User Behavior and Citi Bike: A Study of Behavioral Economics and Data Analytics

Daniel Elmore

Lenoir-Rhyne University

Abstract

[The abstract should be one paragraph of between 150 and 250 words. It is not indented. Section titles, such as the word *Abstract* above, are not considered headings so they don't use bold heading format. Instead, use the Section Title style. This style automatically starts your section on a new page, so you don't have to add page breaks. Note that all of the styles for this template are available on the Home tab of the ribbon, in the Styles gallery.]

Keywords: [Click here to add keywords.]

User Behavior and Citi Bike: A Study of Behavioral Economics and Data Analytics

Originally launched in May 2013, the New York metro area's bike-sharing program, Citi Bike, has experienced remarkable growth, expanding from approximately three hundred to almost two thousand stations over the last decade. With locations strategically spread across four of the five boroughs of New York—Staten Island being the lone exception—as well as in Hoboken and Jersey City, the program has fundamentally transformed the dynamics of commuting in the nation's busiest metropolis. Hence, as communities across the city now embrace Citi Bike as a convenient and eco-friendly alternative to traditional transportation modes. This seismic shift prompts a closer examination of Citi Bike's footprint, exploring its current distribution, user behavior patterns, and the intriguing opportunities for further enhancement.

When examining Citi Bike, there are a spectrum of options catering to varying user needs (See Figure 1). For infrequent users undertaking two rides in a day or less, the "Single Ride" option, priced at $4.49 per thirty minutes, comes with an additional unlock fee of $4.49 and no other specific benefits. Meanwhile, for infrequent users undertaking more than two rides in a day, the "Day Pass," available for $19 per day, provides users with a free unlock with the caveat that rides may only last thirty minutes each. For frequent users or those who plan to use the ride-sharing program over twenty times per year, there are annual memberships; the Citi Bike plan at $205 per year offers a more extended forty-five minute ride time, free unlocks, and additional perks like three guest passes and a rewards program. Similarly, the "Lyft Pink"[1] option, priced at

[1] In 2018, the ride-sharing company Lyft bought the firm behind Citi Bike, Motivate, and incorporated Citi Bike into its subscription program, Lyft Pink, in 2019 (Lekach, 2019). This is why there are two versions of an annual Citi Bike subscription. However, this could soon change as Lyft is currently in the process of exploring a sale of its bike and scooter business (Hawkins, 2023).

$199 per year, mirrors the 45-minute ride time, free unlocks, and boasts additional advantages such as access to all Citi Bike tier perks, benefits when using Lyft, and a complimentary Grubhub+ membership.

Proposing an A/B Test

In the pursuit of enhancing the user experience and optimizing Citi Bike services, a carefully designed A/B test is proposed. This test categorizes users into two distinct groups: customers and subscribers. Customers, constituting one-time or infrequent users, typically opt for the Single Ride or Day Pass plans. Subscribers, representing frequent or daily users, can choose between Citi Bike or Lyft Pink plans. From this clustering approach, it shall be determined whether there is a discernible difference in the biking behavior between subscribers and customers, particularly in terms of trip distance, duration, and station preferences. Hereafter, it will be questioned how these distinctions might inform strategies to enhance the Citi Bike service. Accordingly, the study focuses on key parameters such as trip distance, duration, and station preferences with the primary goal of assessing whether observable differences exist in these aspects, providing valuable insights for strategic enhancements to the Citi Bike service.

To control for seasonal variations in usage patterns, the A/B test exclusively utilizes ride history data from the month of May, spanning the years 2016 through 2023.[2] This specific time frame ensures a focused analysis, enabling a more accurate understanding of user behavior without the confounding effects of seasonal fluctuations. With these controls in place, the null hypothesis asserts the non-existence of observable differences in biking behavior between subscribers and customers with regard to trip distance, duration, and station preferences.

[2] In 2016, the manner in which data was formatted changed, thus why this study only goes back to 2016 and not 2013.

Conversely, the alternative hypothesis proposes the presence of such distinctions. Administered at a significance level of $\alpha = 0.05$, the outcomes of the A/B test will establish a robust foundation for directing strategic enhancements to the Citi Bike service. Stringent statistical analysis guarantees that observed distinctions are not merely coincidental but rather meaningful and actionable for improving the overall user experience.

Methodology and Results

In the process of conducting a comprehensive examination of biking behavior within the Citi Bike system, this study analyzed a substantial dataset comprising 362,100 observations. These observations, meticulously gathered and cleaned, represent user activities throughout the month of May from 2016 to 2023. Within this extensive dataset, 98,067 instances were attributed to customers, while 264,033 instances pertained to subscribers. This extensive dataset serves as the cornerstone for investigating the selected criteria between customers and subscribers. Through a meticulous examination of a diverse range of biking activities, the study endeavors to identify any significant differences in trip distances, duration, or start and end location, thereby shedding light on distinct patterns that may inform strategies for optimizing the Citi Bike service.

Distance Disparities.

In scrutinizing the disparities in trip distance between subscribers and customers within the Citi Bike service, a rigorous methodology was employed. Most importantly, the distribution of distance traveled underwent analysis using a histogram, providing a visual representation of the frequency distribution. This showed that both distance distributions displayed a somewhat normal distribution, but also had a positive skew (See Figures 2 and 3). Complementary to this, a box plot was utilized for outlier analysis, with the objective of identifying and comprehending any extreme values within the dataset (See Figure 4). Following these explorations, a two-sample

t-test was applied to statistically compare the mean distances traveled by customers and subscribers. The outcomes of the t-test reveal noteworthy distinctions. For customers, the mean distance traveled was 1,168.61 meters, with a standard deviation of 930.521 meters. In comparison, subscribers exhibited a mean distance of 1,081.43 meters, with a standard deviation of 763.203 meters. Subsequently, the calculated t-value of approximately 26.3, with 149,722 degrees of freedom, surpasses the critical t-value at a 95% confidence level. Consequently, the null hypothesis is rejected, signifying a discernible difference in trip distance between subscribers and customers within the Citi Bike service.

Furthermore, it should be noted that customer trip distances exhibit greater variance than those of subscribers, although both groups show a significant number of outliers. Notably, customers are about three times as likely as subscribers to initiate and conclude their rides at the same station, constituting 14.86% of customer trips compared to 4.86% for subscribers. However, it is important to acknowledge a potential constraint in the analysis—the use of data exclusively from the month of May. Given New York City's peak tourist seasons in June–August and November–December, this constraint may omit essential information regarding customer usage patterns during busier periods (McCulley, 2020).

Duration Differences and Tardiness Rates.

Delving into the duration differences between subscribers and customers within the Citi Bike service entailed a meticulous analysis of various factors. While it may seem apparent that customers and subscribers would have an apparent difference in ride times due to the fact subscribers benefit from an additional fifteen minutes of ride time without incurring a late fee, the data tells a different story. As observed in the trip distance section, this investigation comprised the assessment of distribution using a histogram and an examination of outliers

through a boxplot (See Figures 5-7). In conducting a two-sample t-test, this revealed a significant disparities in trip duration, but not in the manner one may expect. Customers, with a mean duration of 32 minutes and 54 seconds and a standard deviation of 245 minutes and 45 seconds, displayed markedly different patterns compared to subscribers, who exhibited a mean duration of 10 minutes and 49 seconds and a standard deviation of 78 minutes and 6 seconds. Consequently, the calculated t-value of approximately 27.8 with 105,507 degrees of freedom exceeded the critical t-value at a 95% confidence level, leading to the rejection of the null hypothesis and indicating a noticeable difference in trip duration between the two user groups.

In the histogram, both duration distributions displayed a positive skew, with customer trip durations exhibiting greater variance than those of subscribers. This variance might be attributed to customers potentially misunderstanding the pricing model with time limits and additional fees, which, notably, constitute a substantial source of income for Citi Bike. Moreover, despite subscribers benefiting from an additional fifteen minutes of ride time, their trips were approximately three times shorter than those of customers, although this difference is partly influenced by outliers. Specifically, customers averaged 33 minutes and 4 seconds per trip, while subscribers averaged 10 minutes and 49 seconds. This discrepancy could potentially be attributed to subscribers having a planned route for their ride, while customers simply look to bike through the city without a predetermined start and end location.

Starting and Ending Station Selection.

Finally, the examination of station preferences within the Citi Bike service was bifurcated into two facets: start location and end location preferences. Visual representation through a stacked column chart facilitated the discernment of patterns and trends in user choices (See Figures 8-11). To assess the significance of differences in station preferences between

subscribers and customers, a chi-square test was applied. Outcomes revealed compelling distinctions in both start and end station preferences. The calculated p-value of approximately 2.2e-16 for both components exceeded the critical p-value at a 95% confidence level, leading to the rejection of the null hypothesis. This robust statistical evidence attests to a discernible difference in station preference between the two user groups.

In this analysis, the station called "Grove St PATH" stood out as the unequivocal favorite for both starting and ending trips. Located in Jersey City, NJ, its popularity is attributed to proximity to a subway station, offering seamless access to southern Manhattan. Intriguingly, the top eight starting locations perfectly mirrored the top eight ending locations, underscoring a consistent user behavior pattern. However, despite these prominent preferences, the analysis unveiled a diverse landscape between subscribers and customers when it came to lesser-used stations. This came as less than 50 uses to start a trip were recorded for 90 stations, and less than 50 uses to end a trip were noted for 225 stations. This nuanced exploration emphasizes the varied station preferences among users, with certain stations experiencing significantly higher usage patterns. Overall, understanding station preferences provides valuable insights for optimizing bike distribution, ensuring station availability, and strategically locating new stations. These findings underscore the importance of catering to diverse user needs and preferences within the dynamic landscape of Citi Bike utilization.

Business Recommendations and Market Opportunities

Based on the conducted analysis, this study unequivocally rejects the null hypothesis, revealing a discernible difference in biking behavior between subscribers and customers. Notable variations in trip distance, duration, and station preferences have emerged, providing a foundation for strategic business recommendations aimed at optimizing the Citi Bike payment

model. In response to these findings, four key business recommendations have been proposed to enhance the efficiency and appeal of the Citi Bike service. Firstly, a call to simplify plan options suggests the removal of the Citi Bike plan, positioning Lyft Pink as a more cost-effective alternative with added perks. Moreover, this could also lead to more people using Lyft, thereby increasing revenue in other parts of the business. Secondly, a proposal to increase late fees aims to incentivize timely bike returns and capitalize on users potentially overlooking terms and conditions. Thirdly, the implementation of demand-based pricing for single rides introduces variable unlock fees based on real-time supply and demand, enhancing flexibility. Lastly, a recommendation to analyze existing infrastructure underscores the importance of strategically expanding bike station locations near high foot traffic areas and subway stations, optimizing accessibility.

Resultantly, it is proposed to introduce a revamped pricing model (See Figure 12). For the "Single Ride," a charge of $4.99 per 30 minutes with variable unlock fees and a $5 late fee for every 15 minutes is suggested. Meanwhile, the "Day Pass," priced at $19.99 per day with free unlock and a $5 late fee for every 15 minutes, caters to occasional users. Finally, the aforementioned annual Lyft Pink membership, at $199, offers free unlocks, a 45-minute ride time, and a 20¢ per minute late fee, coupled with additional benefits. Consequently, these strategic adjustments aim to streamline the payment model, incentivize timely returns, and enhance the overall user experience. Nevertheless, regular evaluation of user behavior and market dynamics will be crucial for refining these recommendations and ensuring continued success in meeting the evolving needs of Citi Bike users.

References

Citi Bike Membership & Pass Options. (2023, January 1). Citi Bike NYC. https://citibikenyc.com/pricing

Finley, L. (2023, May 26). *Citi Bike celebrates 10 years in New York City*. Ny1.com; Spectrum News NY1. https://ny1.com/nyc/all-boroughs/transit/2023/05/26/citi-bikes-celebrates-10-years-in-new-york-city

Hawkins, A. J. (2023, July 26). *Citi Bike is for sale*. The Verge; Vox Media. https://www.theverge.com/2023/7/26/23808205/lyft-citi-bike-sale-bike-share-scooter

Lekach, S. (2019, October 29). *Lyft Pink is the ride-hailing app's new monthly subscription plan*. Mashable; Ziff Davis. https://mashable.com/article/lyft-pink-ride-hail-monthly-membership

McCulley, K. (2020, January 13). *The Very Best Time to Visit New York (by a local!) - Adventurous Kate*. Adventurous Kate. https://www.adventurouskate.com/best-time-to-visit-new-york/#:~:text=Generally%20speaking%2C%20high%20season%20in,Christmas%20and%20the%20holiday%20season.

Figures

Figure 1

Plan and Pricing Model for Citi Bike as of November 2023

	Single Ride	Day Pass	Citi Bike	Lyft Pink
Charge/Freq.	$4.49 / 30 minutes	$19 / day	$205 / year	$199 / year
Unlock fee	$4.49	Free	Free	Free
Ride time	30 minutes	30 minutes	45 minutes	45 minutes
Late fee	23¢ / minute	$4 / 15 minutes	17¢ / minute	17¢ / minute
Other benefits?	N/A	N/A	3 guest passes Rewards program	All Citi Bike tier perks Benefits when using Lyft Free Grubhub+ membership

(Citi Bike Membership & Pass Options, 2023)

Figure 2

Plotting Distance Travelled with a Histogram

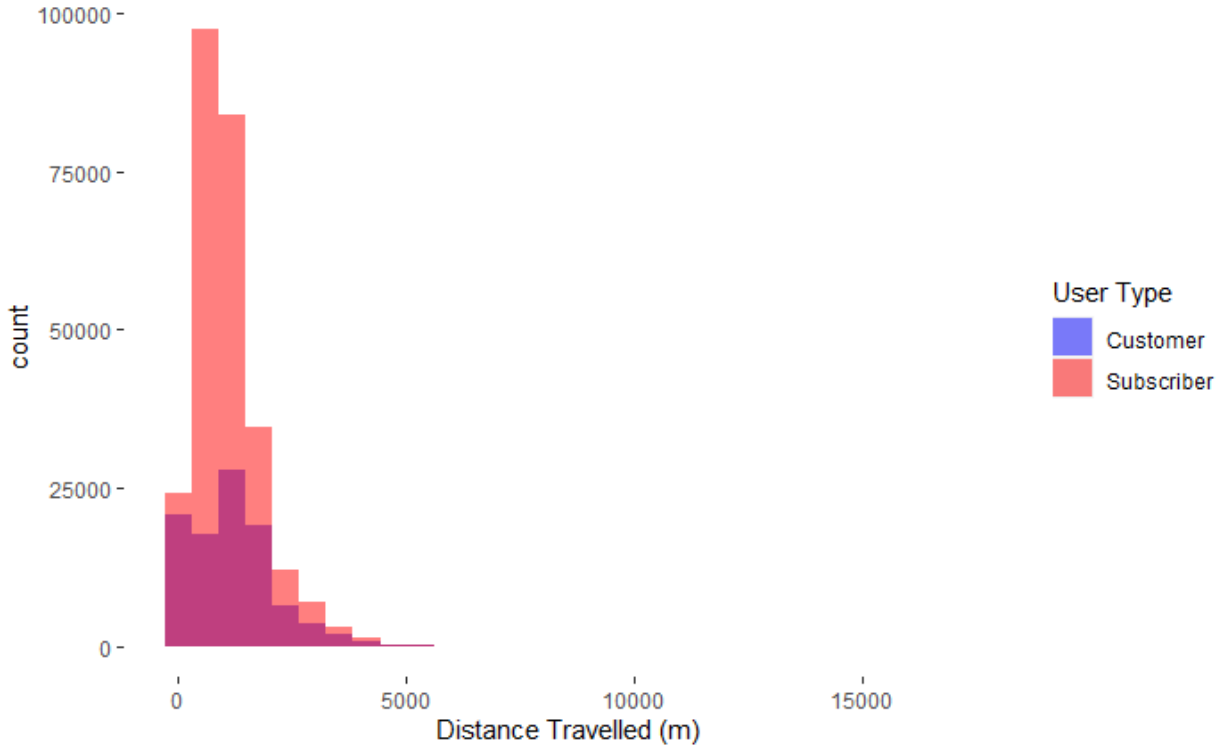
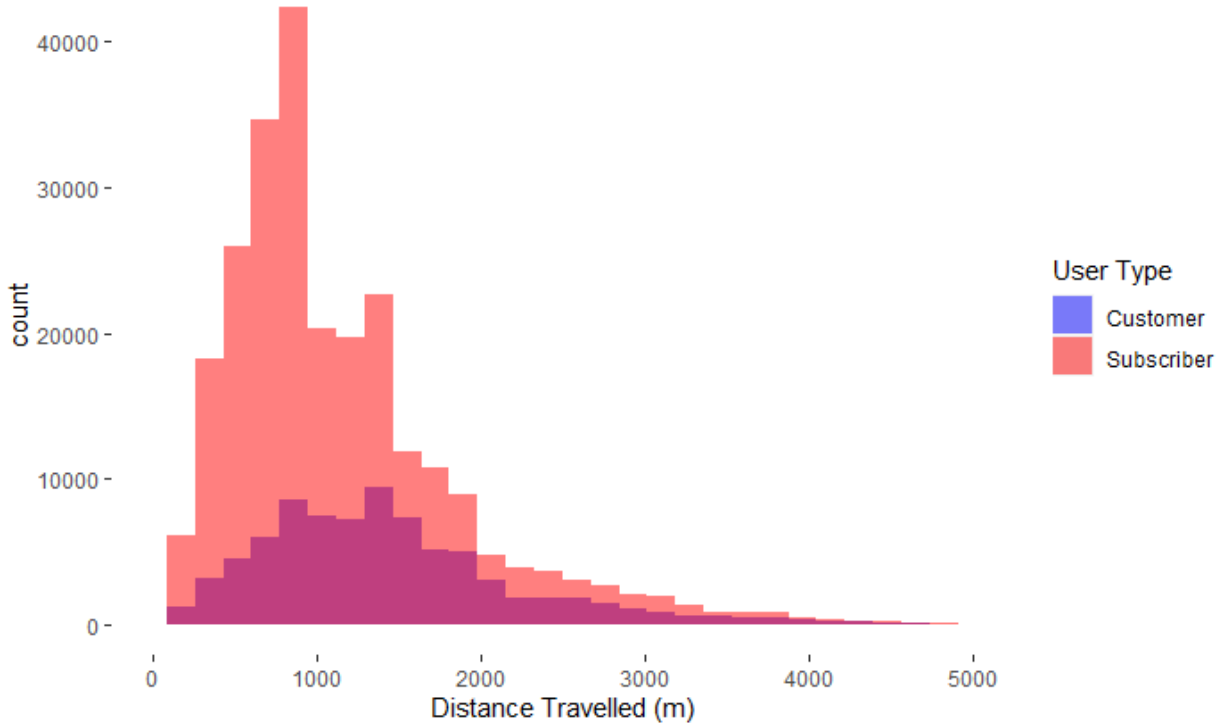


Figure 3

Plotting Distance Travelled via a Histogram with an Upper Limit of 5000



The purpose of adding an additional histogram with an upper limit is for the sole purpose of being able to have better visibility relative to the area with the majority of the data. While this chart does leave out some outlier observations, the previous histogram as well as following boxplot was included for the sake of visualizing the outliers.

Figure 4

Plotting Distance Travelled with a Boxplot

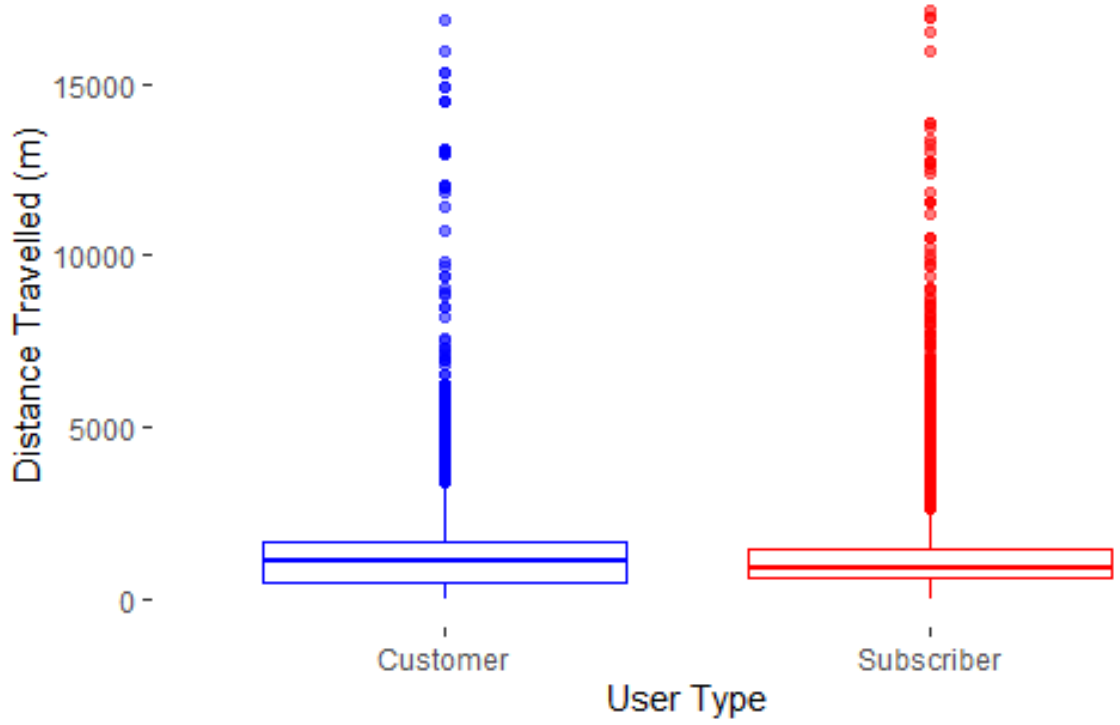
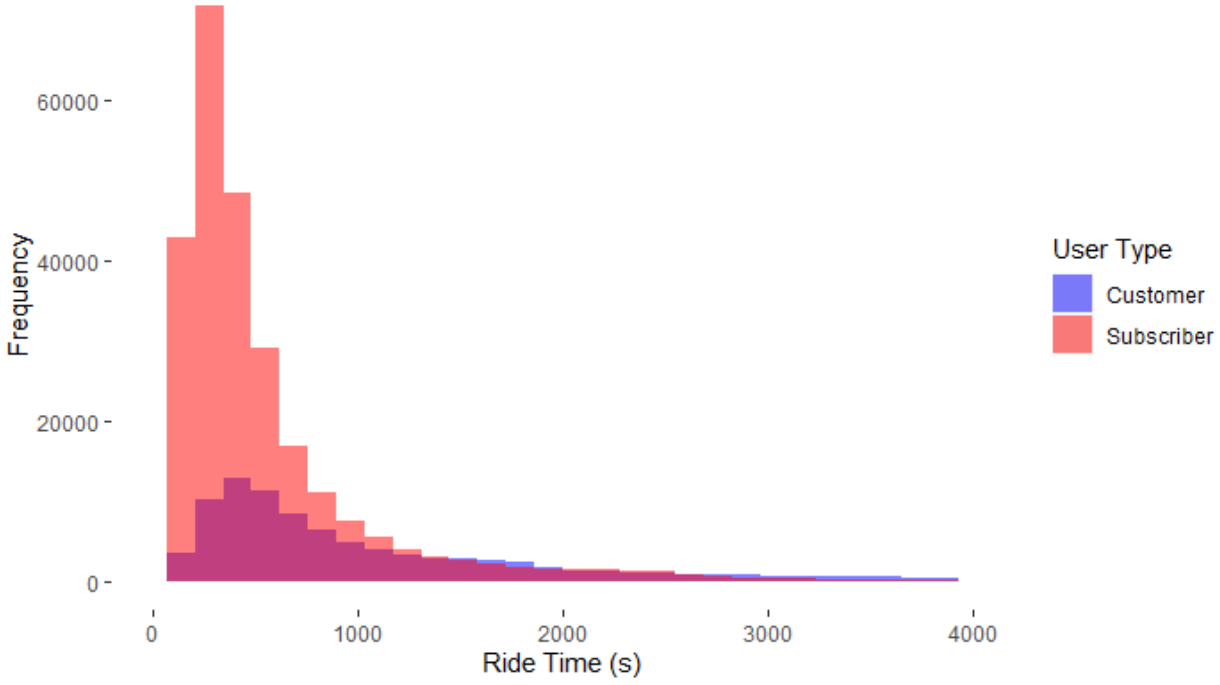


Figure 4

Plotting Travel Duration with a Histogram

Figure 6

Plotting Travel Duration via a Histogram with an Upper Limit of 4000



Once again, the purpose of adding an additional histogram with an upper limit is for the sole purpose of being able to have better visibility relative to the area with the majority of the data. While this chart does leave out some outlier observations, the previous histogram as well as following boxplot was included for the sake of visualizing the outliers.

Figure 7

Plotting Travel Duration with a Boxplot

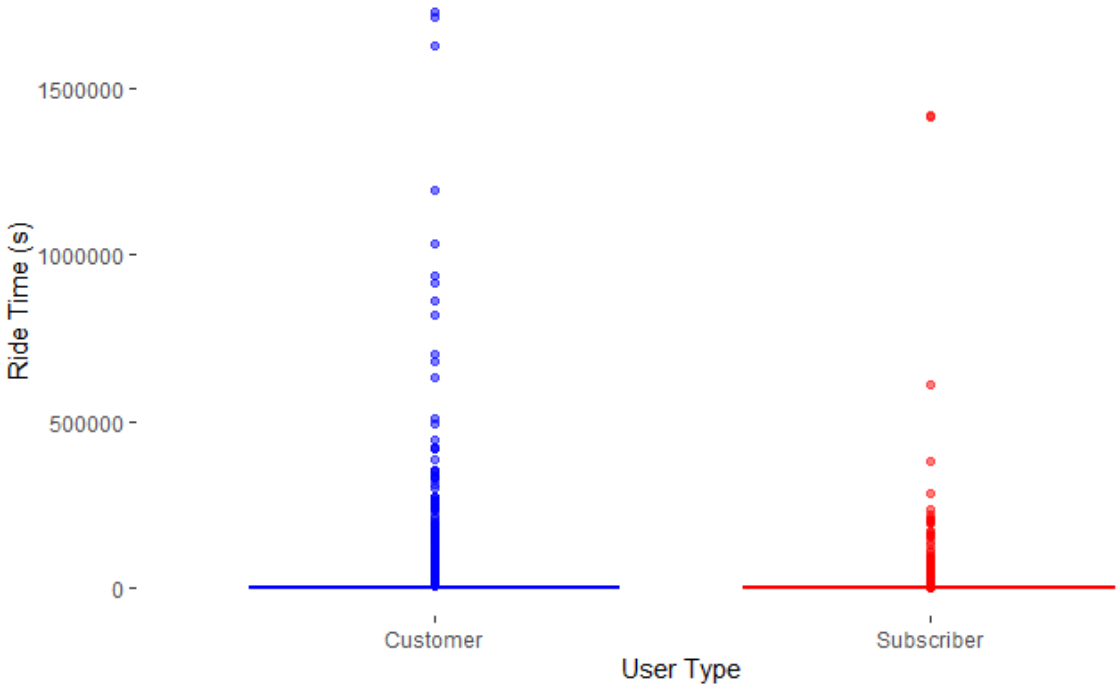


Figure 8

Station Preference for Start Location

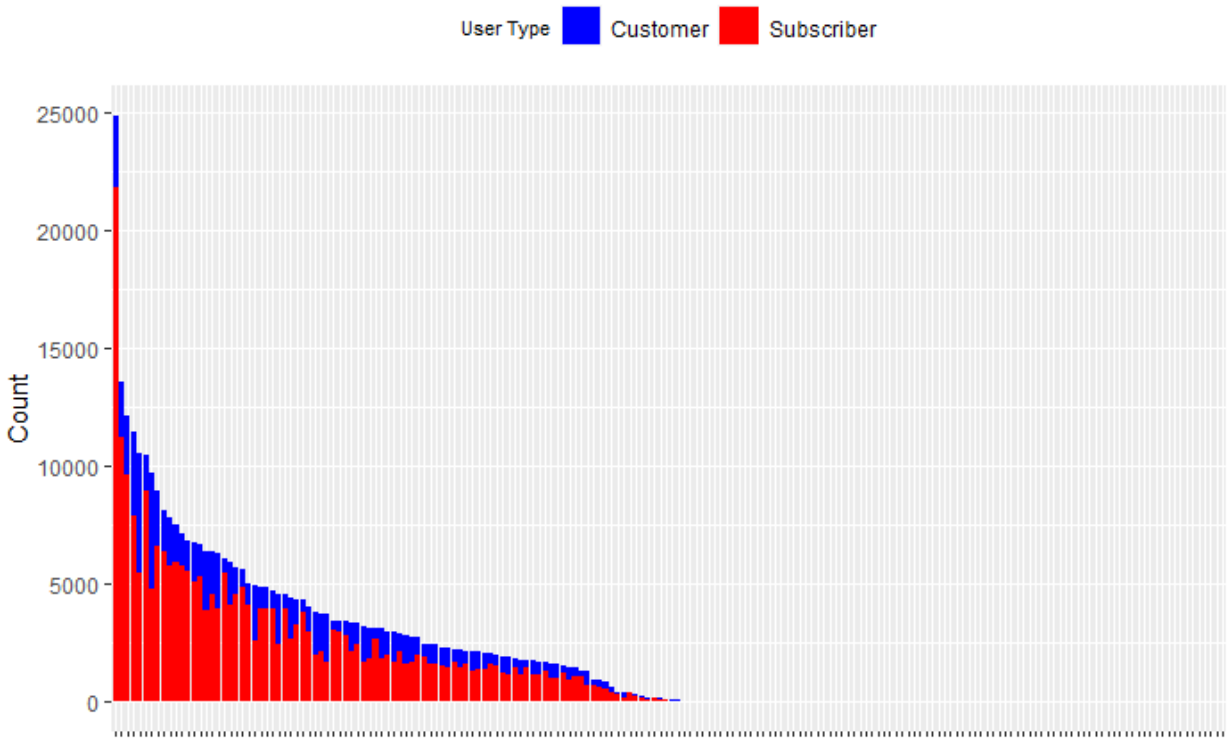


Figure 9

Top 20 Station Preferences for Start Location

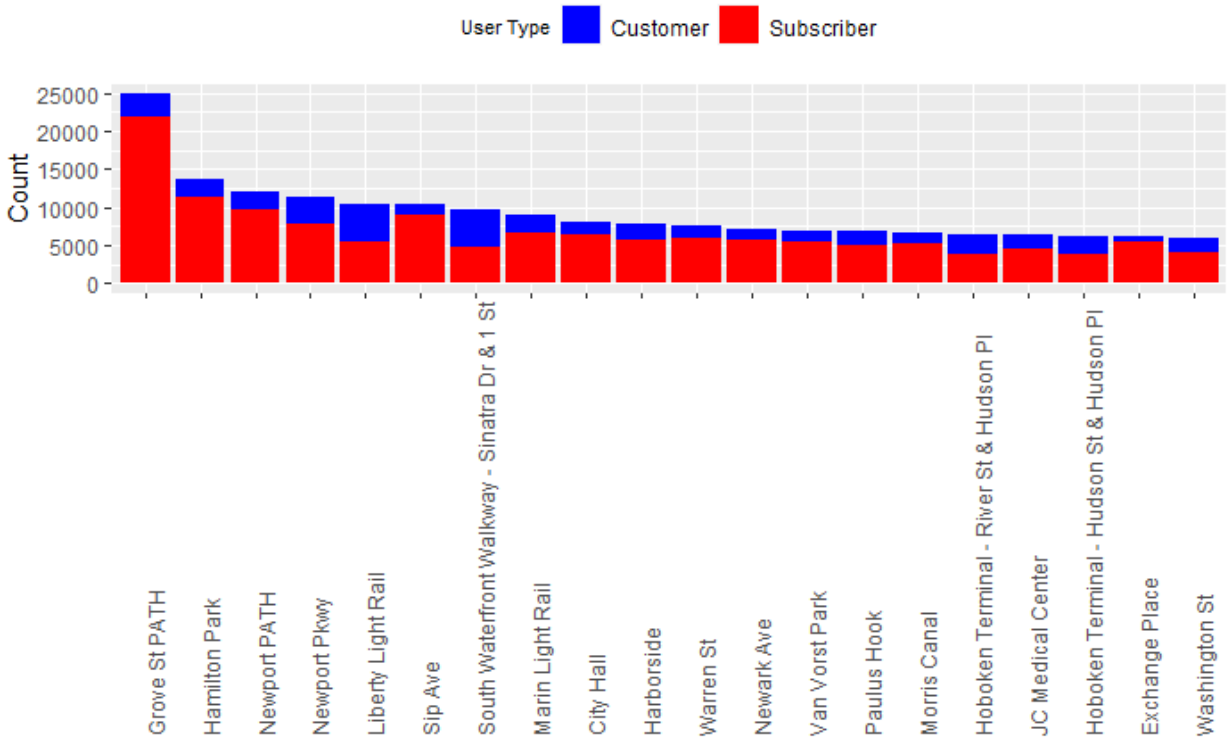


Figure 10

Station Preference for End Location

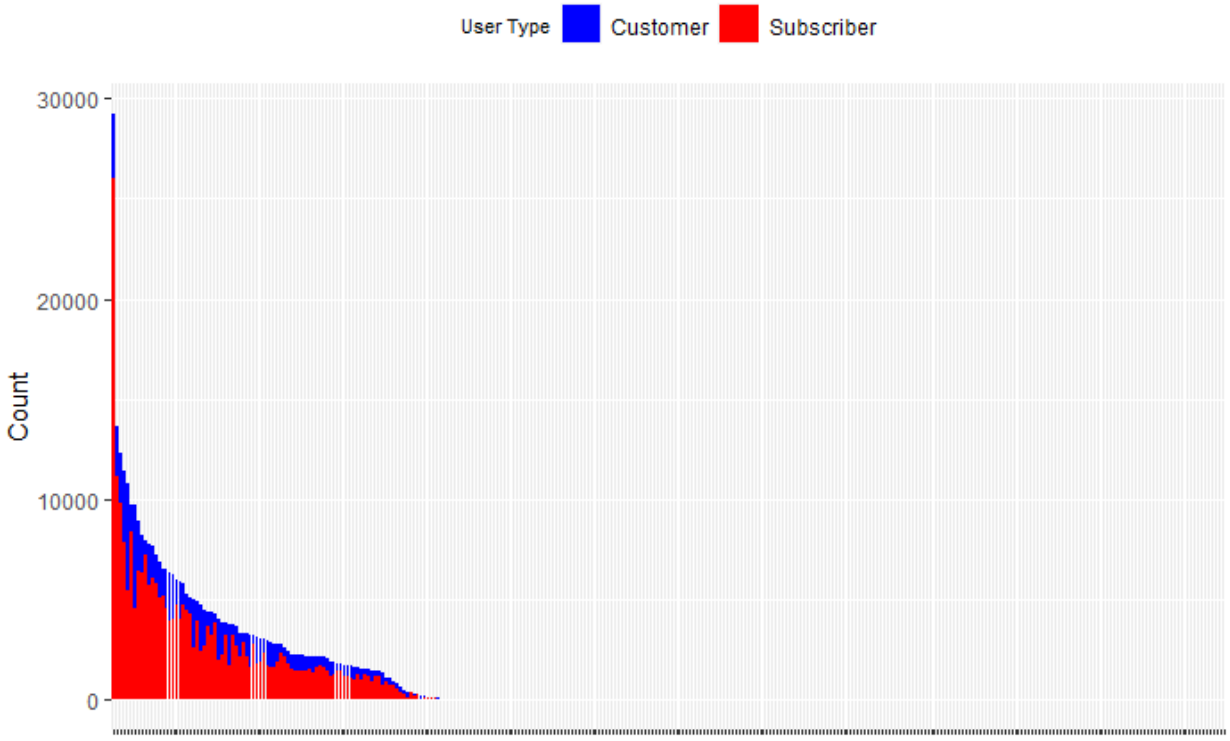


Figure 11

Top 20 Station Preferences for End Location

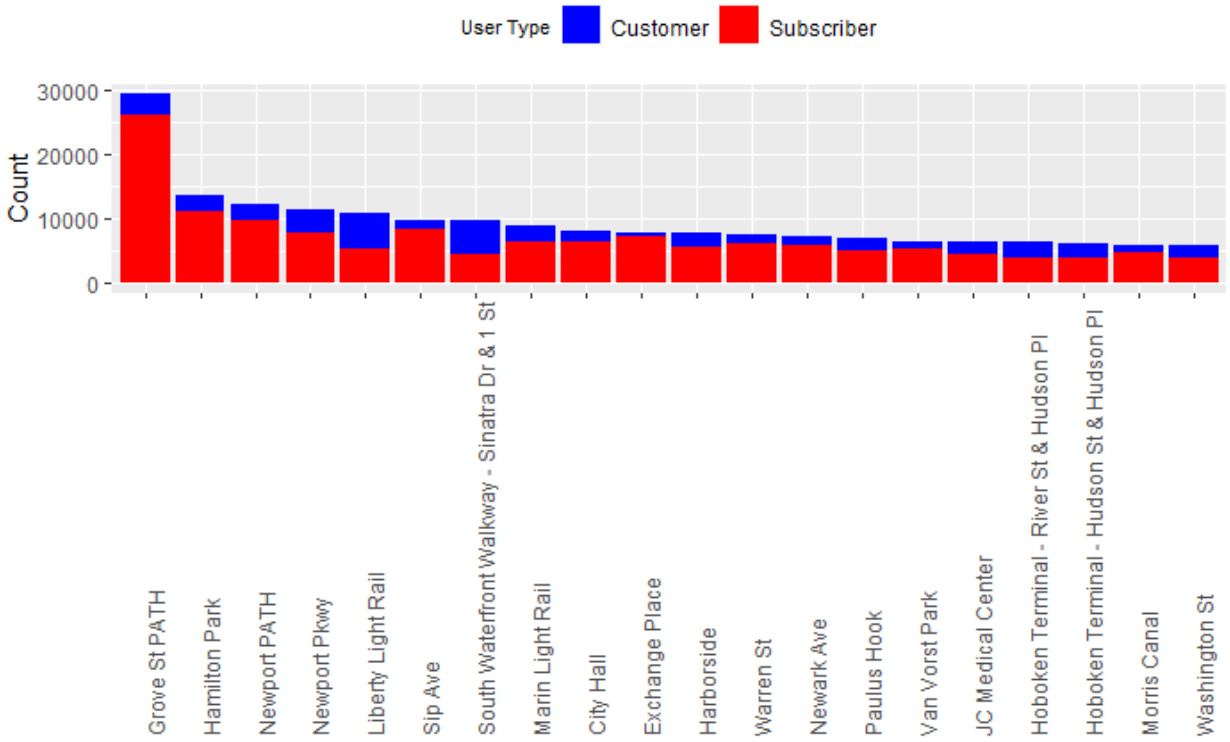


Figure 12

Proposed Plan and Pricing Model for Citi Bike

	Single Ride	Day Pass	Lyft Pink
Charge/Freq.	$4.99 / 30 minutes	$19.99 / day	$199 / year
Unlock fee	Variable	Free	Free
Ride time	30 minutes	30 minutes	45 minutes
Late fee	$5 / 15 minutes	$5 / 15 minutes	20¢ / minute
Other benefits?	N/A	N/A	All former Citi Bike tier perks Benefits when using Lyft Free Grubhub+ membership