User Behavior and Citi Bike:

A Study of Behavioral Economics and Data Analytics

Table of Contents

01

Understanding Citi Bike

Citi Bike Locations in the NYC Area

Plan & Pricing Comparison

	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

Break-Even Point Considerations

Single Ride

Best for infrequent users doing 2 rides in a day or less.

Day Pass

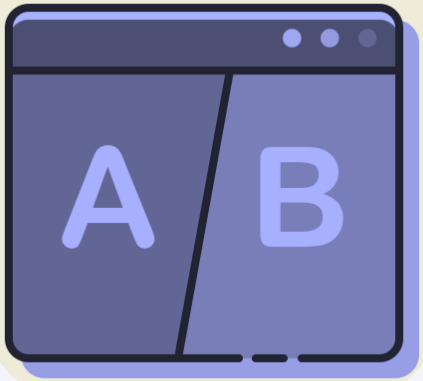
Best for infrequent users doing over 2 rides in a day.

Yearly Plan

Best for those who use the service over 20 times per year.

02

Proposing an A/B Test



Formation of Groups

Customers

Classified as one-time or infrequent users that use one of the following payment plans:

- Single Ride
- Day Pass

Subscribers

Classified as frequent or daily users that use one of the following payment plans:

- Citi Bike
- Lyft Pink

Devising a Research Question

Is there a discernible difference in the biking behavior between subscribers and customers, particularly in terms of trip distance, duration, and station preferences, and how might these distinctions inform strategies to enhance the Citi Bike service?

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in the biking behavior between subscribers and customers in terms of trip distance, duration, and station preferences.

Alternate Hypothesis

There is a discernible difference in the biking behavior between subscribers and customers, with variations in trip distance, duration, and station preferences.

All tests will be conducted with a significance level set at α=.05.

Three Criteria to Consider

Distance

Distance will be calculated using the Haversine formula.

Duration

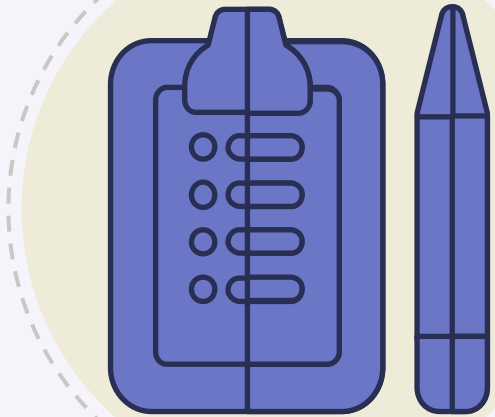
Duration will be calculated by subtracting the end time from the beginning time.

Station Choice

Station preferences will be examined via a revealed preference model.

03

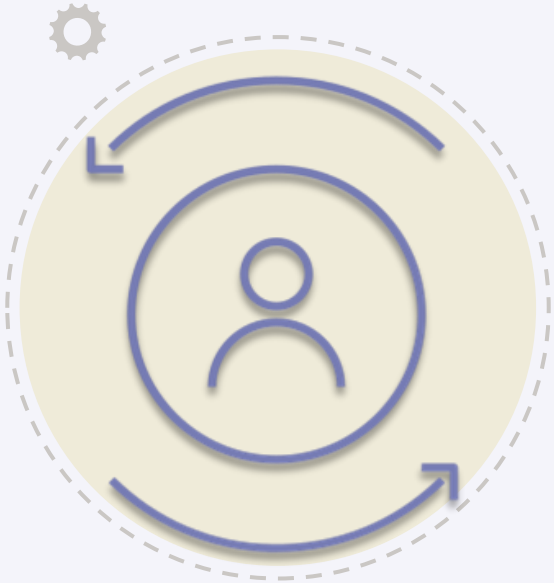
Methodology & Results



362,100

Number of observations in the month of May from 2016-2023 after the data was cleaned.

vs.



98,067

Customer Rides

264,033

Subscriber Rides

Distance Disparities

The Null and Alternate Hypotheses

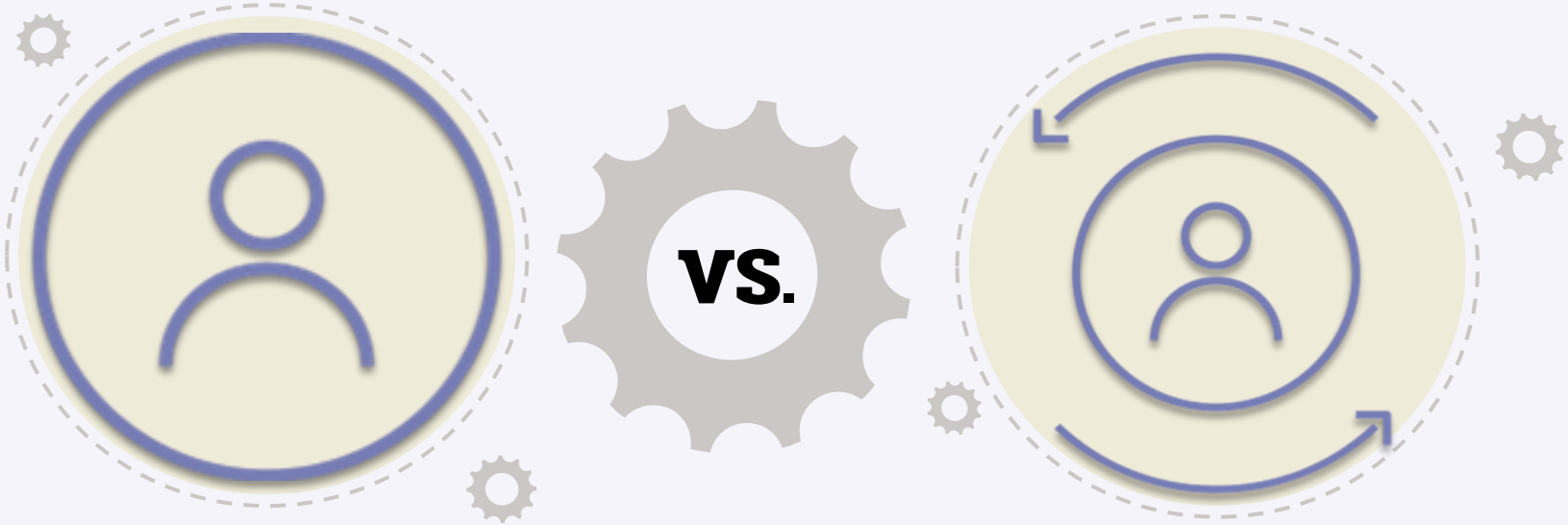
Null Hypothesis

There is **no** discernible difference in trip distance between subscribers and customers within the Citi Bike service.

Alternate Hypothesis

There is a discernible difference in trip distance between subscribers and customers within the Citi Bike service.

Mean Distance Travelled by Each Group



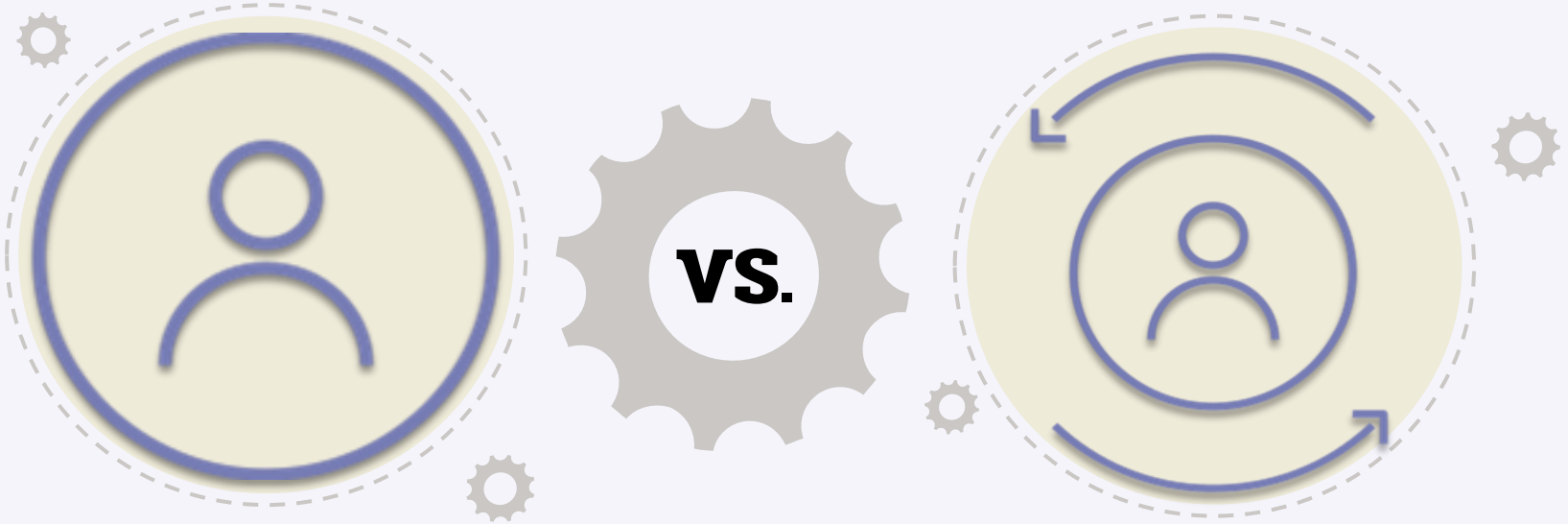
1169.9

Customer Distance (meters)

1082.6

Subscriber Distance (meters)

Standard Deviation of Distance Travelled



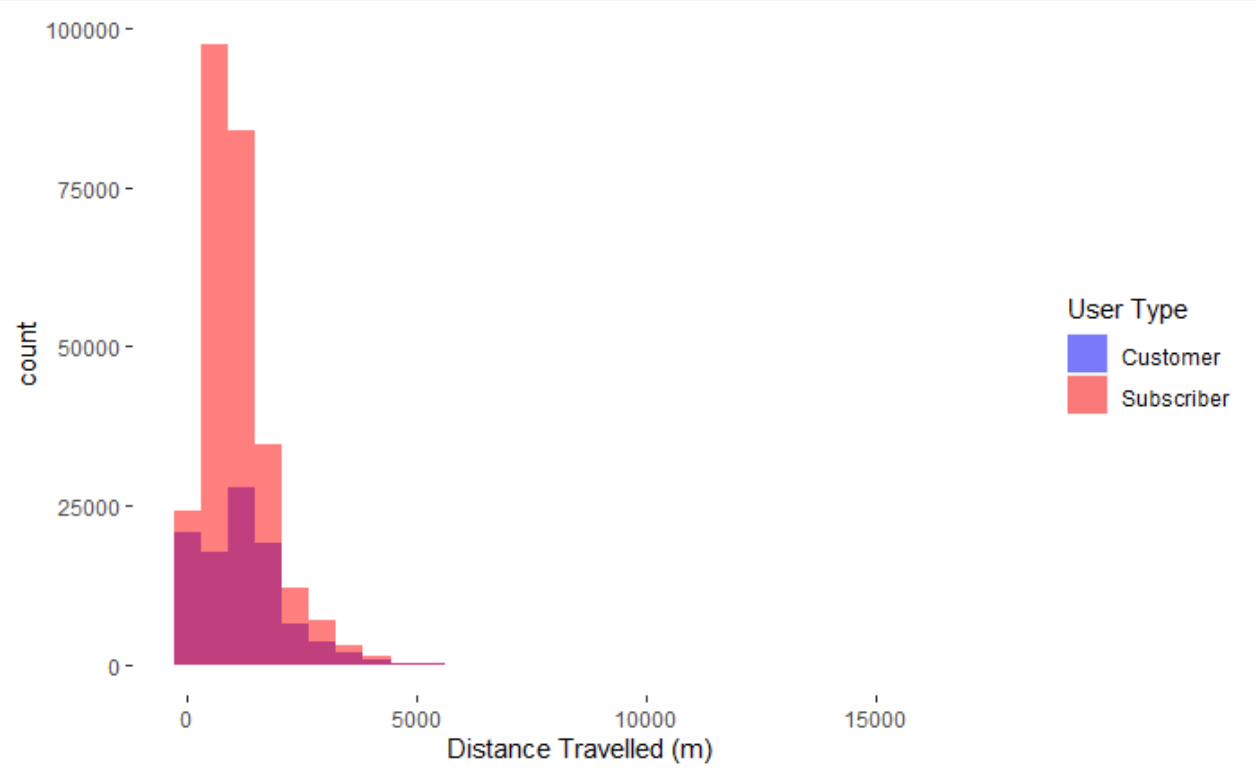
930.5

Customer Distance (meters)

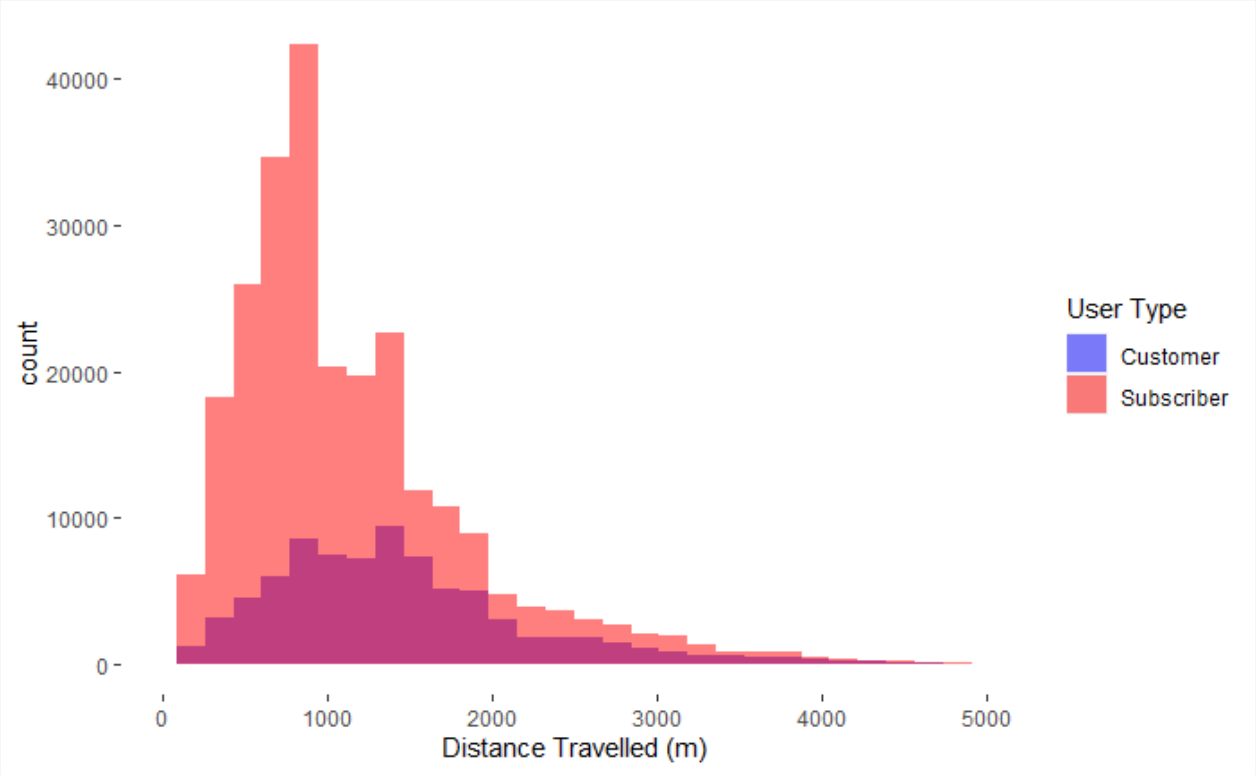
763.2

Subscriber Distance (meters)

Plotting Distance Travelled with a Histogram



Plotting Distance Travelled with a Histogram



Plotting Distance Travelled with a Boxplot

Two Sample T-Test Outcomes

	Customers	Subscribers
Mean	1,168.613	1,081.426
Standard Deviation	930.521	763.203
Observations	98,067	264,033
Degrees of Freedom	149,722	
T-value	26.275	

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in trip distance between subscribers and customers within the Citi Bike service.

Alternate Hypothesis

There is a discernible difference in trip distance between subscribers and customers within the Citi Bike service.

Rejecting The Null Hypothesis

Based on the calculated t-value of approximately 26.3 with 149722 degrees of freedom, which exceeds the critical t-value at a 95% confidence level, we reject the null hypothesis.

Alternate Hypothesis

There is a discernible difference in trip distance between subscribers and customers within the Citi Bike service.

Other Considerations for Findings

- Both distributions exhibited a positive skew.
- Customer trip distance exhibited far more variance than subscribers, but both groups exhibited a great number of outliers.
- Customers were about three times as likely to start and end at the same station.
  - Customers: 14.86%
  - Subscribers: 4.86%
- As New York City's peak tourist seasons are June–August and November–December, the constraint of using data from the month of May could be leaving out important information regarding customer user patterns in busier periods.

Duration Differences

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in trip duration between subscribers and customers within the Citi Bike service.

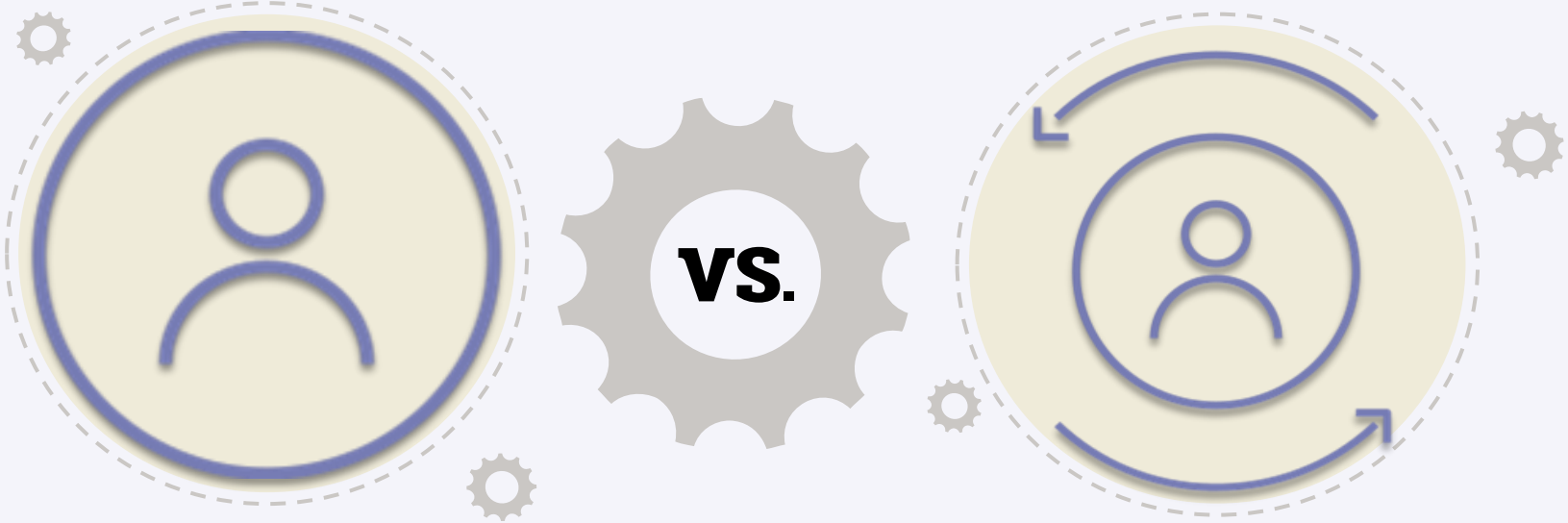
Alternate Hypothesis

There is a discernible difference in trip duration between subscribers and customers within the Citi Bike service.

Plan & Pricing Comparison

	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

Mean Trip Duration of Each Group



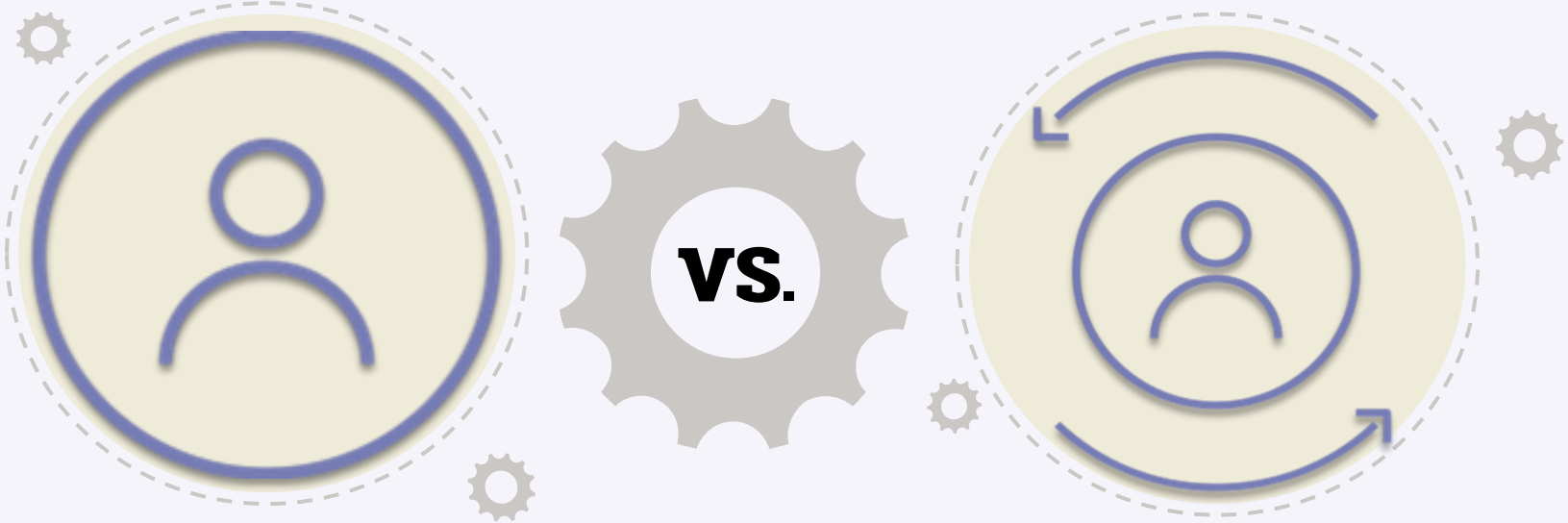
1984.3

Customer Duration (seconds)

649.3

Subscriber Duration (seconds)

Standard Deviation of Trip Duration



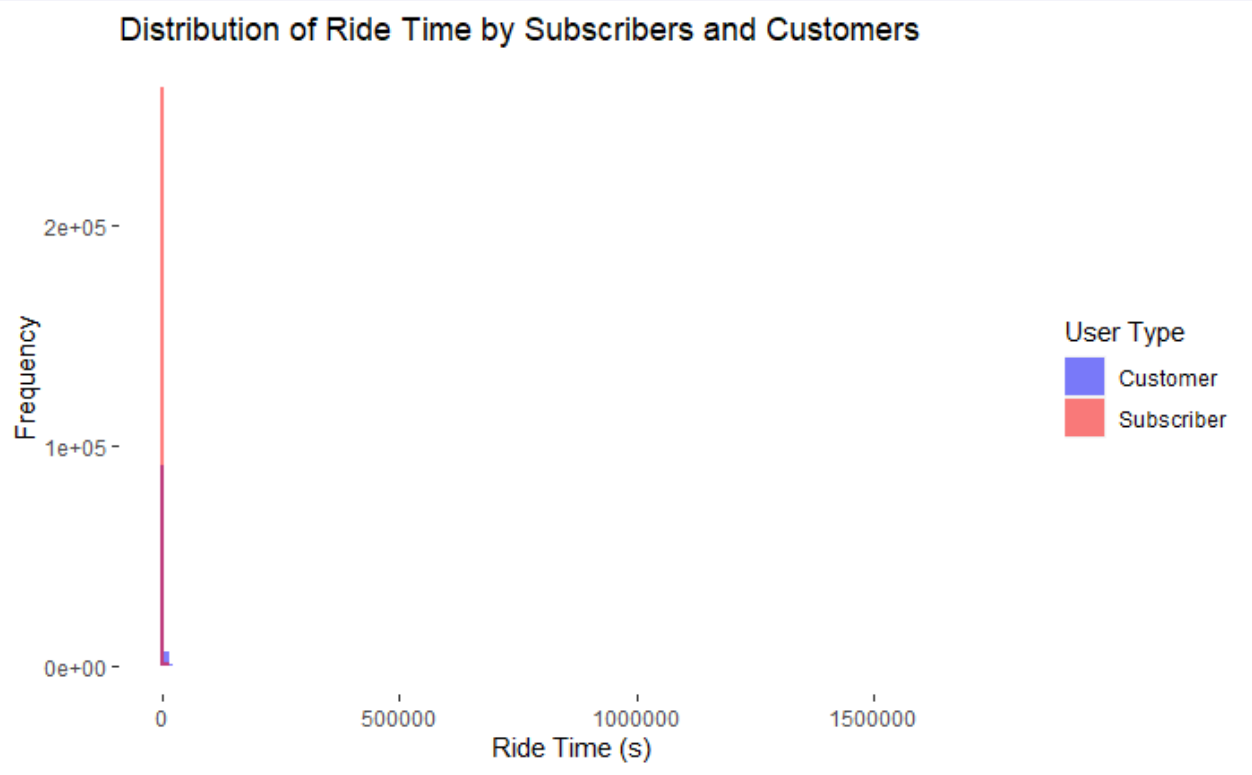
14745.3

Customer Duration (seconds)

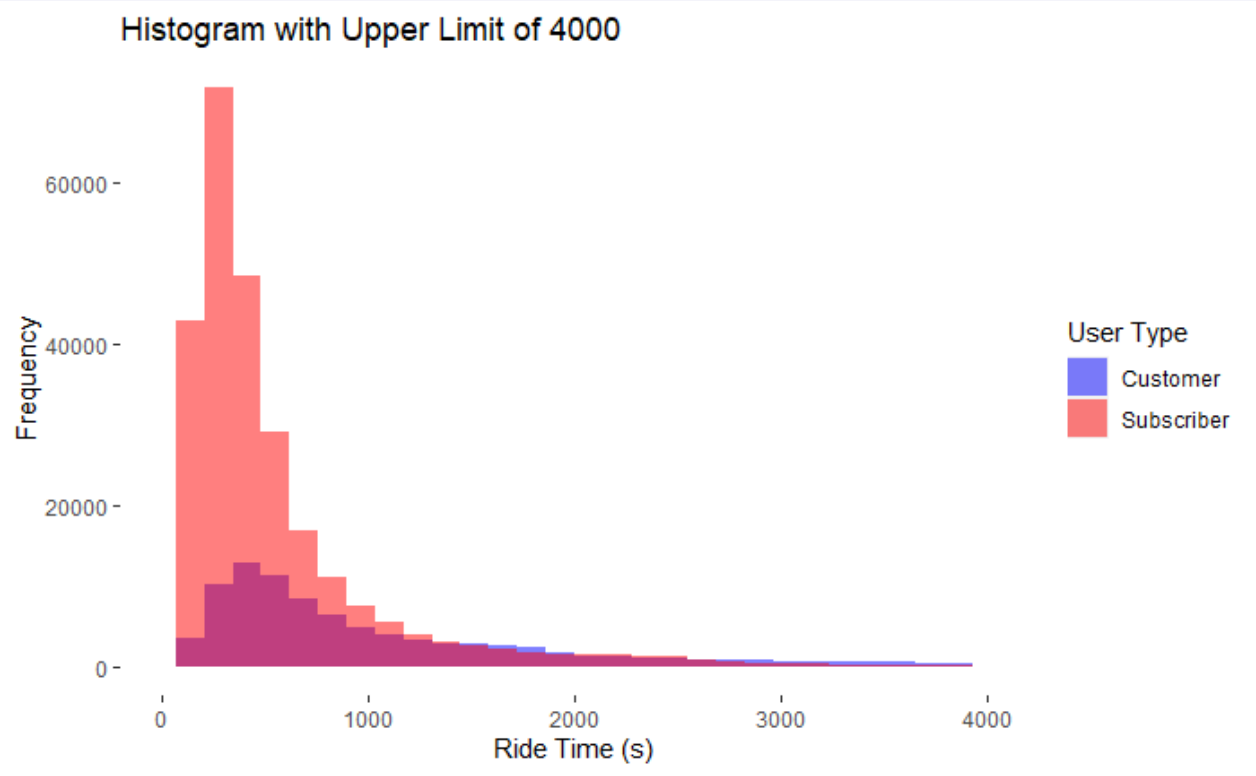
4686.2

Subscriber Duration (seconds)

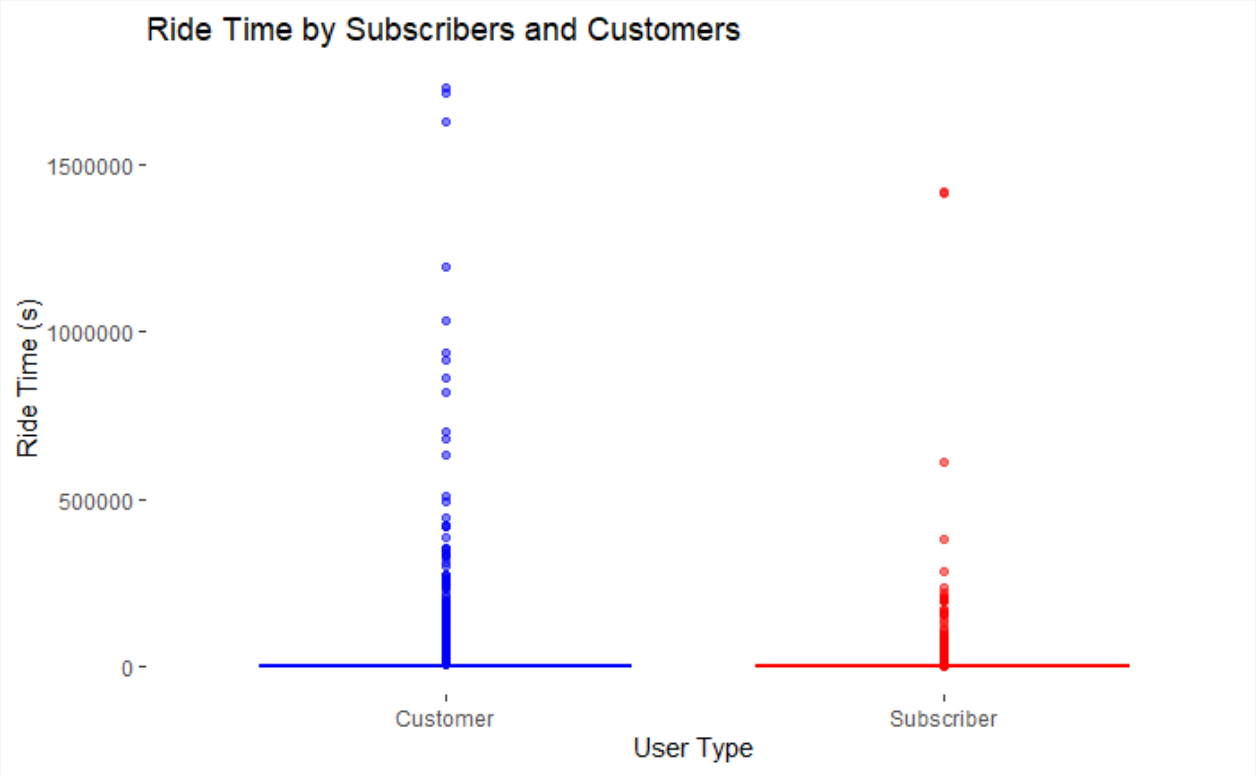
Plotting Travel Duration with a Histogram



Plotting Travel Duration with a Histogram



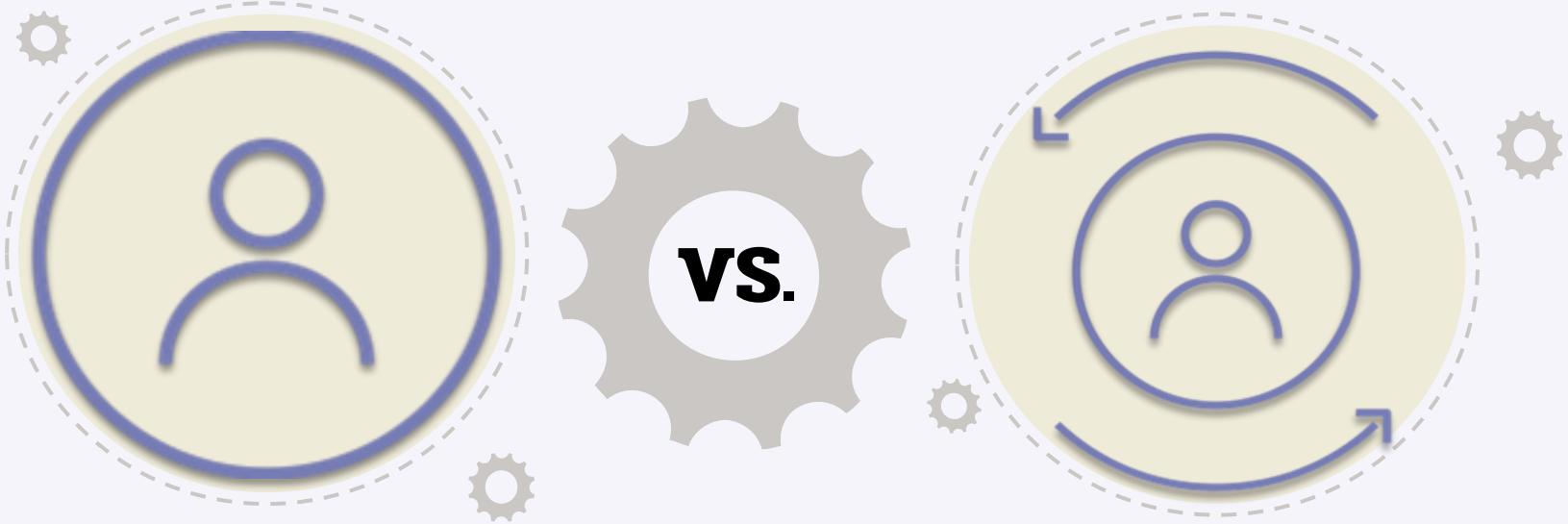
Plotting Travel Duration with a Boxplot



Plan & Pricing Comparison

	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

Tardiness Rate of Each Group



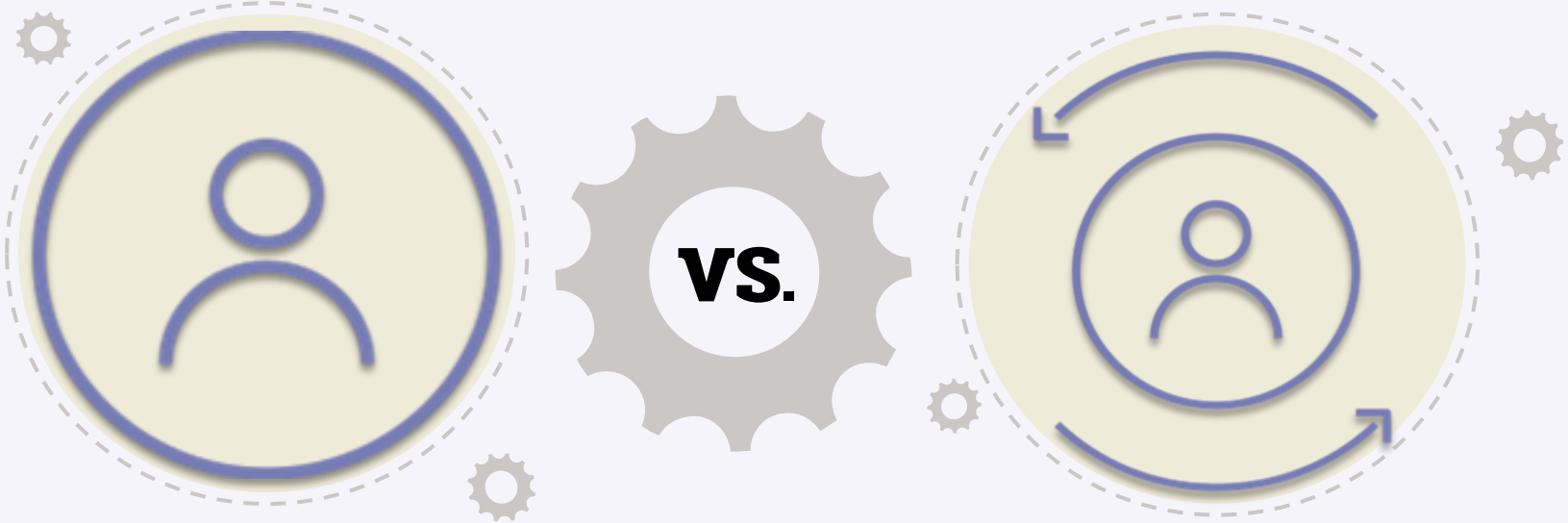
23.02%

Customer Tardiness Rate

1.82%

Subscriber Tardiness Rate

Average Late Fee Applied to Each Group



$3.90

Average Customer Late Fee

26¢

Average Subscriber Late Fee

Two Sample T-Test Outcomes

	Customers	Subscribers
Mean	1,984.341	649.306
Variance	14,745.329	4686.230
Observations	98,067	264,033
Degrees of Freedom	105,507	
T-value	27.836	

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in trip duration between subscribers and customers within the Citi Bike service.

Alternate Hypothesis

There is a discernible difference in trip duration between subscribers and customers within the Citi Bike service.

Rejecting The Null Hypothesis

Based on the calculated t-value of approximately 27.8 with 105507 degrees of freedom, which exceeds the critical t-value at a 95% confidence level, we reject the null hypothesis.

Alternate Hypothesis

There is a discernible difference in trip duration between subscribers and customers within the Citi Bike service.

Other Considerations for Findings

- Both distributions exhibited a positive skew.
- Customer trip duration exhibited far more variance than subscribers, but both groups exhibited a great number of outliers.
  - This could be due to customers not reading or understanding the model of time limits and additional fees.
  - These additional fees are a major source of income for Citi Bike.
- Even with an additional fifteen minutes of ride time, subscribers' trips were about three times shorter than customers, but this is partly attributable to outliers.
  - Customers: 33 minutes and 4 seconds
  - Subscribers: 10 minutes and 49 seconds
- As New York City's peak tourist seasons are June–August and November–December, the constraint of using data from the month of May could be leaving out important information regarding higher tardiness rates when more tourists are in the city.

Station Selection

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in station preferences between subscribers and customers within the Citi Bike service.

Alternate Hypothesis

There is a discernible difference in station preferences between subscribers and customers within the Citi Bike service.

A chi-square test will be conducted instead of t-test.

Most Frequent Stations for Customers

1. Liberty Light Rail – 5062
2. South Waterfront Walkway - Sinatra Dr & 1 St – 4950
3. Newport Pkwy – 3586
4. Grove St PATH – 3052
5. Hoboken Terminal - River St & Hudson Pl – 2535

1. Liberty Light Rail – 5340
2. South Waterfront Walkway - Sinatra Dr & 1 St – 5153
3. Newport Pkwy – 3549
4. Grove St PATH – 3218
5. Hamilton Park – 2524

Top 5 Start Locations

Top 5 End Locations

Most Frequent Stations for Subscribers

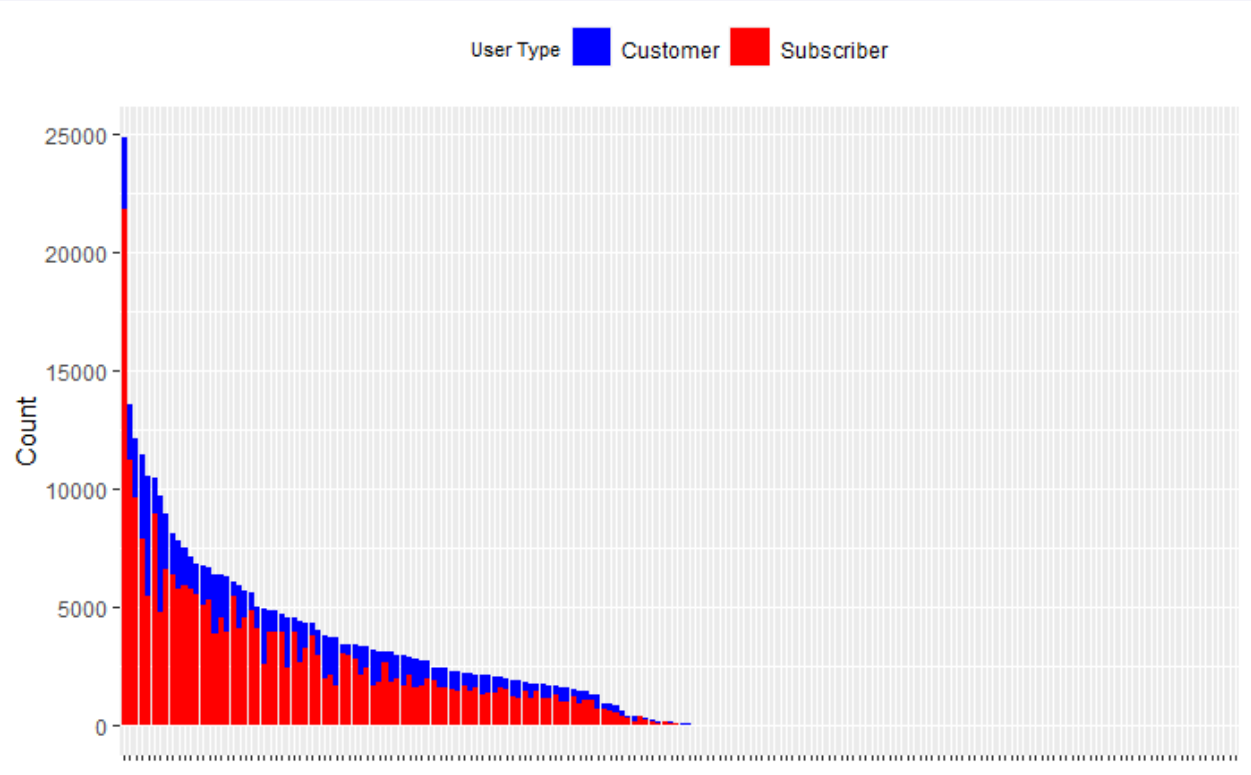
1. Grove St PATH – 21825
2. Hamilton Park – 11228
3. Newport PATH – 9617
4. Sip Ave – 8932
5. Newport Pkwy – 7852

Top 5 Start Locations

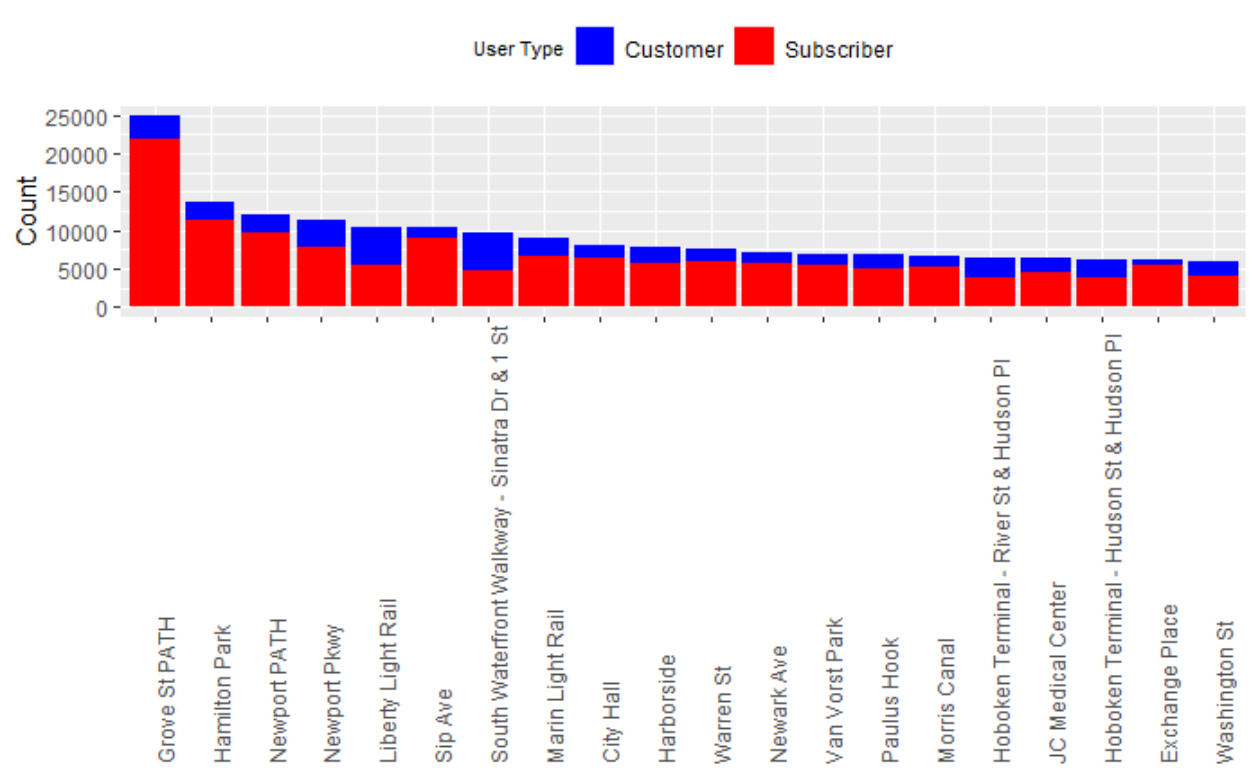
1. Grove St PATH – 26043
2. Hamilton Park – 11106
3. Newport PATH – 9848
4. Sip Ave – 8385
5. Newport Pkwy – 7880

Top 5 End Locations

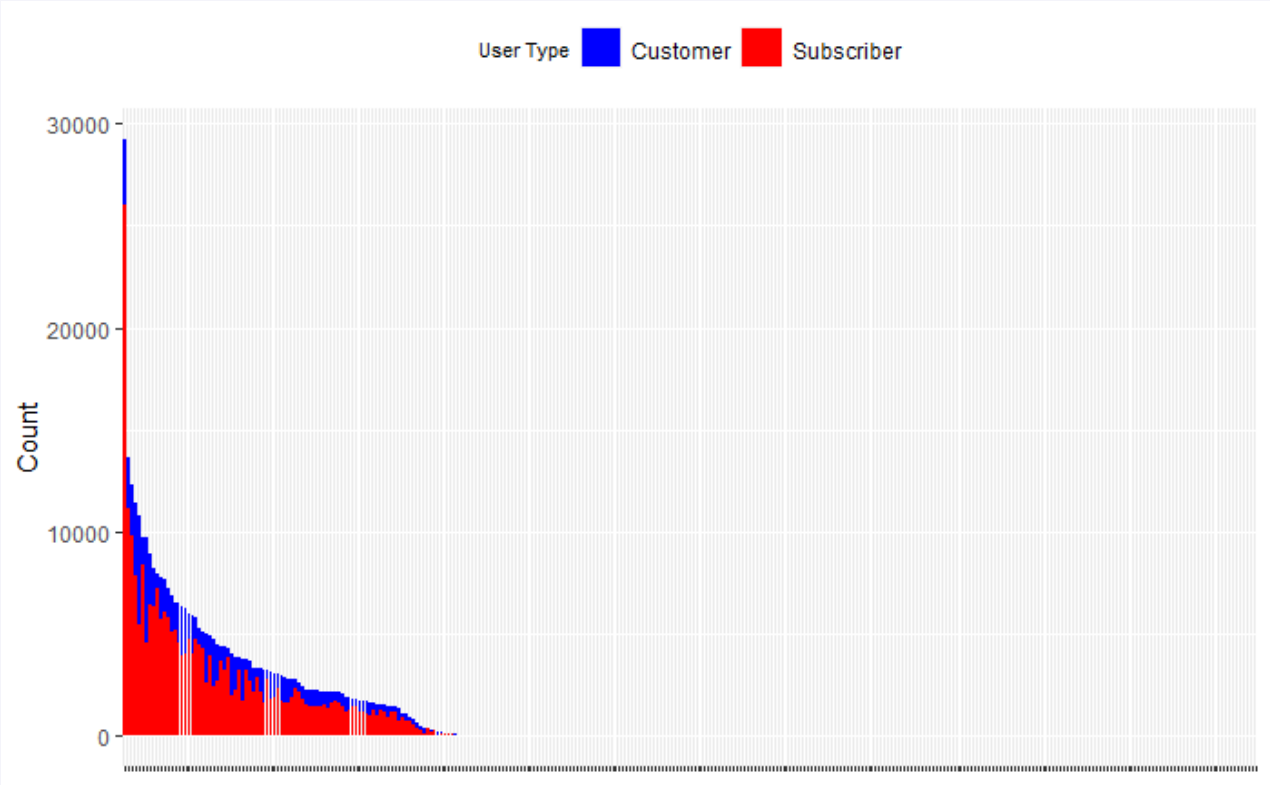
Station Preference for Start Location



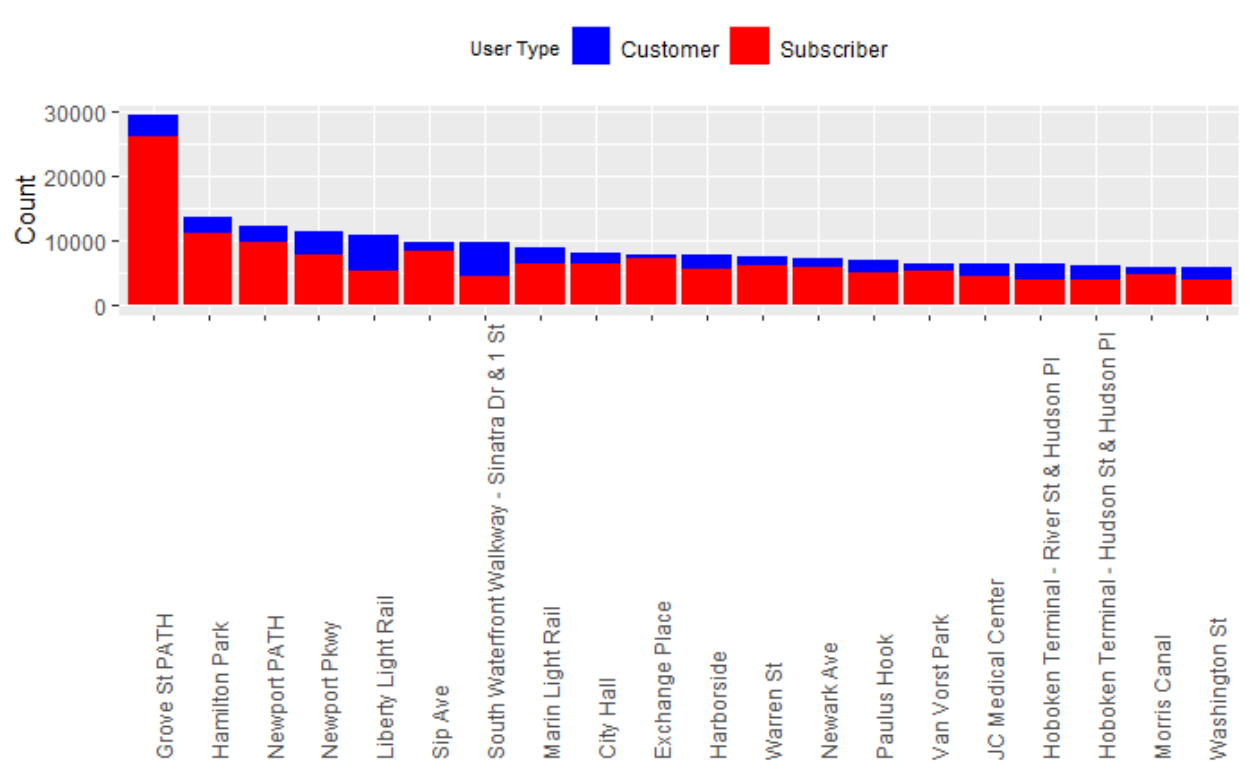
Top 20 Station Preferences for Start Location



Station Preference for End Location



Top 20 Station Preferences for End Location



Chi-Squared Test Outcomes

	Start Station	End Station
X-squared	26427	29395
Degrees of Freedom	183	318
P-value	2.2e-16	2.2e-16

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in station preferences between subscribers and customers within the Citi Bike service.

Alternate Hypothesis

There is a discernible difference in station preferences between subscribers and customers within the Citi Bike service.

Rejecting The Null Hypothesis

Based on the calculated p-value of approximately 2.2e-16 for both start and end station preferences, which exceeds the critical p-value at a 95% confidence level, we reject the null hypothesis.

Alternate Hypothesis

There is a discernible difference in station preference between subscribers and customers within the Citi Bike service.

Other Considerations for Findings

- Grove St PATH is by far the most popular station for starting and ending one's trip.
  - This station is not located in New York City, but rather in Jersey City, NJ.
  - It is right next to a subway station which can take people into southern Manhattan via subway
- The top eight starting locations are also the top eight ending locations.
- 90 stations were used less than 50 times to start a trip.
- 225 stations were used less than 50 times to end a trip.

04

Market Opportunities

The Null and Alternate Hypotheses

Null Hypothesis

There is **no** discernible difference in the biking behavior between subscribers and customers in terms of trip distance, duration, and station preferences.

Alternate Hypothesis

There is a discernible difference in the biking behavior between subscribers and customers, with variations in trip distance, duration, and station preferences.

Rejecting The Null Hypothesis

Alternate Hypothesis

There is a discernible difference in the biking behavior between subscribers and customers, with variations in trip distance, duration, and station preferences.

Plan & Pricing Comparison

	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

Optimizing the Citi Bike Model

Simplify plan options

Remove the Citi Bike plan as Lyft Pink is a currently a cheaper option with more perks.

Demand-based pricing for single rides

Consider varying the price of unlock fees for single riders based on the supply and demand of bikes.

Increase late fees

Give people a greater incentive to return their bikes on time while capitalizing on those who fail to read the terms and conditions.

Analyze existing infrastructure

Focus on adding stations near areas with high foot traffic and subway stations.

Plan & Pricing Comparison

	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

Thanks!

Do you have any questions?