Citi Bike Demographics and Marketing Strategy Suggestions for New York City between August 2015 and October 2015

Data Bootcamp

Professor Benjamin Zweig

December 20th, 2018

Charles Chansa, Cecilia Figueroa Arrivillaga, and Martin Smit

Table of Contents

Research Question

What demographic profile of Citi Bike Users should Citi Bike target in order to maximize profits?

Abstract

The purpose of this study is to identify who Citi Bike's most valuable subscribers are. The study makes an in-depth study based on gender and age for a period of three months during the year 2015. Since Citi Bike's is based on subscription with a yearly fee fixed at $149 (3-month fee of $37.25), it follows that the population groups that use the Citi Bike's the least are not maximizing the utility they could have from using the bikes or in other words, aren't getting a bang for their buck. Therefore, it makes sense that Citibike's marketing should focus on the population groups that use the Citi Bike service the least relative to the number of people in that demographic that have subscribed to the Citi Bike's monthly service.

We came to the conclusion that Citi Bike should mainly focus on marketing their program to males between the ages of 18-30 and above 80 given that these users experience the highest average cost per minute from riding. If Citi Bike attracts customers who use the program least, they will be able to spread their assets amongst more customers. Unfortunately, our OLS regressions were unable to provide significant correlations between variables in our data and therefore, we recommend a further investigation with a larger dataset and more demographics to improve the test significance of the OLS regression.

Code

Our code can be found through the following links to our Github repositories:

https://github.com/CeciliaFigueroaA/Data_Bootcamp_Final_Project

https://github.com/smitm1997/Data_Bootcamp_Final_Project/tree/master

https://github.com/kchansa/Data_Bootcamp_Final_Project

Background Information - Citi Bike

Citi Bike prides itself on being "the largest bike share program" in the United States. It started operating in Manhattan, Brooklyn, Queens, and Jersey City in May of 2013 with 6,000 bikes and 332 stations[1]. It has now expanded to a total of 12,000 bikes and 750 stations across these same boroughs and plans to triple its number of bikes in the next 5 years[2].

The service operates under 3 main types of passes or plans. Users have the option of buying a Single Ride Pass ($3/ride), a Day Pass ($12/day or $24/ 3 days), or an Annual Membership ($169/year). Rides are limited to 30 minutes for single-ride and day users, and to 45 minutes for annual membership users. Riders can get a bike from any station while their pass is active and return it in any other station[3].

Moreover, this service is part of a highly competitive market. In New York City, there are several transportation options that generally serve similar customer segments. Among these, major ones are Uber, Lyft, the New York City Subway and Bus services, and yellow taxis, among others.

Considering this competition, and Citi Bike's remarkable growth since it's launch in 2013, we will consider Citi Bike's riders in order to understand the company's current market segment and possible areas for growth in terms of customer segmentation. In particular, with this research we aim to answer the following research question: "Who should Citi Bike market to?"

Data Understanding and Preparation

In order to do this, we downloaded datasets from Kaggle.com about Citi Bike rides in New York City.

The datasets included information on the gender of riders, their age, the start and stop time of each ride, whether each rider had a subscription or not, and more.

The information was only available on a monthly basis. Therefore, we merged 3 of the monthly databases to create one database with information from August to October of 2015. We

[1] https://www.citibikenyc.com/how-it-works
[2] https://www.citibikenyc.com/blog/citi-bike-is-going-to-dramatically-expand

[3] https://www.citibikenyc.com/pricing

chose to work only with data from these months because we consider weather to be more favorable for riders. Furthermore, doing this rather than working with yearly data allowed us to reduce the amount of data from about 9.6 million data points to under 4 million. We did this by using a for loop where each monthly database was opened and appended to the data frame "Citi".

In order to further clean and prepare our data, we sorted the "Citi" data by "starttime", beginning with the earliest trips and dropped the rows in each data set that had missing values for "starttime" and "stoptime", allowing for a smooth analysis ahead.

Unfortunately, the data provided by the Citi Bike platform does not provide user identification for subscribers. Therefore, we are restricted in our analysis. To facilitate our investigation and gather data on the average Citi Bike rider, we need the number of subscribers for 2015. We were able to find the number of subscribers to Citi Bike in 2016 and the increase in Citi Bike trips between 2015 and 2016[4]. We have used this 40% growth rate in trips to predict the amount of Citi Bike subscriptions in 2015. We have estimated the number of Citi Bikes yearly subscriptions in 2015 to be 71,500 (for calculations see *Exhibit 1).*

Findings on Citi Bike

According to the data, there were 3,942,470 trips in the months of August, September, and October of 2015. Restricting our data to only the quarter of the year where we believed Citi Bike usage would be least affected by climate conditions allowed us to work more efficiently with a smaller dataframe. With this information, we decided to study the demographics of these rides more closely. We particularly focused on differences in riders as related to age group and gender. Due to the availability of information, we had to study gender as a division between male and female users.

We structured our analysis by looking at differences in gender first, then focusing on differences in age groups, and finalizing with a conjunctive analysis.

1. *Citi Bike and Gender*

[4] https://www.nytimes.com/2017/07/30/nyregion/new-yorkers-bike-lanes-commuting.html

We first calculated the average time duration of each ride to be 736 seconds or approximately 12 minutes. Moreover, this average time varies depending on gender, as shown in **Figure 1**.

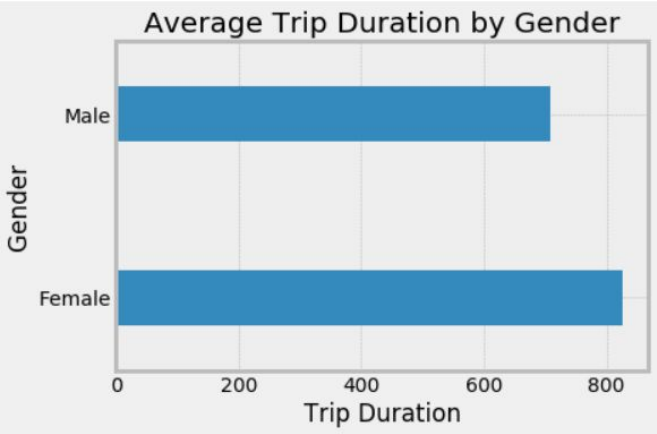

Figure 1

However, although females take longer rides on average, **Figure 2** shows that men use the Citi Bike service more than 3 times as much as women. We measured this usage in terms of the number of trips taken by male vs female riders.

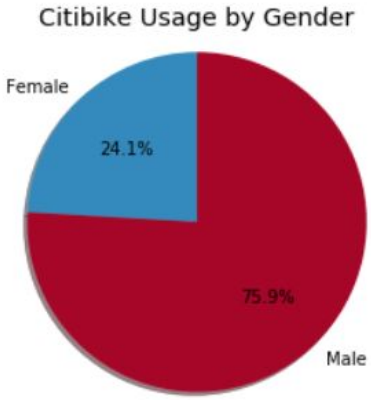

Figure 2

In our gender analysis, we also wanted to have an understanding of how male and female riders differ in different areas where Citi Bike operates. To do this, we created the heat map in **Figure 3**. This map shows which areas across Manhattan, Brooklyn, and Queens have more

starting points for rides by males vs. females. It is interesting to note that in areas that are outside of Manhattan, starting points by males are much more common than those by females. This disparity is less clear in what seems to be midtown and downtown Manhattan. Regardless, There is still a clear dominance of males because of the overwhelming size of the red circles respective to the blue ones.

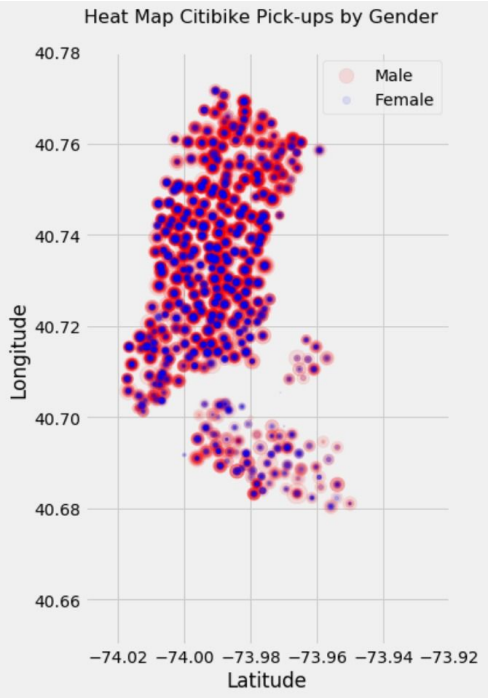

Figure 3

2. *Citi Bike and Age Groups*

Next, we conducted a similar analysis for different age groups represented in Citi Bike riders. We started by looking at the average trip duration by age group. Recall that the overall average trip duration was about 736 seconds, or 12 minutes. By looking at **Figure 4**, we can see that the age groups that are most significantly below average seem to be from 18-30, and from 82-90. In other words, these age groups may be more convenient users for Citi Bike since they use each bike for less time on average.

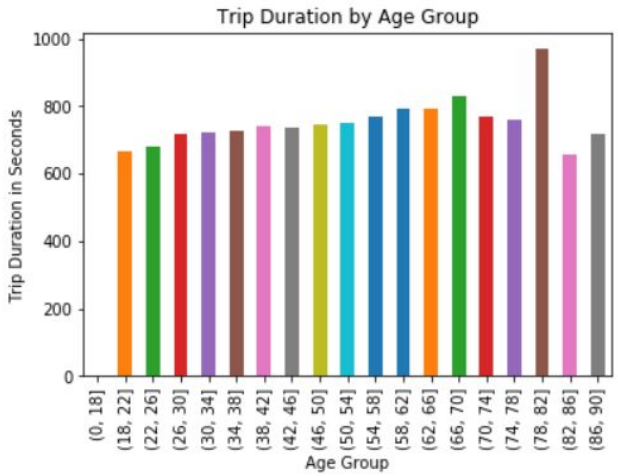

Figure 4

Moreover, we also looked at Citi Bike usage for different age groups. **Figure 5** shows the number of rides each age group took. We see that the most common users of this service are in the 30-34 age group.

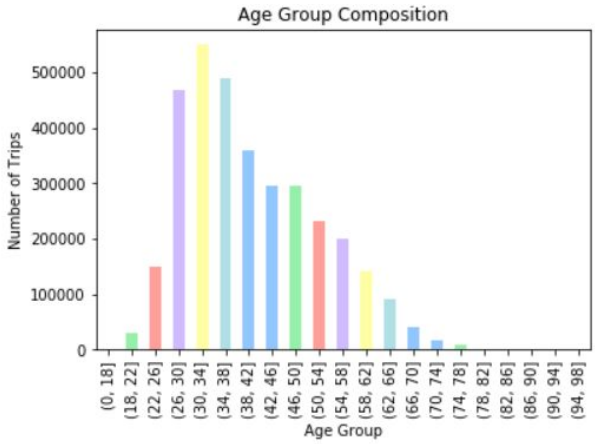

Figure 5

3. *Conjunctive Analysis*

We then used the previous information as the first step to analyze both gender and age together. We wanted to understand which gender and what specific age group would potentially be more profitable or convenient for Citi Bike to target.

To do this, we first looked at differences in gender within each of the age groups. **Figure 6** shows the ratio of men to women in each age group in terms of total rides. Below that, **Figure 7** shows the differences in the length of rides for both men and women in each age group.

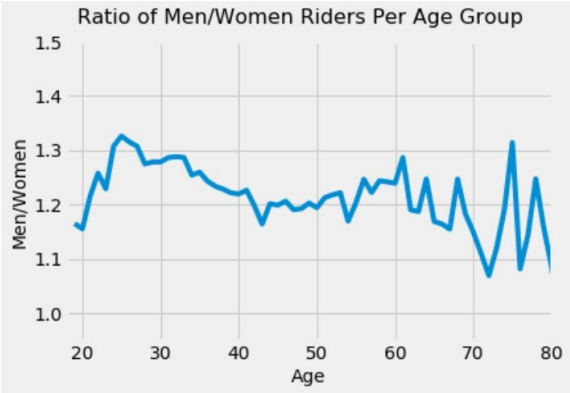

Figure 6

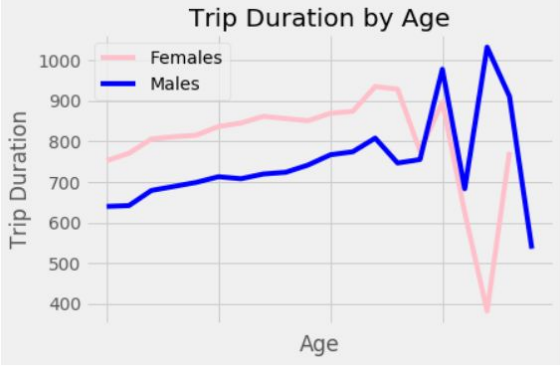

Figure 7

Lastly, as part of our analysis we also calculated the costs of different customer segments. In order to do this, we first created 2 dataframes: one for age groups and one for gender. In these data frames, we included the sums of trip duration for each group, the number of individual data points per group, and the average time in minutes riders used Citi Bike per capita per trip in each group. These data frames are shown in **Figure 8** and **Figure 9**.

Gender	sum_duration	count	duration_capita	cost_capita
Female	670605848	811170	39033.707008	0.057258
Male	1809229479	2556766	33410.810800	0.066895

Figure 8

Age	Sum Duration	Count	Duration Per Capita	Cost Per Capita
(0, 18]	0	0	NaN	NaN
(18, 22]	19413981	29153	31442.400098	0.071082
(22, 26]	102345843	150166	32179.770100	0.069454
(26, 30]	334952362	468224	33776.427749	0.066170
(30, 34]	398047114	550314	34151.380207	0.065444
(34, 38]	356793551	490350	34355.411786	0.065055
(38, 42]	266437785	359939	34950.327599	0.063948
(42, 46]	217672946	296277	34688.926380	0.064430
(46, 50]	220924758	295664	35280.138541	0.063350
(50, 54]	173443014	230895	35467.177261	0.063016
(54, 58]	152783232	199241	36206.064618	0.061730
(58, 62]	112295491	141826	37384.433675	0.059784
(62, 66]	71533896	90126	37475.386839	0.059639
(66, 70]	34652630	41678	39256.689524	0.056933
(70, 74]	12986034	16910	36259.120869	0.061640
(74, 78]	7266648	9576	35828.970376	0.062380
(78, 82]	1644655	1697	45759.082655	0.048843
(82, 86]	300469	458	30975.514997	0.072154
(86, 90]	43074	60	33895.987856	0.065937
(90, 94]	0	0	NaN	NaN
(94, 98]	0	0	NaN	NaN

Figure 9

We then used this information to calculate the average cost of a ride per minute per person. We did this by dividing the total fee of their subscription ($37.25 for 3 months) by the average amount of minutes per ride per capita in each group. We then plotted this information in order to understand it visually. **Figure 10** shows the cost per minute for a riders who are male or female. **Figure 11** shows the same information, but for each individual age group.

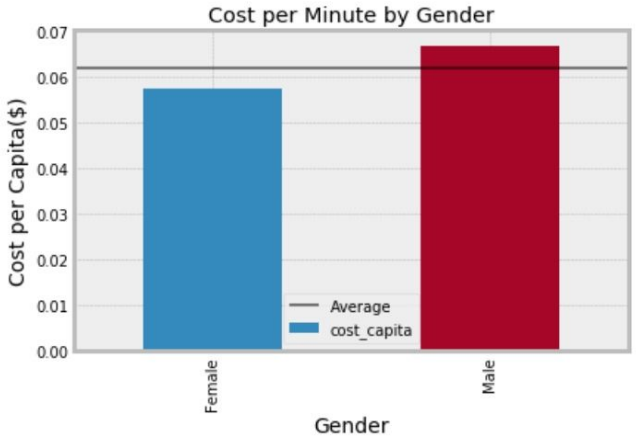

Figure 10

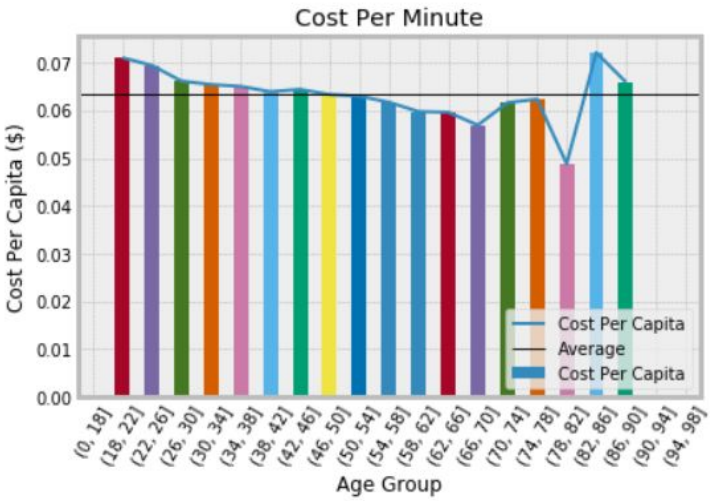

Figure 11

From these graphs, we can see that males pay a higher price per minute than females. We also calculated an OLS regression on this data to quantify the difference and found that for men, each ride costs approximately 1 cent on average more than for women (calculations in *Exhibit 2*). Likewise, the age groups in the extremes of **Figure 11** also pay a higher price per minute than average. We ran a similar regression for age that is included in (*Exhibit 3).* This is useful information when considering who Citi Bike should target as customers because it gives insight into which customers are "cheaper" for Citi Bike to support.

Conclusion

By analyzing the data on Citi Bike rides in New York City for the months of August to October of 2015, we were able to clearly see a difference in the company's customer demographics in terms of both gender and age group. Overall, it seems like Citi Bike's current market is focused on men of ages 26 to 42.

However, after analyzing the value of each type of customer by looking at the time duration of their trips and their average number of trips, we were able to determine the cost per minute of each type of customer. In other words, this cost represents the amount of revenue per minute Citi Bike can get from each different customer type. With this analysis, we realized that Citi Bike should aim to attract more male customers and customers who belong to the extremes

of the age group spectrum in order to attract more revenues per person. This is because, as seen in **Figures 10** and **11**, males, along with riders who are younger than 42 but older than 82, pay the highest price per minute as a result of using the service less on average.

However, there are certain limitations in this analysis that must be acknowledged and addressed. Firstly, since it was not possible for us to work with data for the entirety of the year, we were limited to information on only one season of the year. Therefore, these observations may be affected by outside factors such as weather, differing tourism levels in the city, demand, etc. It would be necessary to make similar observations taking into account the rest of the year in order to make a more objective conclusion.

Moreover, we realize that there are greater complexities when considering who the best segments to market Citi Bike to are, such as the market size. For example, it may be important to take into account that there are more women than men in New York City[5], and therefore, limiting the targeting of women may risk the company's total outreach in terms of market. This is also why it is important to consider if Citi Bike is more interested in achieving a larger market share in the transportation market, or simply increase revenues.

Lastly, another thing we were not able to consider with the data we have available is the impact of external factors such as safety within New York City. For example, the differences in density in different boroughs as seen in **Figure 3**, may be the result in significant demographic differences within each respective area.

[5] https://www.states101.com/gender-ratios/new-york

Appendix

Exhibit 1:

Data:

Citibike subscribers 2015 is 100,000
Citibike trips in 2014 is 10 million
Citibike trips in 2015 is 14 million

$$Trip\ growth\ rate_{2016} = \frac{14M}{10M} = 40\%$$

Assume 40% growth in Citibike subscribers:

$$40\% = \frac{subscribers_{2016}}{subscribers_{2015}} = \frac{100{,}000}{subscribers_{2015}}$$

Therefore,

$$subscribers_{2015} = 71{,}429$$

Exhibit 2:

OLS Regression Results

Dep. Variable:	cost_capita	R-squared:	1.000
Model:	OLS	Adj. R-squared:	nan
Method:	Least Squares	F-statistic:	0.000
Date:	Thu, 20 Dec 2018	Prob (F-statistic):	nan
Time:	17:56:45	Log-Likelihood:	75.265
No. Observations:	2	AIC:	-146.5
Df Residuals:	0	BIC:	-149.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0573	inf	0	nan	nan	nan
Gender2	0.0096	inf	0	nan	nan	nan

Omnibus:	nan	Durbin-Watson:	0.200
Prob(Omnibus):	nan	Jarque-Bera (JB):	0.333
Skew:	0.000	Prob(JB):	0.846
Kurtosis:	1.000	Cond. No.	2.62

Exhibit 3:

OLS Regression Results

Dep. Variable:	tripduration	R-squared:	0.011
Model:	OLS	Adj. R-squared:	0.011
Method:	Least Squares	F-statistic:	1.927e+04
Date:	Thu, 20 Dec 2018	Prob (F-statistic):	0.00
Time:	18:10:50	Log-Likelihood:	-2.6006e+07
No. Observations:	3372554	AIC:	5.201e+07
Df Residuals:	3372551	BIC:	5.201e+07
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	483.6462	1.434	337.383	0.000	480.837	486.456
Age	2.4898	0.026	96.298	0.000	2.439	2.540
gender	121.5974	0.686	177.197	0.000	120.252	122.942

Omnibus:	1883764.503	Durbin-Watson:	1.916
Prob(Omnibus):	0.000	Jarque-Bera (JB):	24671729.627
Skew:	2.431	Prob(JB):	0.00
Kurtosis:	15.326	Cond. No.	218.

.. .

Works Cited

"Citi Bike." Wikipedia, Wikimedia Foundation, 6 Dec. 2018, en.wikipedia.org/wiki/Citi_Bike.

Hu, Winnie. "More New Yorkers Opting for Life in the Bike Lane." The New York Times, The New York Times, 30 July 2017, www.nytimes.com/2017/07/30/nyregion/new-yorkers-bike-lanes-commuting.html.

Levy, Nicole. "These Are the Most Popular Citi Bike Routes by Age and Gender." DNAinfo New York, DNAinfo New York, 27 Feb. 2017, www.dnainfo.com/new-york/20170227/midtown/citibike-cycling-route-gender-age-demographics/.

Motivate International, Inc. "Citi Bike Is Going to Dramatically Expand!" Citi Bike NYC, www.citibikenyc.com/blog/citi-bike-is-going-to-dramatically-expand.

Motivate International, Inc. "Citi Bike: NYC's Most Popular Bike Rental Program." Citi Bike NYC, www.citibikenyc.com/how-it-works.