Citi Bike: Analysis of the bike-sharing program in New York City

Farahin Choudhury, Clayton Monroe, Patrick Bush, Leigh Sausville, Destiny McDaniel

IST687 | GROUP #2

Ridership in 2019 vs Ridership in 2020 (the year of Covid-19)

Overview:

How has Covid-19 impacted ridership in the Citibike program?

Supporting questions

- Why use the bikes instead of the subway/bus/Uber etc?
- What are the most frequent times people are riding?
- Which age group is the program popular among?
- Which locations see the most/least activity?

Difficulties with the data

We learned (albeit the hard way) that our data wasn't perfect.

- Wrong years (1800s?)
- Missing data for gender (0 = unknown gender)
- Stations not broken down by borough
- Inconsistencies in data collection
    - as a result, outliers needed to be removed

Changes we made to the data

tripduration

Trip duration was collected and recorded in seconds, so it was converted from seconds to minutes

gender

Using the guide provided on the website, the numbers in the "gender" column were changed to "male", "female" and "unknown"

Removed outliers

Part of the data munging processes involved cleaning the data and removing any outliers that did not make sense for this analysis

Data analysis methods

Descriptive statistics

Tools included t-tests on trip duration and comparing the numbers of user types (customers vs subscribers)

Plotting on maps using ggplot2

The data made the most sense to be plotted on a map, so the top ten start stations were plotted for both 2019 and 2020

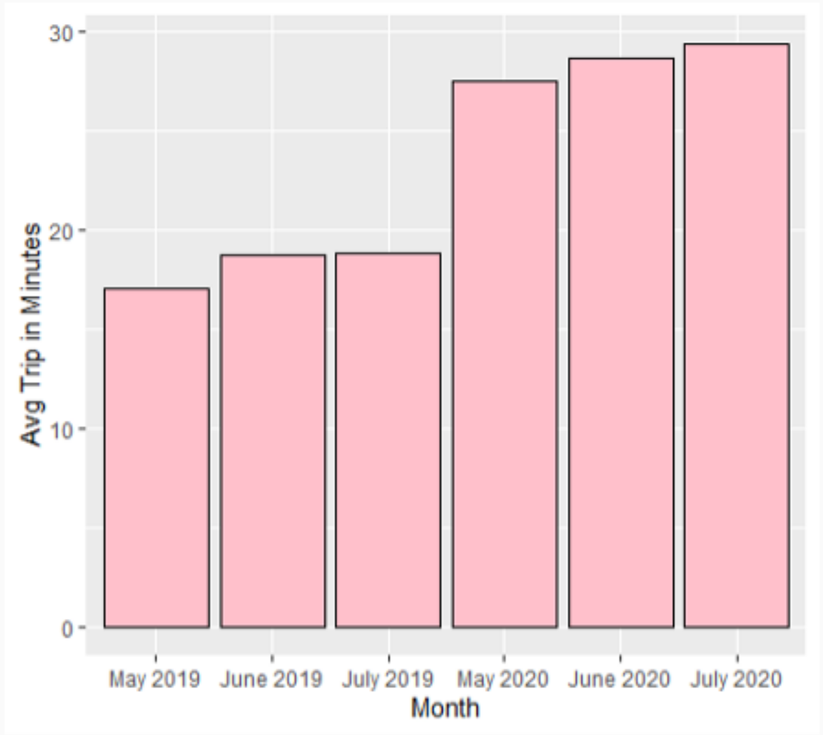
Merging large datasets

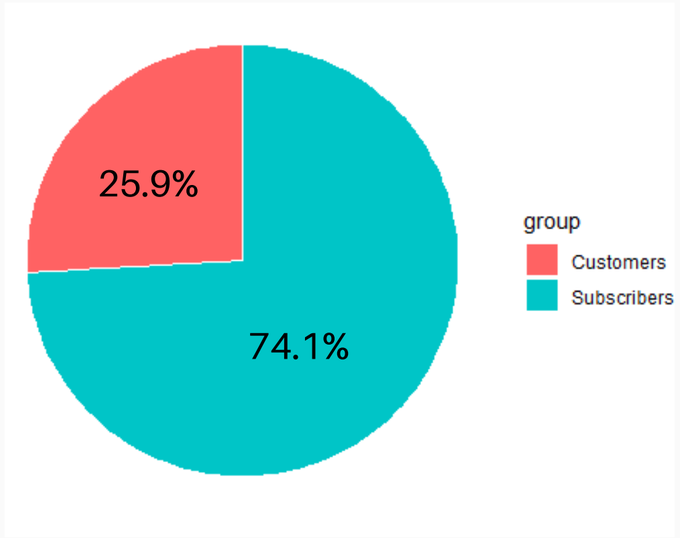
1) Datasets for May through July 2019 were combined together

2) Datasets from May through July 2020 were combined together

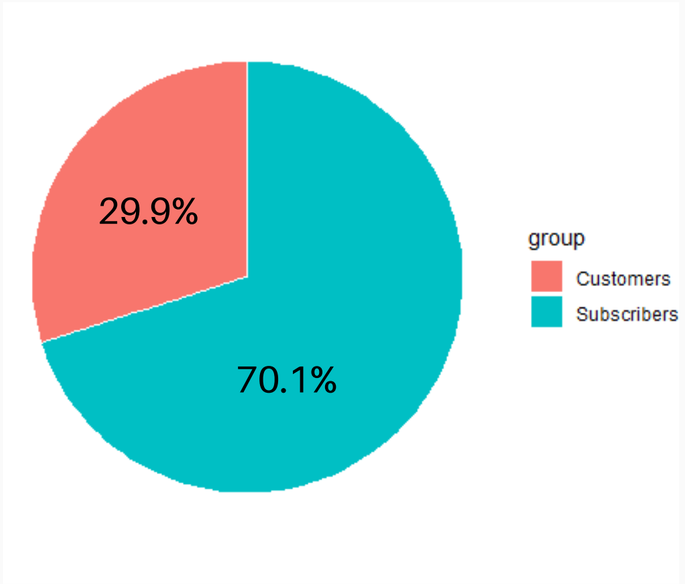
3) Both combined datasets were combined into one master dataset

Visualizations



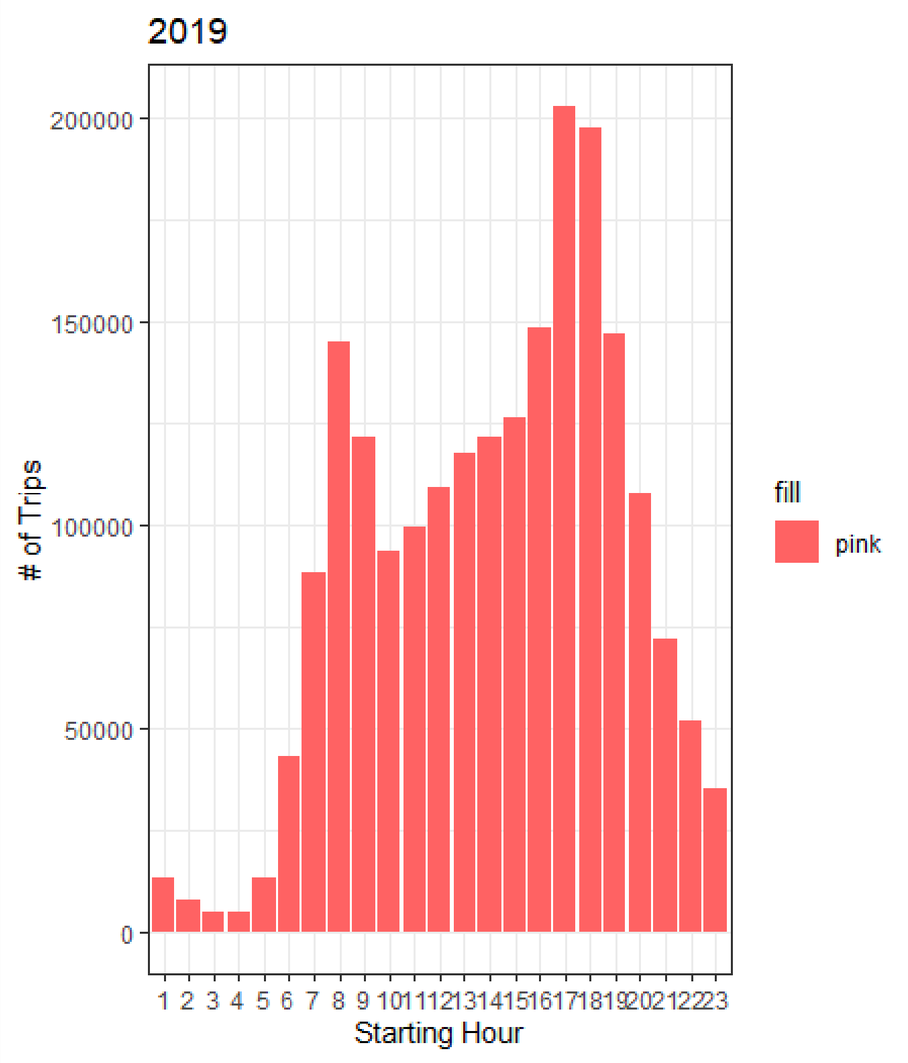


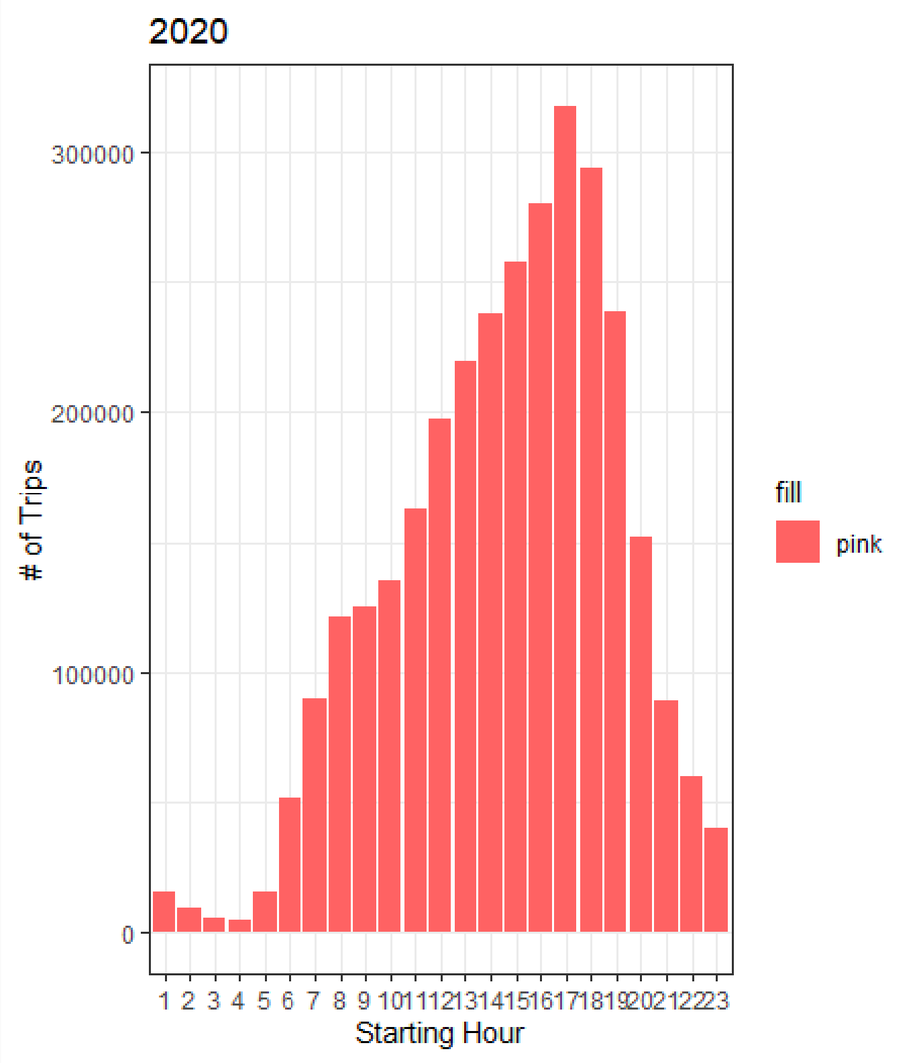
2019



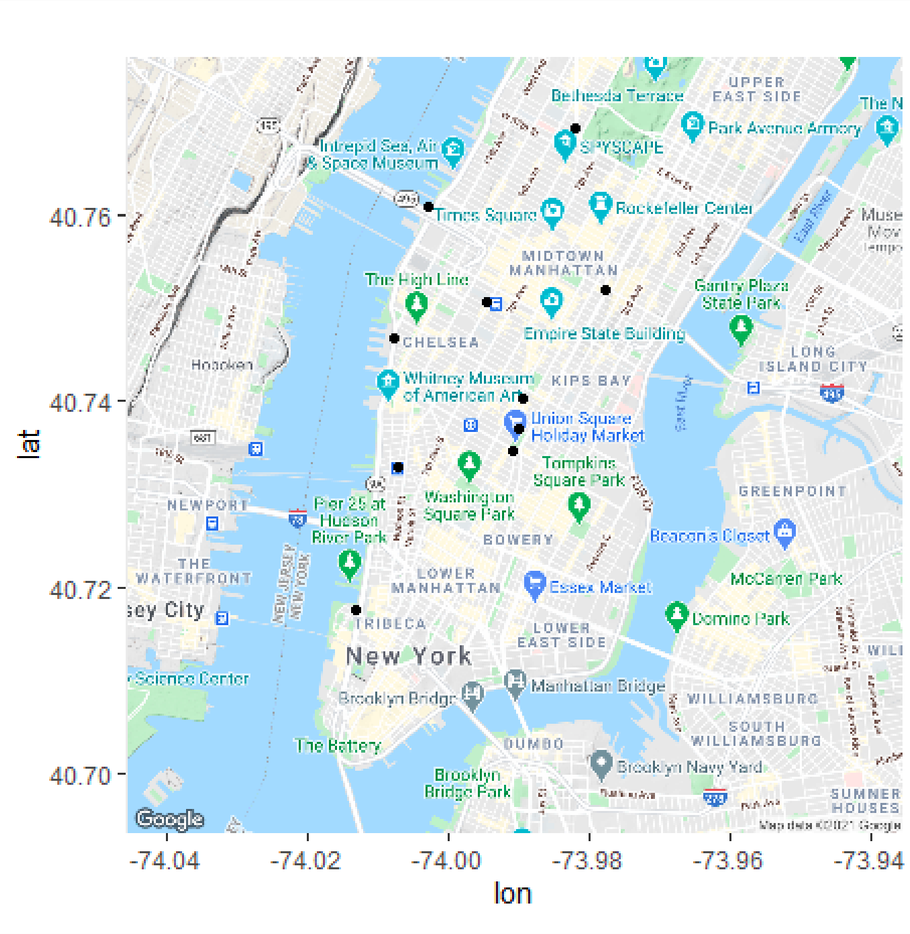
2020

Visualizations pt.2



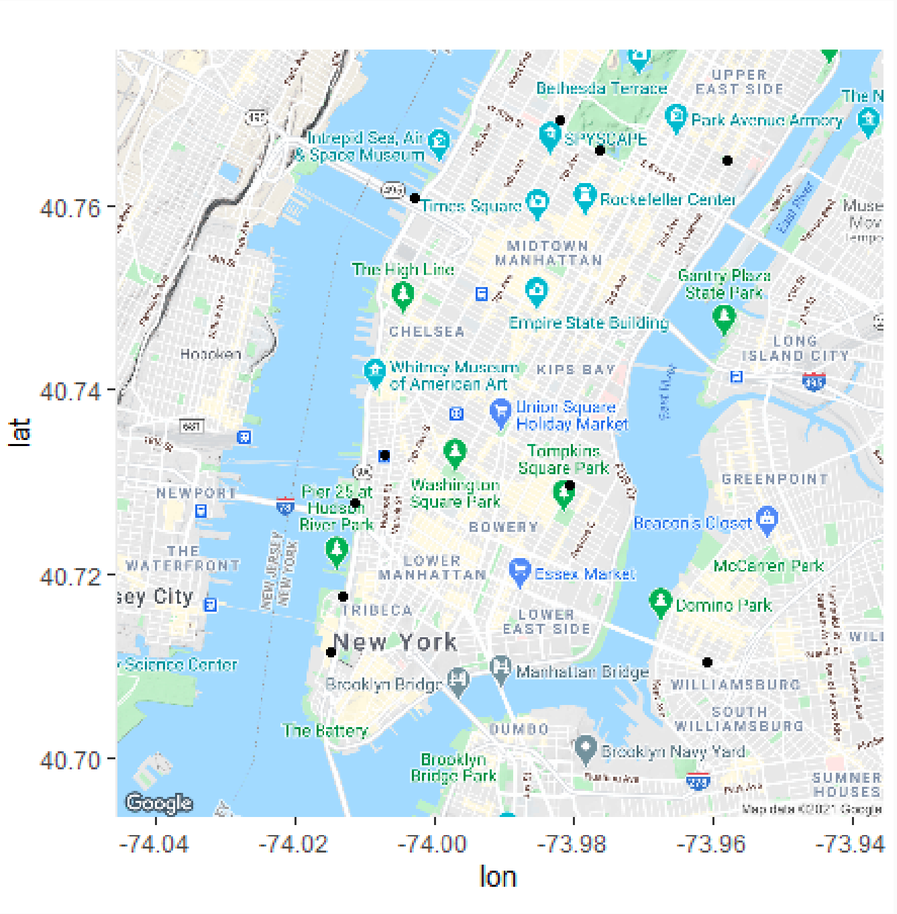


top 10 destinations 2019



1. **Pershing Sq N**
2. **West St & Chambers St**
3. **E 17 St & Broadway**
4. **12 Ave & W 40 St**
5. **Broadway & E 22 St**
6. **Broadway & E 14 St**
7. **8 Ave & W 31 St**
8. **Broadway & W 60 St**
9. **Christopher St & Greenwich St**
10. **W 20 St & 11 Ave**

top 10 destinations 2020



1. **12 Ave & W 40 St**
2. **West St & Chambers St**
3. **1 Ave & E 68 St**
4. **Broadway & W 60 St**
5. **Pier 40- Hudson River Park**
6. **Christopher St & Greenwich St**
7. **Central Park S & 6 Ave**
8. **S 5 Pl and S 5 St**
9. **West St and Liberty St**
10. **E 13 St & Avenue A**

Summary: key takeaways from our analysis

- We saw longer average trip durations during COVID restriction in 2020.
- Type of accounts users paid for demonstrated a decline in subscriptions and rise in daily use passes.
- Our analysis showed that there was a shift from riding during rush hour and that there was a correlation with recreational bike-riding (biking to a park vs. to an office building)
- Top docking station locations changed away from secondary travel hubs to focus more around parks and points of interest.

Recommendations

Better data collection

Accuracy in the collection of the data is key to help making decisions about the best course of action in terms of planning for the growth of the program. Giving the collection process care and attention will lead to less cleanup.

Increase bike usage at scale

As the company sees growth year after year, the number of stations and bikes will need to support the growth. It will also need to keep pace with the competition with other modes of transportation available in the city.

Financial analysis of users

To anticipate usage in terms of customers, the company will have to create a budget accordingly. More users of both the subscriber and the customer typer will make the company profitable.

Use of predictive models

Predictive models can be guides for planning new bike stations, which is correlated with better data collection.

Thank you!

Learn more about the program at citibikenyc.com