A Geospatial and Hypothetical Analysis of Capital Bikeshare (CaBi)

Tegveer Ghura, Ahmed Khair, Anthony Moubarak, Raghav Sharma, Karan Uppal

ANLY 511 Final Project

Georgetown University

11th December, 2022

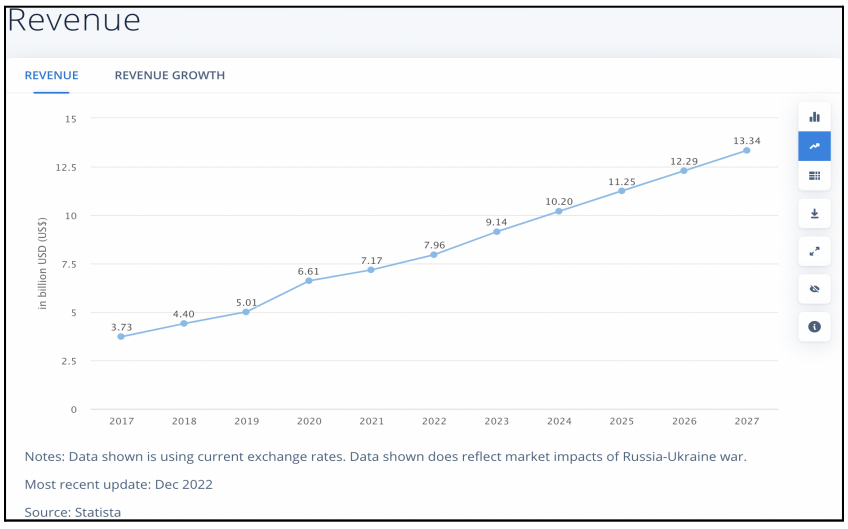
Table of Contents

I. Introduction

In 2008, The District of Columbia (D.C) became the first city in America to create a bike-sharing platform named SmartBike DC (Capital Bikeshare, n.d). Similarly, Arlington County was also working on launching its own bikeshare system by joining forces with D.C, Alexandria, VA, and Montgomery County, MD. To further their efforts, in 2010, Arlington County and The District of Columbia chose an operator and their expanded efforts brought forth Capital Bikeshare to serve the Washington DC metro area with a fleet of 5000 bikes and 600 stations (Capital Bikeshare, n.d).

With rising fuel costs and weather crises, capital bike-share can be seen as an alternative method of transportation as it's cost-effective and environmentally friendly. Additionally, a $19 million investment by the district aims to add 80 new stations and 3500 electric bikes (Lazo, 2021). This investment was done to expand Capital Bikeshare's presence in the Washington D.C. metro area by reducing car congestion and providing a more viable transportation method for commuters in a post-covid world.

Figure 1: Worldwide Bike-sharing market growth projection from 2017 to 2027 in USD Billions



As shown in *Figure 1*, the bike-sharing segment in the world is projected to grow by 11.48% (2022-2026), resulting in a market volume of US $12.29B in 2026 (Statista, 2022). Since the

bike-sharing industry is growing, companies are seeing how bike-sharing can transform the way we commute and decided to grow their presence in this market. For example, In 2018, Lyft completed the acquisition of Motivate, the biggest bike-share operator in the U.S. which operates the likes of Capital Bikeshare, Citi Bike, and many other bike-share companies (Siddiqui, 2018).

With bike-sharing becoming increasingly important to the DC metro area, this report aims to better understand Capital bikeshare (the most popular DC bike-sharing system) and its attributes/characteristics. We will do that by analyzing the attributes/characteristics of Capital Bikeshare.

The Capital Bike system has many facets to it. For example, there are two types of users: members (who pay $7.92/mo for unlimited 45min rides or $95 billed upfront annually) and casual riders (who pay per trip and pay a $1 unlocking fee) (Capital Bikeshare, n.d). Additionally, users can choose between different types of bikes: “classic” bikes (standard bikes) and electric bikes. These bike types have different costs, which lead to them being used for different purposes. For example, it costs $1 for casual users to unlock a bike and are charged 0.05/min for a classic bike and 0.15/min for an e-bike (Capital Bikeshare, n.d). On the other hand, members are charged 0.10/min for an e-bike with no unlock fee (Capital Bikeshare, n.d).

To analyze how these bike types are used differently, we plan to look at comparisons of how ride duration and distance traveled differs between electric and classic bikes. We can also look at the interplay between Member Type and Bike Type to see if members are using electric bikes more than casual riders are. Another aspect of bike-sharing systems is when trips are taken. This is important to bike-sharing companies because they have to ensure that there are enough bikes at their locations throughout the day. To do that, they must understand when and

where users are taking most of their rides.

Finally, we can combine these characteristics (member type, bike type, and ride distance) to see if they have an effect on ride duration. Since ultimately revenue is determined by how long rides last, it is important to understand how each characteristic contributes to ride duration.

In summary, our research project aims to answer the following questions:

1) *At what time does Capital Bikeshare experience the most amount of riders?*
2) *At what time does Capital Bikeshare experience the least amount of riders?*
3) *Do ride durations differ between members and casual users?*
4) *Are electric bikes used for longer trips than classic bikes?*
5) *Are electric bikes used to travel further distances than classic bikes?*
6) *Do members have a different split of electric vs. classic bikes than casual riders?*
7) *How do member type, bike type, and ride distance impact ride duration?*

II. Data

The data was found and collected from the CaBi Amazon S3 Bucket. All the datasets were organized yearly from 2016 to September 2022 and, for each year, the CSV files were either present quarterly or monthly. For our analysis, we appended geospatial features to the yearly datasets not containing those features (2016-March 2020) by sampling on the feature of Station ID and then concatenated yearly data frames into a combined CSV file that we used for Descriptive and Inferential Statistics. More information about each of these steps is described in detail under Section III Data Cleaning.

Details of useful columns to our project have been listed in *Table 1* below:

Table 1: CaBi Dataset variable descriptions

Key Attribute(s)	Definition
Duration (minutes)	Response Variable present across all years
Start Date and End Date	Start and end date of each trip -> extract date features (hour, month, etc.)
Start Station # and End Station #	837 Levels: Station Addresses encoded as ID's
Bike Type	3 Levels: Classic or Docked or Electric
Rider Status	2 Levels: Casual or Member
Start and End Geospatial Coordinates	2016-2020 March missing these variables -> Data Cleaning and Feature Engineering!
Distance (miles)	Feature engineered using Python 's Haversine package

---

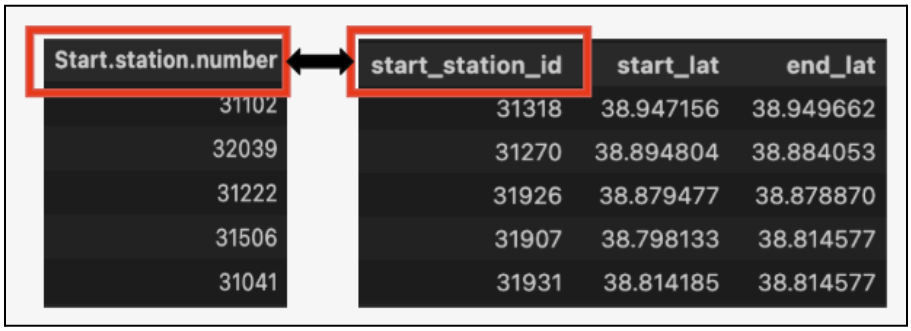
III. Data Cleaning

Considering that the data was gathered from pre existing CSV files, there was no extensive cleaning required, but rather minor pre-processing for specific needs of this project. All data cleaning was completed using Python. The correct sequence to follow in our Github Repository for data cleaning is *"511 Geospatial Features Append.ipynb"* followed by *"Joining_data_teg.ipynb"*.

It is important to mention all NA values were dropped since there were a total of 20,000 NA values out of 20 million rows of data we had gathered from all the CSVs, so dropping them would not be significant and/or change our analysis. In addition the column called "Bike.number" was dropped as it represents every ID for every bike Capital Bikeshare owns. With that being known, it can be said that running an analysis on every bike is useless since all of them are the same and there are a lot of bikes in stock.

Next, not all of the datasets (2016 - 2020) had longitude and latitude columns, which motivated us to conduct geospatial analysis for which left joins were used. Left joins were performed between every yearly dataset provided from the source on the column called Station.number (see visual example below), but distinctions were made between the 2021 and 2022 datasets that involved longitude and latitude columns with ones that did not. The only column that all the datasets share in common as shown in *Figure 2* below.

Figure 2: Merging datasets based on Station ID

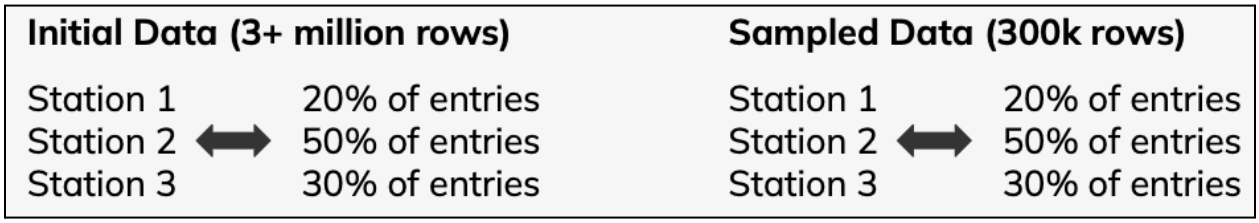


One last task that had to be completed before joining the datasets was sampling, since every dataset has around 3 million rows, and joining them as is would have resulted in 21 million rows, which was not within the scope of the project and would require much more computational power. For that reason, we had to come up with a sampling strategy that is reflective of the

population data, which is why we decided to sample on station number, as it guarantees the spread of rides around the DC. Maryland, and Virginia area remains consistent with that of the population data.

As shown below in *Figure 3*, presented is a small example that explains that sampling strategy:

Figure 3: Initial Data set v Sampled Data set



Assume the initial dataset had 3 million rows and all rides are coming from 3 stations with 20% of bike rides from the first station, 50% from the second, and 30% from the third. The sampling was conducted using Python functions that guarantee that this spread among stations is the same when x (in this case 10) % is being sampled from the initial data. Obviously, The data set had much more than just three stations but this is just a small example to explain the sampling strategy.

Finally all 7 datasets were joined together into one final cleaned dataset that had approximately 2.1 million rows of data, which was representative of the population data.

IV. Exploratory Analysis

Figure 4: Capital Bikeshare usage per Year (10% data sampled from each year)*

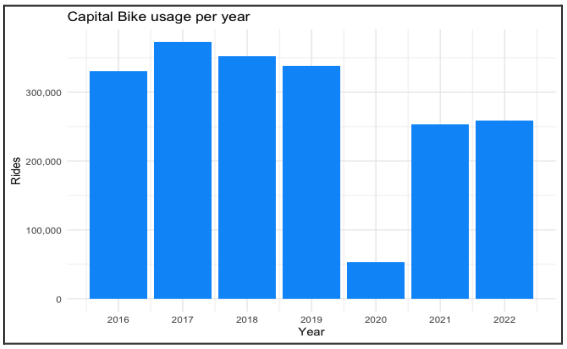
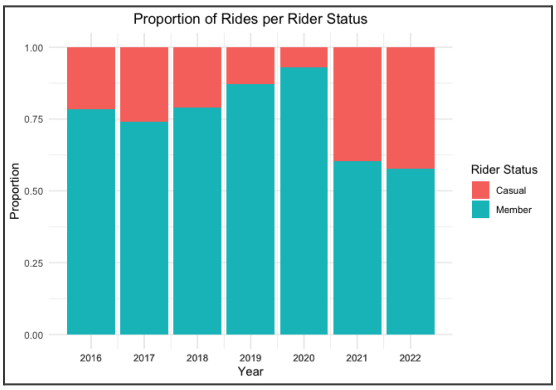


**Figure 4* above was generated using our sampled data, so the original number of rides for each year is ten times more than that is showcased on the y-axis. We notice that the demand for CaBi stayed relatively similar across 2016 to 2019. The drop in demand in 2020 is a cause of the COVID-19 pandemic and the consequent lockdown imposed. However, the key finding here is that post-covid demand levels have not yet caught up to pre-covid demand!

Figure 5: Proportion of Rides per Rider Status



A positive sign of growth in members over years 2016 to 2020 is seen in *Figure 5* above. The fact that CaBi had the highest number of members in 2020 signifies that the number of casual riders, mainly tourists and folks residing in D.C for a few months at maximum, declined significantly due to the imposed lockdown. As a result of the lockdown, we notice a drop in members from 2020 to 2021 by almost 25%! Therefore, the pandemic indubitably affected CaBi's revenues and has now put them in a period of recovery.

Figure 6: Line plot of the number of start rides per hour of the day by Rider Status (2016-2022)

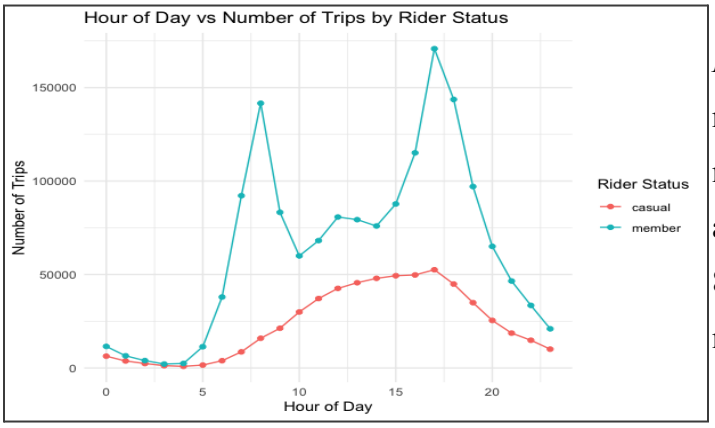


Figure 6 on the left highlights that there is relatively no activity on CaBi bikes amongst members and casual riders at 3AM and 4AM across the years 2016 to 2022. However, at 8AM we see a spike in rides started for members and not for casual riders, implying

that members mainly commute for work using CaBi. Another peak is seen at 5PM, this time for both members and casual riders, which denotes that members mainly commute from their workplaces back home but casual riders set off for a leisurely trip, reinforcing our observation from Figure 6 that casual riders comprise mainly tourists.

Figure 7: Ridge plot of log ride duration across 2016-2022

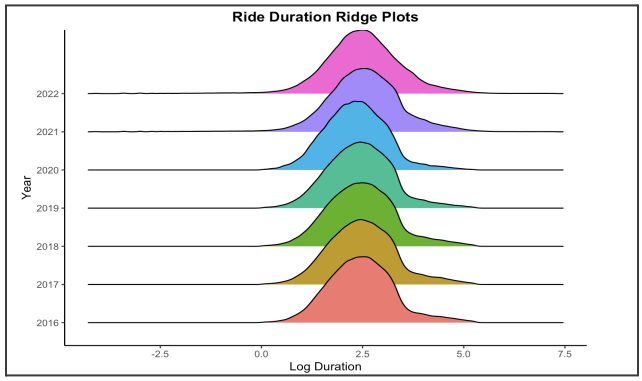
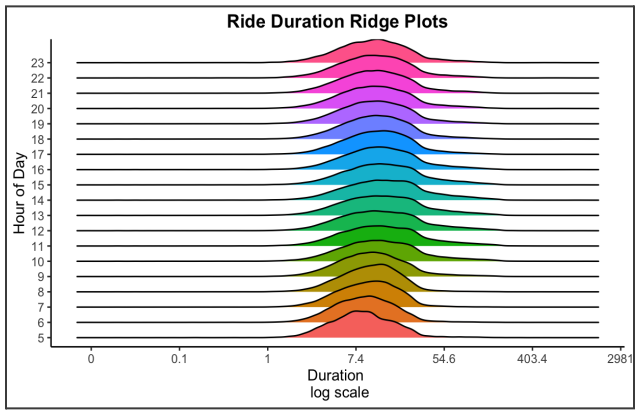
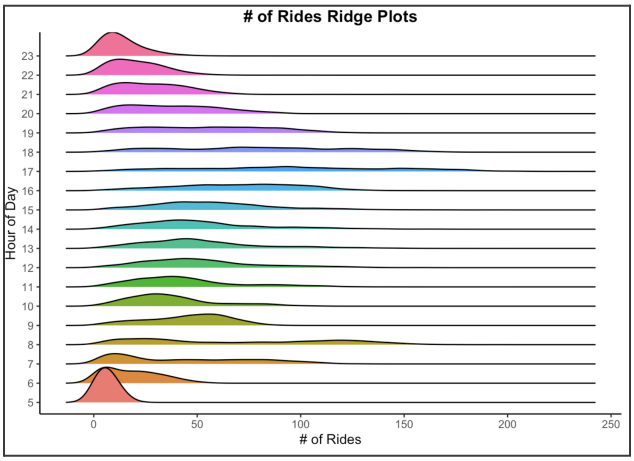


Figure 8: Ridge plot of log ride duration (2016-2022) per hour of the day



From *Figure 7*, we now know that not only did frequency of rides reduce in 2020, but so did duration of each ride as the peak of 2020 is slightly shifted left relative to other ridges. *Figure 8* accounts for outliers by using a log-transformation on duration in minutes. The main takeaway from this plot is the heavier tails

Figure 9: Ridge plot of number of start rides (2016-2022) per hour of the day



Analogous to *Figure 8, Figure 9* above helps us visualize the distribution of the number of rides started over a day across all years and members in our data. At 5AM, there is relatively no activity for CaBi, but once the clock strikes 6AM and riders start their day, the activity increases gradually until it reaches a peak at 8AM. Due to work hours, activity is low from 9AM to 4PM, but we see even greater activity at 5PM and 6PM as riders leave their workplaces. Activity after 6PM starts reducing gradually and the cycle begins again at 6AM the next day.

for hours 10AM to 11PM, implying riders use CaBi for greater durations half the day compared to the other half.

Figure 10: Proportion of Rider Status (2016-2022)

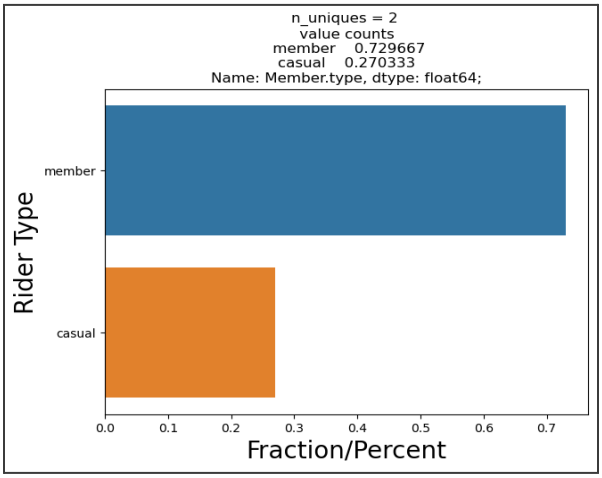


Figure 10 above signifies that CaBi members account for approximately 73% of the whole data and casual riders account for 27%

Figure 11: Proportion of Bike Type (2016-2022)

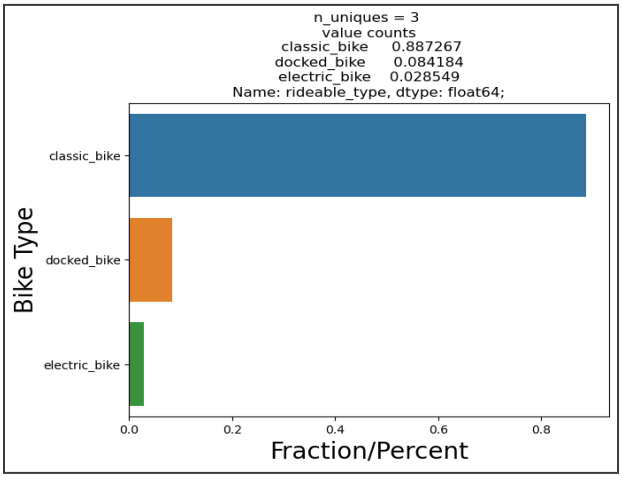
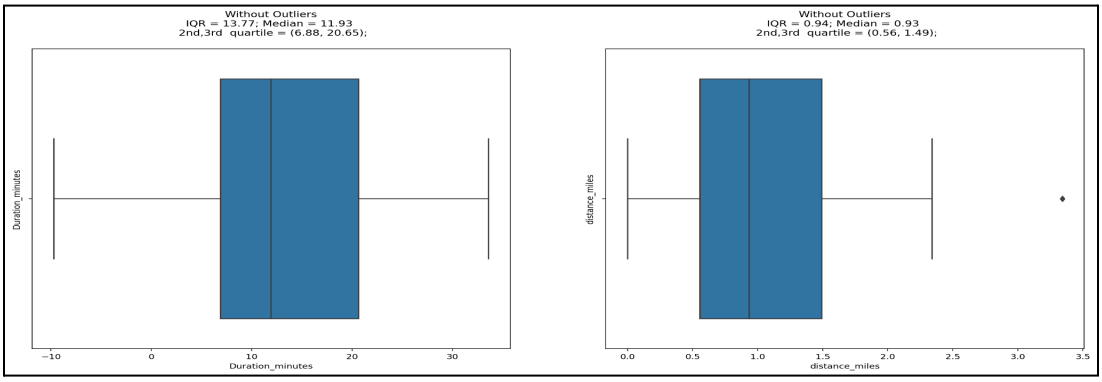


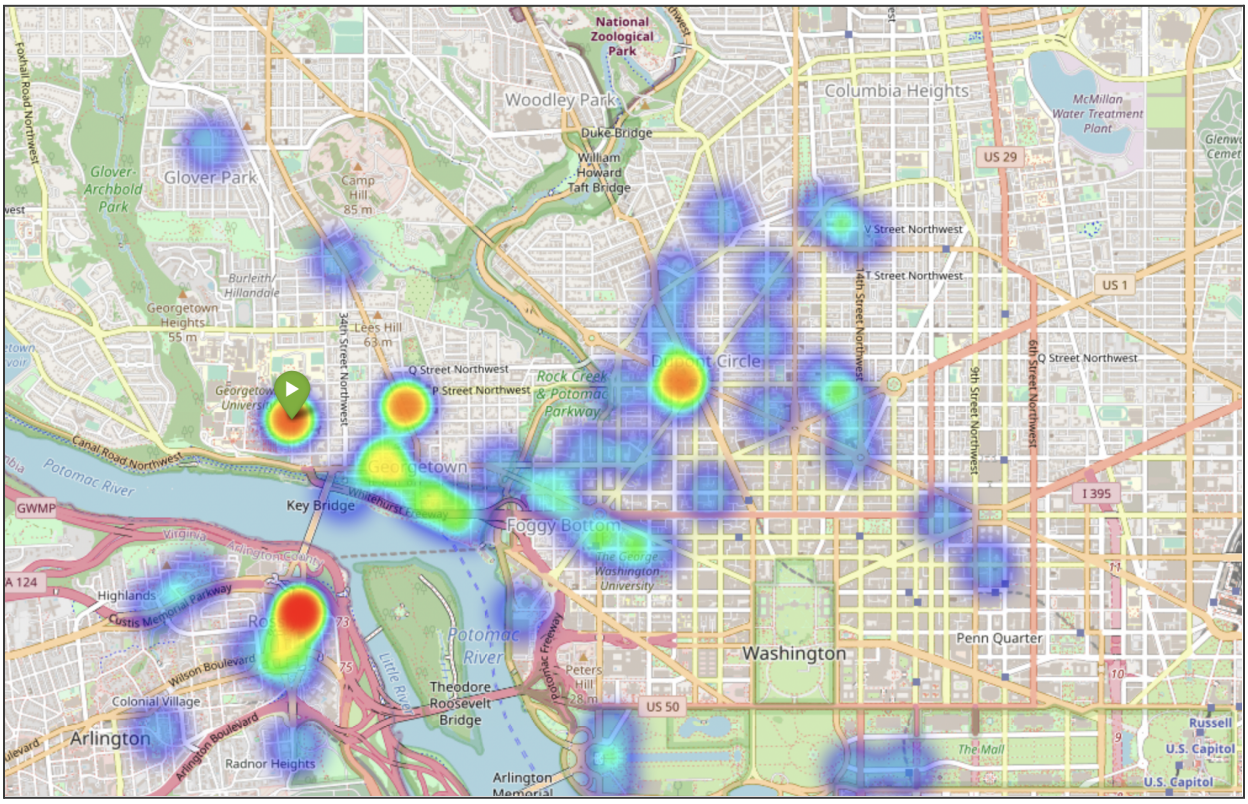
Figure 11 above encapsulates another categorical variable, bike type. Classic bikes were present since the inception of CaBi and comprise approximately 90% of the data. Electric bikes were introduced in 2020 and because we could not find conclusive information about the meaning of docked bikes, we decided to drop it entirely for statistical analyses.

Figure 12: Boxplots of both trip duration in minutes (left boxplot) and distance traveled in miles (right boxplot) (2016-2022)



From *Figure 12* in the previous page, the numeric variables of duration in minutes and distance in miles still had a right-skewed distribution after removing heavy outliers. Because the median is closer to the left of the box (lower duration or distance) and the whisker is shorter on the left end of the box, Their distributions are right-skewed. The median duration is approximately 12 minutes and the median distance covered is approximately one mile among members and casual riders. Therefore, using log-transformations on these features is better for t-Tests as they would then follow normality strictly.

Figure 13: End Stations for trips starting from Georgetown University (37th & O St NW)



The Folium package in Python helped us generate the geospatial visualizations because our data contained latitude and longitude features for start as well as end dates of each trip. Therefore, we created a function that takes in the station address as a string and outputs a heatmap of stations where rides ended. From *Figure 13* above, most trips ended around the Rosslyn Metro Station and Dupont Circle, indicating that the Georgetown community uses CaBi as a substitute for the Georgetown University Transportation Shuttle. Moreover, 8000 trips, including 5500 members and 2500 casual riders, were started from *37th & O St NW* across 2016-2022.

V. Statistical Methods and Results

A. Hypothesis Testing

Four distinct hypothesis tests were completed throughout the project in the order presented below:

- 2 Two Sample t-tests
- 1 Bootstrap test
- 1 Chi squared test

Each test, its respective hypotheses, and its results are explained below:

- **Two sample t-test (Membership status v Bike ride duration)**

The first hypothesis test included an analysis of seeing what statistical inference can be extracted from the comparison of Membership status and Bike ride duration result in. A t-test was chosen because its measurement is a direct result of comparing the means of the two populations that are set as variables, as in this case between membership status and bike ride duration. This comparison was chosen specifically to identify which members are using the bikes the most and as a direct result identify any decisions that could be taken to address this relationship. The null hypothesis and alternative hypothesis that were determined are listed below:

H_0: The mean trip duration of casual riders is the same as that of members across 2016-2022

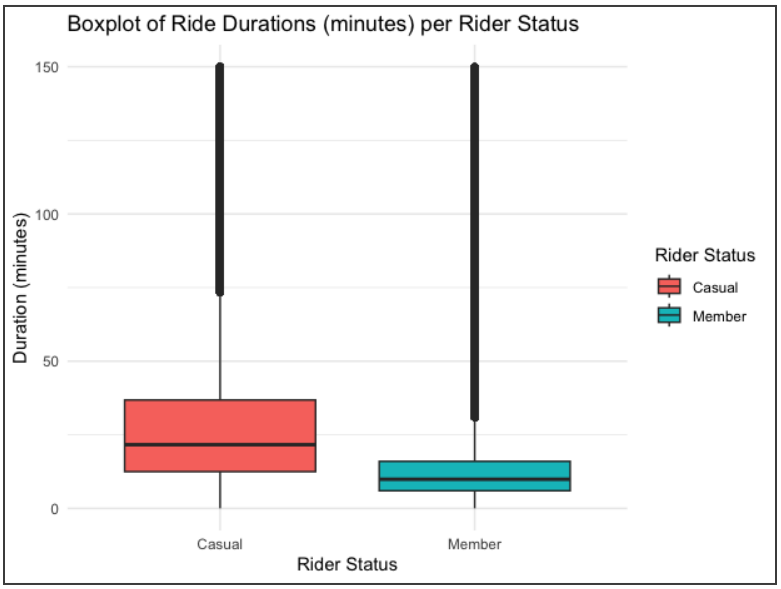
H_1: The mean trip duration of casual riders is more than that of members across 2016-2022

Table 2: Statistical significance of t-test between Membership status and mean trip duration

P-Value	95% Confidence Interval: Lower Bound	95% Confidence Interval: Upper Bound	Mean Trip Duration of Casual Riders across 2016-2022	Mean Trip Duration of Members across 2016-2022
$< 2.2 \times 10^{-16}$	17.8. minutes	Infinity	30.4 minutes	12.6 minutes

With these entities set, the statistical inferences gathered from the test can be used to help identify which statement we can accept, above in *Table 2*, since the p-value is less than 0.05, we ran reject the null hypothesis with 95% confidence and conclude that that Mean trip duration of casual riders is Mean trip duration of members, which can be visualized in *Figure 14* below:

Figure 14: Boxplot of ride duration of different membership types



- **Two sample t-test (Bike type v Bike ride duration)**

The second comparison was again conducted using another two sample t-Test. Its relation included bike type and bike ride duration as we sought to investigate what ways users would prefer when commuting across the local area. Although it seems that the company seemed to overwhelmingly have classic bikes in stock (89%) in our exploratory analysis as depicted with *Figure 11*, it was advisable to make this comparison to see what distinction can be drawn if there are so many classic bikes compared to electric bikes. The null hypothesis and alternative hypothesis that were determined are listed below:

H_0: The mean trip duration on classic bikes is the same as that on electric bikes

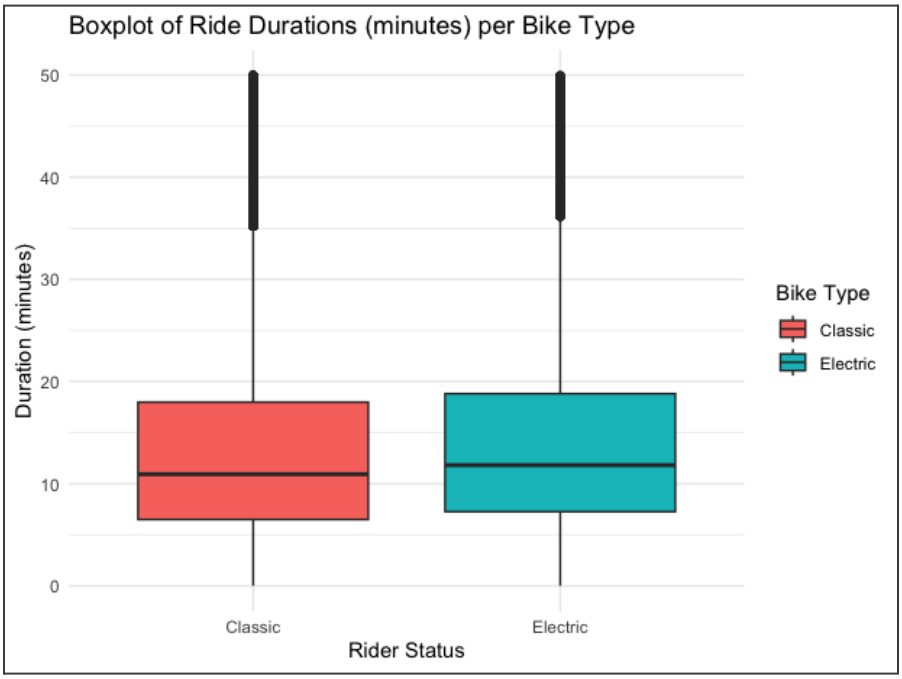
H_1: The mean trip duration on classic bikes is not the same as that on electric bikes

Table 3: Statistical significance of t-test between bike type used and mean trip duration

P-Value	95% Confidence Interval: Lower Bound	95% Confidence Interval: Upper Bound	Mean Trip Duration of Rides Completed on Classic Bikes across 2016-2022	Mean Trip Duration of Rides Completed on Electric Bikes across 2016-2022
$< 1.2 \times 10^{-8}$	Negative Infinity	-0.26	16.54 minutes	16.92 minutes

As *Table 3* above shows, the p-value is less than 0.05, with that being said, we can reject the null hypothesis with 95% confidence and conclude that the mean trip duration on electric bikes is greater than the mean trip duration on classic bikes, which can be visualized in *Figure 15* below:

Figure 15: Boxplot of Ride duration(minutes) per Bike type



❖ **Bootstrap test**

With the third hypothesis test, it was chosen to conduct a bootstrap comparing the distance traveled with the type of bike used by the user. The motivation behind this test came from *Figure 12*. Its findings show that most bikes traveled at a median of 0.93 miles with outliers removed. This was the initial outlook based on all bikes used, but a distinction needed to be made to see which type of bike was used the most and in order to find that, the bootstrap method allowed the variables to be sampled without replacement and create a distribution that would imitate the data. The null hypothesis and alternative hypothesis that were determined are listed below:

H_0: The mean distance covered per trip on classic bikes is the same as that on electric bikes.

H_1: The mean distance covered per trip on classic bikes is not the same as that on electric bikes.

Figure 16: Boxplot of bike type and distance traveled (miles)

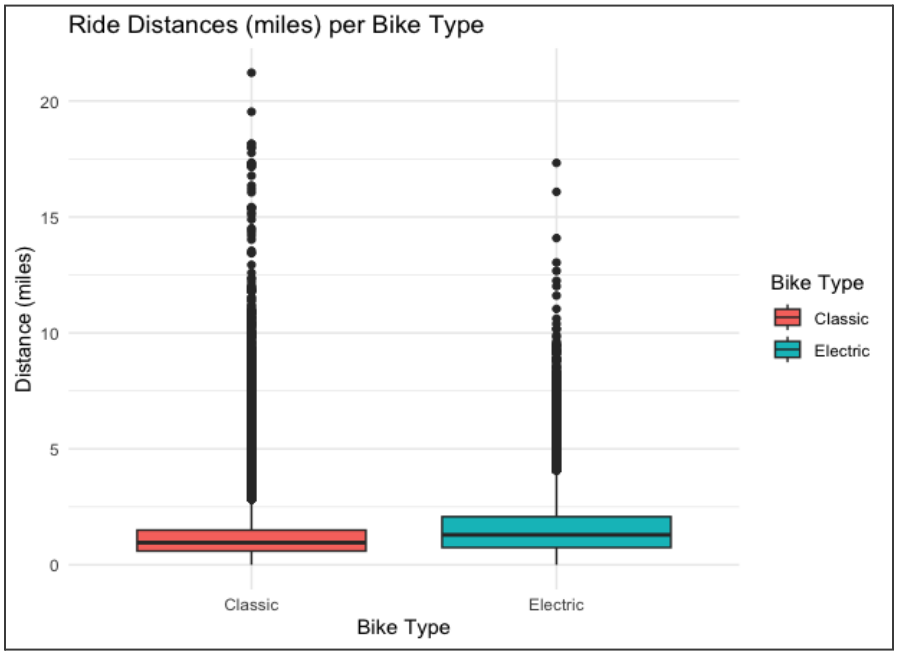
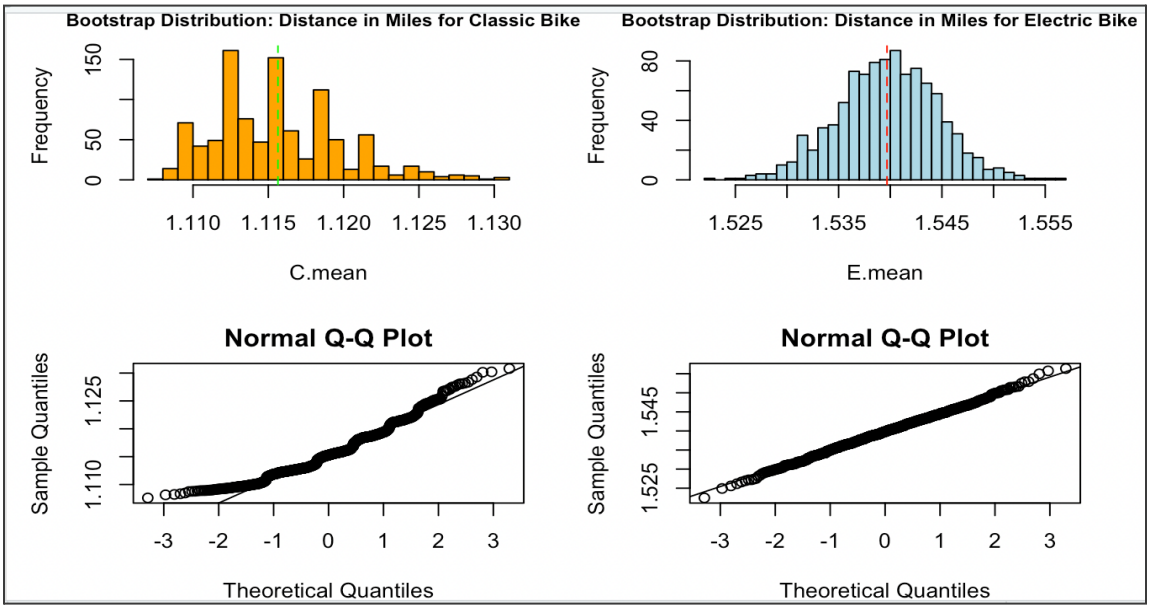


Table 4: Confidence intervals determined from bootstrap test between bike type used and mean distance traveled

95% Confidence Interval (Difference in Means): Lower Bound	95% Confidence Interval (Difference in Means): Upper Bound
0.41	0.44

After sampling, as seen in *Table 4* above, it was found that 0 is not within the confidence interval, and we can further reject the null hypothesis and conclude the mean distance traveled on electric bikes is greater than the mean distance traveled on classic bikes. Furthermore, in *Figure 16*, the classic bike mean distance traveled from the sampled distribution was around 1.116 miles whereas the electric bike sample distribution was much larger at 1.540 miles. Both distributions seemed to follow a normal gaussian distribution, yet the electric bike was more representative of that statement.

Figure 17: Bootstrap Distribution and Quantile plots for classic and electric bikes



❖ Chi Squared test

For the last hypothesis test, a chi square test of independence was performed to determine the relationship between bike rider status and bike type. This relationship has many implications as riders may be more inclined to use specific bikes based on their personal needs. The chi-square test can show a level of interdependence, which was why this test was examined. The chi-square test looks to evaluate experimental and theoretical values where the sum of all those differences will be compared to the test statistic. The null hypothesis and alternative hypothesis that were determined are listed on the next page:

H_0: Bike rider status and bike type used are independent

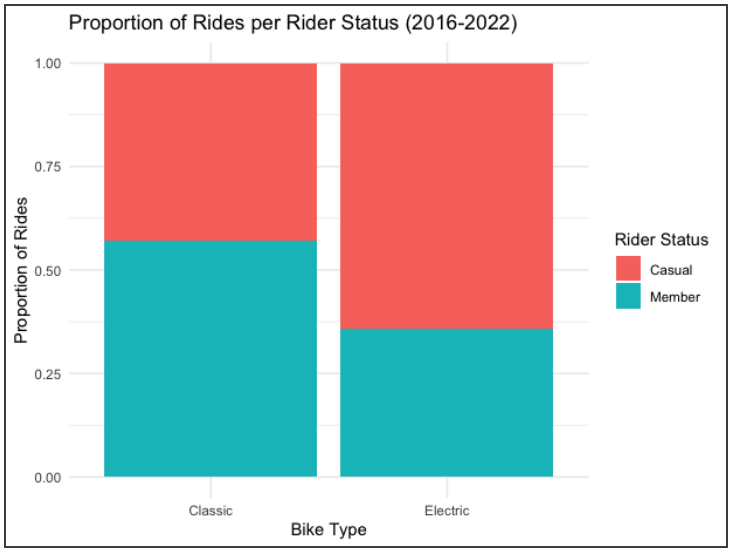
H_1: Bike rider status and bike type used are dependent

Table 5: Table categorizing the cumulative number of people based on bike type used and membership status

Two-Way Table (Bike Type vs Rider Status)	Casual	Member
Classic Bike	440,846	1,428,789
Electric Bike	25,592	34,566

The chi-square test value was 105,876 highlighting the difference between the observed and expected frequencies of the outcome, signifying how well the sample data matches the known characteristics of the larger population. The statistical results also presented us with a p-value less than 0.05 where we could reject the null hypothesis and conclude with 95% confidence that Bike Type and Rider Status are dependent.

Figure 18: Stacked Bar plot of the Proportion of Bike type used based on membership status

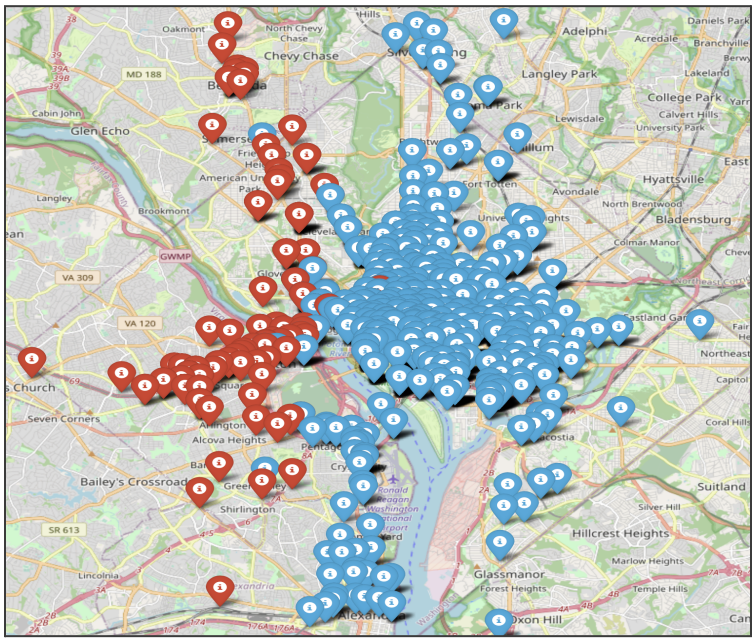


B. Clustering

Since our dataset has more than 700 stations, using station ID was avoided as including it in statistical and machine learning models as a categorical variable would be very hectic and would probably require supercomputers. Consequently, we opted to cluster the stations into 2 separate groups that we would eventually be used as a variable in the machine learning part of this project, reducing the number of variables from 700 (total stations) to 2 (clusters).

K-means clustering was used since it is the most popular clustering algorithm and fits the problem well. The results can be visualized in *Figure 19*.

Figure 19: K-Means Clustering of BikeShare Stations



The algorithm's performance is near perfect, as there is no visible overlap between the two clusters and all stations are assigned to a cluster in a visually logical manner. The red cluster

is representative of bike stations in Georgetown, Chevy Chase (Maryland), and Arlington (Virgina) and the blue cluster is representative of bike stations in both Washington D.C, bordering the states of Maryland and Virginia, and the vicinity of Ronald Reagan Washington National Airport (Virginia). This has now helped us factor in the location of each station through the resulting labels of the clustering models.

C. Linear Regression

In an effort to predict the amount of time users spend on the bike, a Linear regression model was used to input multiple variables that may contribute to a desired prediction. To initiate this process, it was devised to cluster specific chosen variables first. As shown in *Figure 19*, the distances in miles, bike type, and membership status were all used to train the model, and they were also used as variable inputs for the clustered group, as shown by the distinct clustered group they were placed into“k_1” and “k_0”, while *Figure 20* depicts the chosen target variables that were used for predicting the model, distance in miles traveled.

Details about input variables:

- Bike type(r_classic_bike, r_docked_bike, and r_electric_bike): Provides the type of ride being used. These include ‘electric’, ‘classic’, and ‘docked’. With the hypothesis testing, we came to know that the mean distance traveled on electric bikes was more than that of casual bikes. So it’s pretty interesting to know how long people ride different types of bikes for their CaBi journey.
- Member type(M_member and M_casual): Now knowing the method in which members used CaBi as compared to a casual rider. The hypothesis tests also

provided an insight on how much a member uses an electric bike in comparison to the latter. Such fascinating aspects contribute to the significance of this particular column in predicting the dependent variable.

- Distance of miles traveled (distance_miles): This is a feature-engineered column using the geospatial data that we had in the initial dataset. Start and end lats and longs along with other details such as Station name and number contributed to calculating the distance covered by a rider.
- K-means variable (k_1 and k_0): By using k-means, we divided certain areas of the map into different clusters. Based on the geospatial data, k-means clustering was achieved helping not only in Exploratory Data Analysis but also in Dimensionality Reduction.

Details about the dependent variable:

- Duration (minutes): How long, on average, does a user ride CaBi bikes across the DC-Arlington area? The response variable can help in providing better recommendations according to the bike-type. It can also let us know whether casual riders want to become CaBi members.

Figure 20: Variable set used to train Linear Regression model

	distance_miles	r_classic_bike	r_docked_bike	r_electric_bike	M_member	M_casual	k_1	k_0
0	0.482018	1.0	0.0	0.0	0.0	1.0	0.0	1.0
1	0.468603	1.0	0.0	0.0	0.0	1.0	0.0	1.0
2	1.791614	1.0	0.0	0.0	1.0	0.0	0.0	1.0
3	0.614417	1.0	0.0	0.0	0.0	1.0	1.0	0.0
4	0.498364	0.0	1.0	0.0	0.0	1.0	0.0	1.0

Performance metrics used to provide accuracy of our model:

- Mean absolute error: This is the average of absolute errors of all the data points in the given dataset. The models mean absolute error was set at 14.8.
- Mean squared error: This is the average of the squares of the errors of all the data points in the given dataset. The model's mean squared error was set at 25,133.51, which was relatively high.
- Median absolute error: This is the median of all the errors in the given dataset. The main advantage of this metric is that it filters any outliers within the data. A single bad data point in the test dataset wouldn't skew the entire error metric, as opposed to a mean error metric. The model's median absolute error was 4.97.
- Variance score: This score measures how well our model can account for the variation in our dataset. A score of 1.0 indicates that our model is perfect. Our model's variance score was 0.01, putting to emphasis its insignificant value.
- R^2-score: This score refers to the coefficient of determination. This tells us how well the unknown samples will be predicted by our model. The best possible score is 1.0, but the score can be negative as well. The model's R^2-score was 0.01, which restated that our model was insignificant.

Using this devised structure, the analysis and results shown from the linear regression model unfortunately do not provide sufficient evidence to predict the amount of time users will spend on the bikes. The statistics gathered from the model show a high mean squared error alongside an R^2-value that highlighted the model is not fit for our data.

Figure 21: Linear Regression model based on prediction and test results

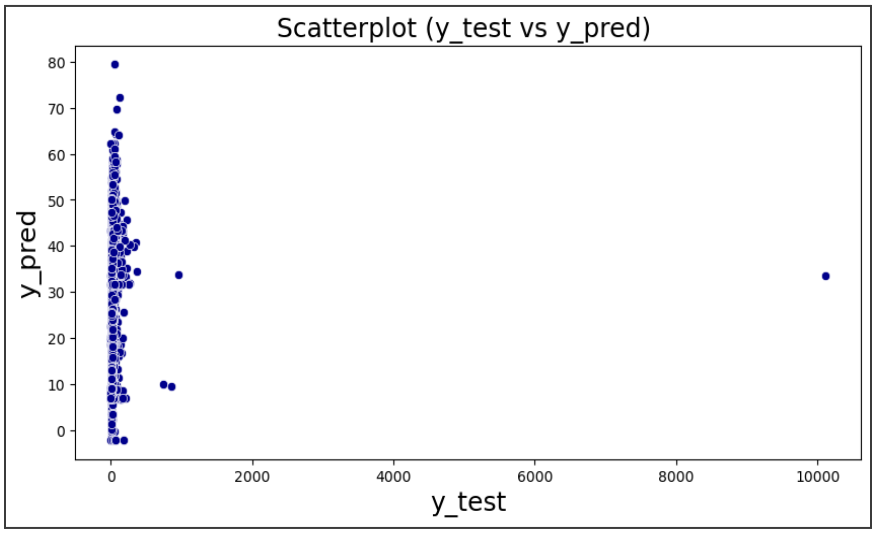


Figure 21 also shows how poorly linear regression performs for this problem, since the predictions with respect to the actual values are all over the place and there does not seem to be any consistency in performance. Also, the figure also showcases the presence of extreme outliers in the test data, with some durations exceeding 10,000 minutes. This information could be used later on when developing other models to improve their performance and the quality of the test data.

VI. Conclusions

Our study focused on exploratory data analysis and the implementation of several statistical tests, both of which have furthered our understanding of Capital Bikeshare, the components that affect its overall usage, and the behavioral differences between casual riders and members.

The first hypothesis test proved that a casual user will spend, on average, more time riding the bikes than that of membership holders. Possible implementations CaBi could work on given these results are developing strategies that focus on converting casual riders to members. This would operate not only with the company's benefit, but with the user as well since long-term spending for casual users would equate to a greater expenditure than if they take part in being a member.

The second hypothesis test demonstrated that all members spend greater time and greater distances traveling electric bikes than the time spent and distance traveled riding classic bikes. This finding not only helps identify the bike preferences of riders, but it also would be in the company's benefit to increase the stock of electric bikes as, eventually, its prevalence across the D.C metro area. Although a cost-benefit analysis would need to be conducted, we can clearly see this as a limitation on the study as no revenue or pricing data was provided.

The last hypothesis test showed that the type of bike used is dependent on membership status. This makes it vital for the company to take necessary steps in order to make bike-sharing more convenient for members' use.

Finally the use of Linear Regression showed how some systems are too complex for our chosen setting. Since one of the variables revolves around membership types, it is safe to say that classification would suit the project a lot more than regression, which is something we aim to work on and further develop in the near future.

Limitations associated with our study involve not being able to evaluate cost benefit analysis because a revenue or pricing feature was not provided for each ride in our dataset. Some columns were dropped in order to "clean the data", as done with the Bootstrap test. Specifically,

the docked bike type was removed because of its bike type misidentification possibility, as this would disrupt the known bike type values while conducting the hypothesis tests. Although only 10% of the data was used, the data used was adjusted by representing the same proportional ratio of data within each column of the initial 20 million row dataset covering 2016 through 2022. Moreover, the study could have been approached with the help of ANOVA models, as the comparison of means with different variable groups could be more significant and help further identify other relationships that are meaningful in a Multiple Linear Regression Model.

VII. Works Cited

Capital Bikeshare. (n.d.). About Capital Bikeshare. Retrieved December 9, 2022, from https://ride.capitalbikeshare.com/about

Capital Bikeshare. (n.d.). Choose your plan. Retrieved December 9, 2022, from https://capitalbikeshare.com/pricing

Chester, C. (2013, October 21). *Are you getting the most for your money from capital bikeshare?* WAMU 88.5 American University Radio . Retrieved December 12, 2022, from https://wamu.org/story/13/10/21/are_you_getting_the_most_for_your_money_from_capital_bikeshare/

Corin, C. (2021, October 25). *Mayor Bowser and Lyft Announce Free Capital Bikeshare memberships for all DC residents*. DC News Now Washington, DC. Retrieved December 7, 2022, from https://www.dcnewsnow.com/news/local-news/washington-dc/mayor-bowser-and-lyft-announce-free-capital-bikeshare-memberships-for-all-dc-residents/

District Department of Transportation . (2020). (rep.). *Development Plan Update* . Government of the District of Columbia. Retrieved December 12, 2022, from https://ddot.dc.gov/sites/default/files/dc/sites/ddot/page_content/attachments/23397_Capital_Bikeshare_Plan_Update_v4_051220_WEB.pdf.

Lazo, L. (2021, July 1). *Capital Bikeshare gears up for expansion as commuters resume pre-pandemic routines*. The Washington Post. Retrieved December 9, 2022, from

https://www.washingtonpost.com/transportation/2021/07/01/electric-bike-dc-capital-bikeshare/

Lazo, L. (2021, October 16). *D.C. wants you to ride bikes. This month it raised the cost to Ride Capital Bikeshare.* The Washington Post. Retrieved December 7, 2022, from https://www.washingtonpost.com/transportation/2021/10/16/dc-capital-bikeshare-fee-increase/

Lazo, L. (2022, April 29). *Meet the capital bikeshare rider who visited all 683 stations*. The Washington Post. Retrieved December 12, 2022, from https://www.washingtonpost.com/transportation/2022/04/29/capital-bikeshare-top-rider/

Russell , E. (2015, August 19). *To bike across the Potomac, most use the 14th street bridge or key bridge*. Greater Greater Washington. Retrieved December 12, 2022, from https://ggwash.org/view/38995/to-bike-across-the-potomac-most-use-the-14th-street-bridge-or-key-bridge

Siddiqui, F. (2018, July 2). *Lyft gets into bike-share business, acquiring operator of Capital Bikeshare and Citi Bike*. The Washington Post. Retrieved December 9, 2022, from https://www.washingtonpost.com/news/dr-gridlock/wp/2018/07/02/lyft-gets-into-bike-share-business-acquiring-operator-of-capital-bikeshare-and-citi-bike/

Statista. (2022, November). *Bike-sharing - worldwide*. Statista. Retrieved December 7, 2022, from https://www.statista.com/outlook/mmo/shared-mobility/shared-rides/bike-sharing/worldwide

---

VIII. Appendix

Link to Github Repository:

https://github.com/TegveerG/GU-ANLY511-FinalProject