 27 is 1.0207; it suggests that the station is in balance. Another example is station 275, its index flow is 1.0567, that also suggests a balanced station. Nevertheless, station 27 is a very intensive traffic one (about 1.5% of the arrivals and departures occurs in it). Meanwhile, station 275 is a very low traffic station given that this station receives and departs about 0.08% of the bicycles, see Table II.

C. *Traffic among Stations*

In Fig. 5 the number of trips among stations is plotted. The aim of this graph is to show the number of from-to trips and to identify some clusters based on traffic. The darker areas in Fig. 5 suggest higher traffic.

The darkest area is the one in the bottom-left corner. Therefore, between station 1 and 80, the traffic is very intensive. One important characteristic is that it seems there are many trips that start and end in the same stations, as shown by the dark area formed by the diagonal line from the bottom-left to the upper-right corner.

The lower part of Fig. 5 shows that from stations 1 to 80, a lot of bicycles depart to all the stations given the darker areas. Even though, four clusters are identified based on the traffic activity represented by the dark areas:

- Stations 1 to 80
- Stations 110 to 170
- Stations 190 to 250

To classify the other stations, advanced clustering techniques should be applied. On the other hand, there are light spots which indicate inactivity among those stations. For example, the traffic is very little between stations 100 – 180 and 200 – 270. Other light spots are identified such as the one from stations 160 – 220 to stations 80-110 but above mentioned is the biggest one

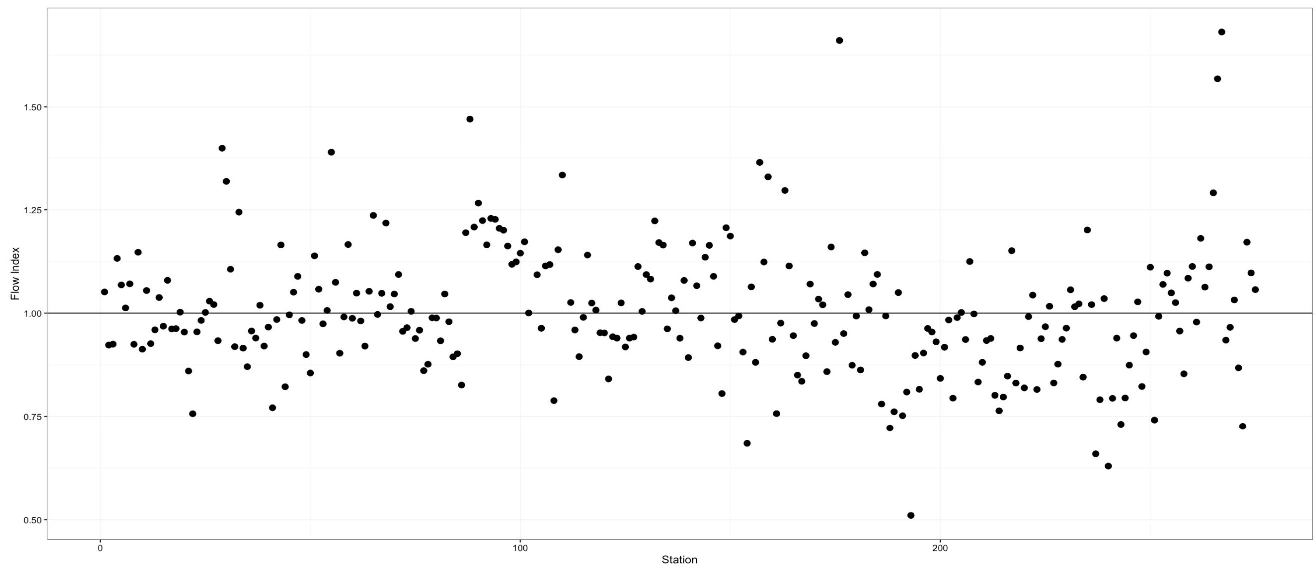

Fig. 2. Fow index (Eq. 1) per stage.

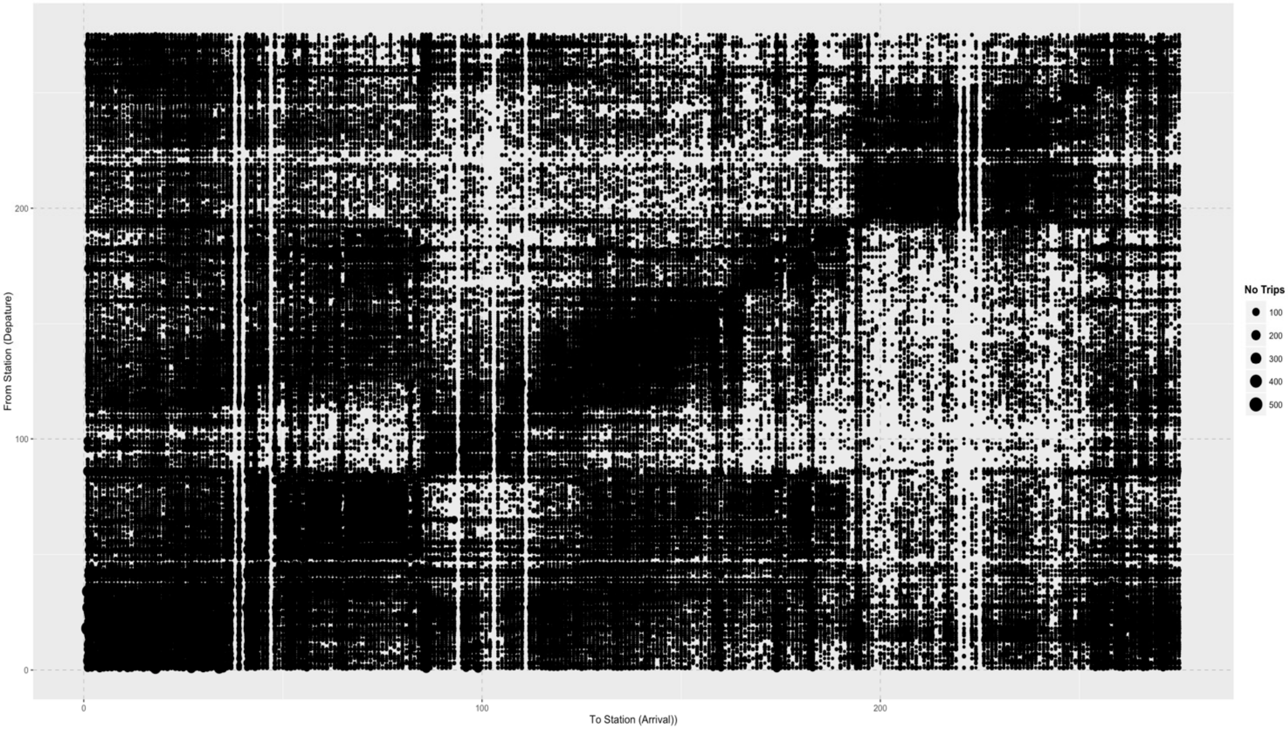

Fig. 3. Number of trips among stations in January 2015

D. *Ten-minute Interval Activity per station*

Fig. 2 and 3 shows the traffic among stations without taking into account the day of the week and the time in which an arrival or a departure takes place. Hence, every day of the week is divided into intervals of ten minutes, so a detailed view, of the number of incoming and outgoing bicycles, is provided.

In this work, we use station 27 as an example because it receives the highest percentage of incoming and outgoing trips (see Table II and Fig. 2). Moreover, we plot the data of Friday and Sunday given that the first one is the day with the highest traffic of the week; while Sunday is the day with less traffic day.

If the reader wants to do a similar analysis for any day of the week, for any station, and for anytime, s/he is encouraged to download the R script from the link provided in section I.

In Fig. 4, the median value of all inbound (dotted-blue line) and outbound (solid red line) trips done on Fridays from Nov 2014 to Jan 2015 is plotted. For example, the median value of the number of departure bicycles on Fridays from 18:09:50 to 18:19:59 is 18 bicycles. Meanwhile, 13 bicycles arrive at the same time on Fridays. Therefore, at this time more bicycles depart than those that arrived. On the contrary, from 08:49:59 to 08:59:59, **13** bicycles arrive and **7** depart. We can see that from 05:39:50 to 12:09:59 the number of arrivals is bigger than the number of departures. From 12:09:59 to midnight most of the time there are more departures than arrivals. This suggests that this station 27 is "balance"; thus, its

flow index is near to one (ϕ_{27} = 1.0207). We can see that on Fridays the traffic is heavy from 14:29:59 to

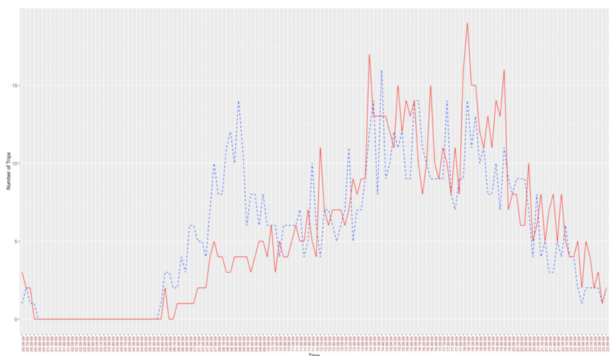

Fig. 4. Meadian of arriving bickes (dotted-blue line) and of departing bicycles (solid red line). All **Fridays** from Nov 2014 to Jan 2015

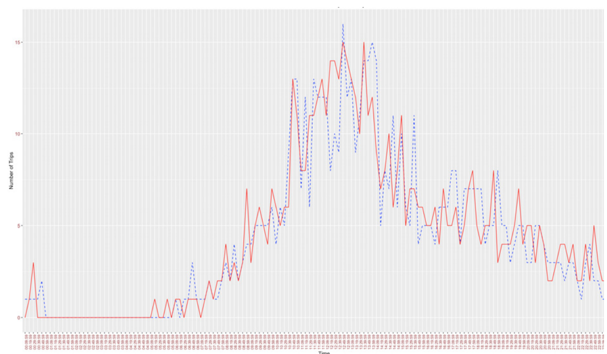

Fig. 5. Meadian of arriving bickes (dotted-blue line) and of departing bicycles (solid red line). All **Sundays** from Nov 2014 to Jan 2015

21:09:50.

The median value on Sundays is computed for the same period and stations, the results are plotted in Fig. 5. On Sunday, the traffic is heavy from 10:29:59 to 16:19:59 and it seems that the number of arrivals and departures follows the same pattern. Therefore, the station seems to be balanced. On Sundays, the number of departures and arrivals are about **15** bicycles at 12:59:59.

According to Fig. 4 and 5, we can recommend having 5 bicycles without any analytical technique. We think that if we draw a horizontal line in 5 the area below this value and the area above is very similar; thus, we could balance the number of stock out and surplus units

IV. DISCUSSION AND CONCLUSIONS

According to the Figures and Tables presented in the previous sections, we conclude that:

- The meantime, a user uses a bicycle, is 12.68 minutes (Table I).
- Station 27 receives the largest number of bicycles. The largest number of bicycles ends in this station (Table II).
- According to Table II, most of the stations receive less than 0.5% of the total number of bicycles. This proportion is the same for the number of arrivals.
- According to the flow index (Eq. 1), stations could be divided into "supply" and "demand" stages. Stations with index $\phi < 1$ (Fig. 2) need bicycles and stations with $\phi > 1$ hold bicycles. We can use this information to rebalance the systems.
- From Fig. 3, three clusters are identified based on the from-to matrix. We conclude that many trips begin and end in the same station. Manager could focus on the management of the stations given that they know where the trips begin and end.
- Computing a plot such as Fig. 4 and 5, we could compute the amount of bicycles in stock. Those plots could be done using our R script.

After this study, the bike-sharing transport system is described not only by the users' profile but also by the moving patterns among stations. Our aim is to provide to scientist and practitioners of the moving patterns to develop analytic method for mainly rebalance the systems.

For the theoretical point of view, we conclude that a single station could need or store bicycles during the day. This is important since most theoretical models suppose a station have the same necessity of excess of bicycles for a long period, usually six-hour period.

From the practical point of view, the managers must take into account that a trip lasts about 12 min. Thus, we suggest that every 12 min the systems must be balanced

ACKNOWLEDGMENT

"Financial support from the Asociación Mexicana de Cultura, A.C. is gratefully acknowledged". The National Council of Science and Technology (CONACyT) from México.

REFERENCES

[1] "Bike-sharing Blogsplot," 2016. .
[2] R. Noland and M. Ishaque, "Smart Bicycles in an Urban Area: Evaluation of a Pilot Scheme in London," *J. Public Transp.*, vol. 9, no. 5, pp. 71–95, Dec. 2006.
[3] J. Bachand-Marleau, J. Larsen, and A. El-Geneidy, "Much-Anticipated Marriage of Cycling and Transit," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2247, pp. 109–117, Dec. 2011.
[4] E. Fishman, "Bikeshare: A Review of Recent Literature," *Transp. Rev.*, pp. 1–22, Apr. 2015.
[5] E. Fishman, S. Washington, and N. Haworth, "Bike Share: A Synthesis of the Literature," *Transp. Rev.*, Apr. 2013.
[6] J. Schuijbroek, R. Hampshire, and W.-J. van Hoeve, "Inventory Rebalancing and Vehicle Routing in Bike Sharing Systems," *Tepper School of Business*. 2013.
[7] R. Nair and E. Miller-Hooks, "Fleet Management for Vehicle Sharing Operations," *Transp. Sci.*, vol. 45, no. 4, pp. 524–540, Nov. 2011.
[8] T. Raviv, M. Tzur, and I. A. Forma, "Static repositioning in a bike-sharing system: models and solution approaches," *EURO J. Transp. Logist.*, vol. 2, no. 3, pp. 187–229, Jan. 2013.
[9] P. Vogel, T. Greiser, and D. C. Mattfeld, "Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns," Cornell University, 2011.
[10] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervasive Mob. Comput.*, vol. 6, no. 4, pp. 455–466, Aug. 2010.
[11] E. O'Mahony, "Smarter Tools For (Citi)Bike Sharing," Cornell University, 2015.
[12] "Laboratorio de Datos CDMX." 2016 .
[13] R. Perez-Lopez, "Encuesta Ecobici 2014," 2015.