Citi Bike Program
Data Analysis

Marc Pitarys
Data Analytics Student
20 JUL 2018

Presentation Overview

- Background
- Data Analysis
- Results
- Observations

Background

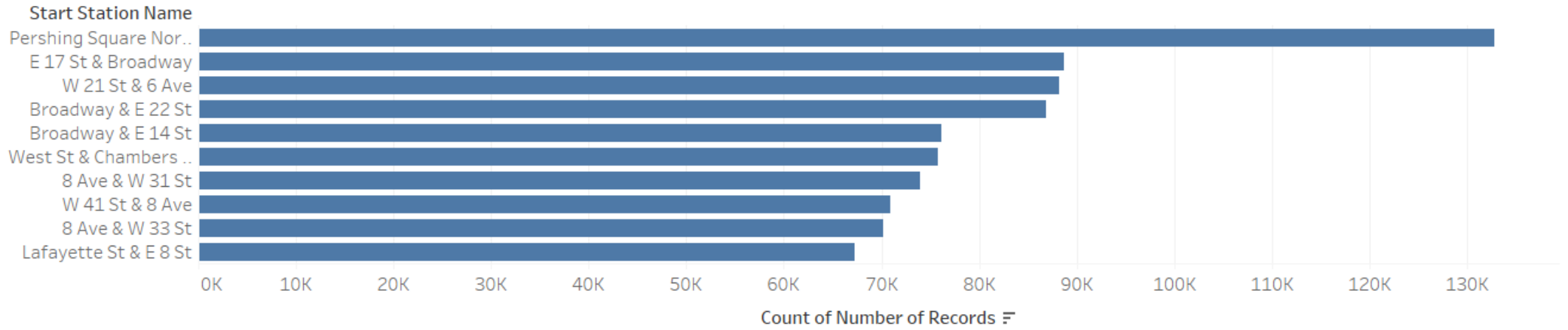
- The purpose of this analysis is to measure the performance of the Citi Bike Program in New York City. The data analysis also identifies performance trends, if any, over the subject time span.
- Bike ridership data and trip histories were obtained from the Citi Bike System Data Web Site. Analysis was conducted on Data collected over the time period of AUG 2017 to MAY 2018.
- A commercial data visualization tool called Tableau was used to analyze approximately 13.4 million records
- Observations were made from the visualizations and a interactive dashboard was deployed to allow further investigation of the ridership data

Data Analysis Overview

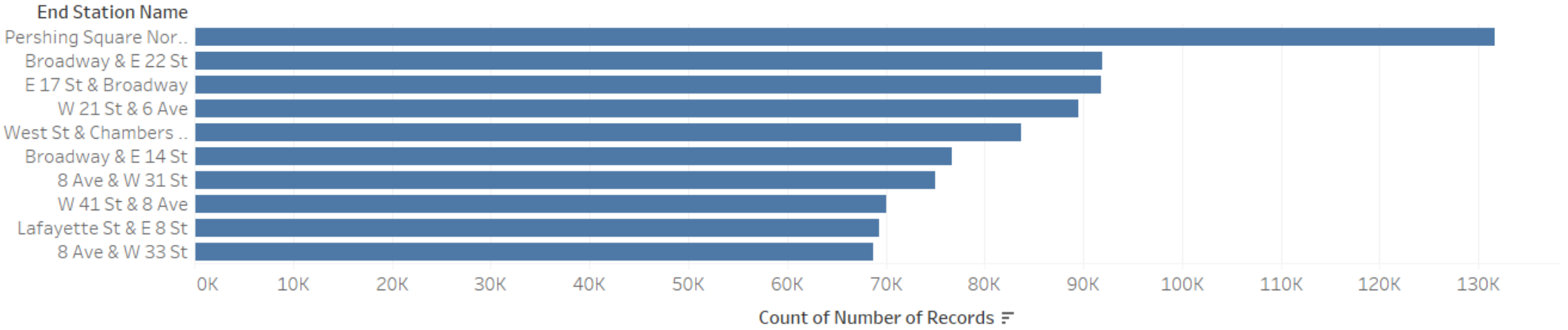
- Approach:
  - Download CSV files from Citi Bike System Data Web Site
  - Extract data from CSV files and load into a Panda data frame then create a single CSV file from the data frame
  - Load the CSV file into Tableau
  - Using Tableau, create visualizations from the data and store on worksheets
  - Generate an interactive dashboard from the Tableau worksheets
  - Make observations and determine any trends from the visualizations
- Data Analysis Software Tools/Resources:
  - Python 3.6
  - Panda scientific computing library
  - Jupyter for Python code creation and execution
  - Tableau Public (Limited to 15 million records)

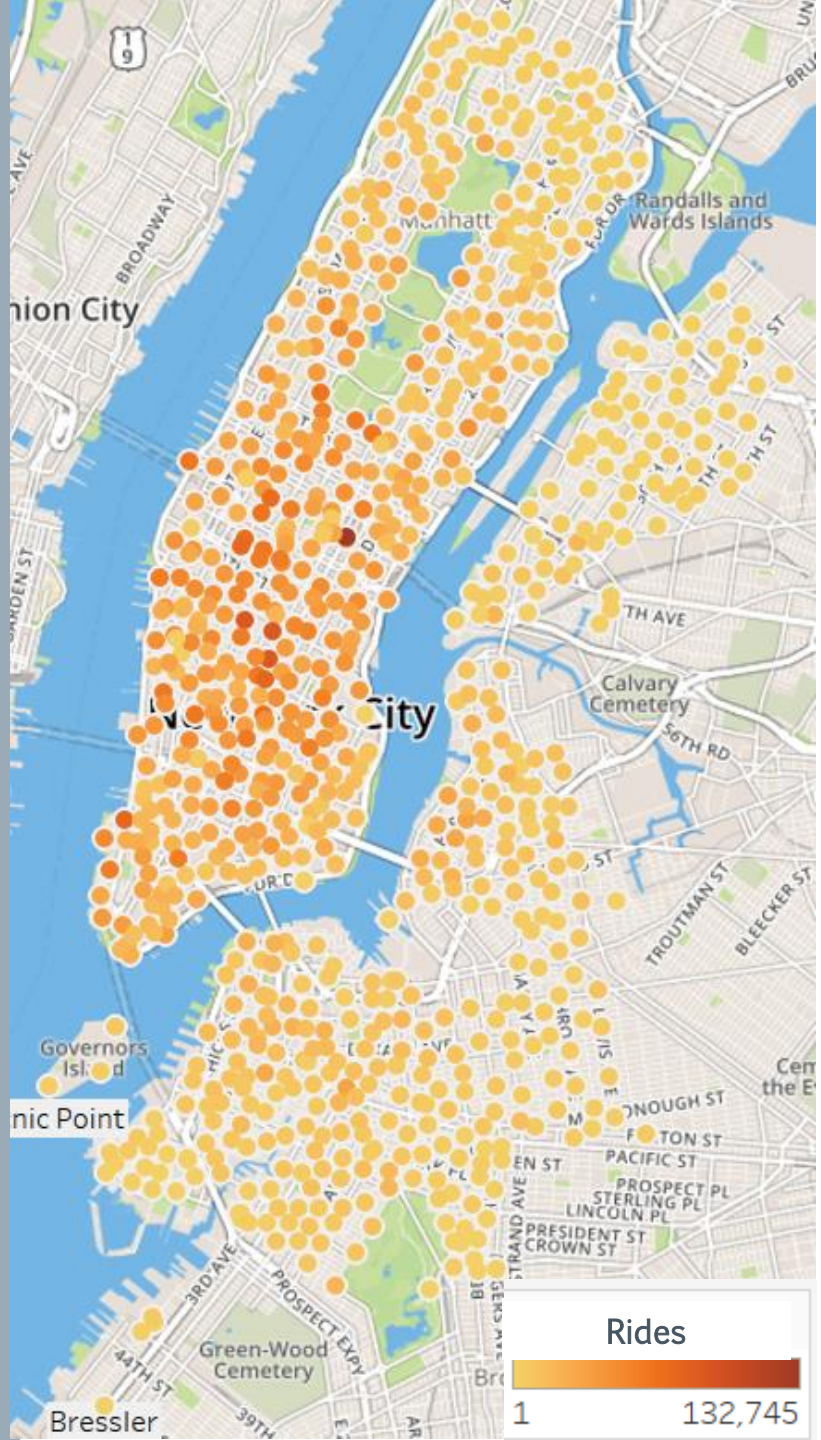
Ridership (AUG 2017 to MAY 2018)
Top Ten Stations

Rides by Start Station

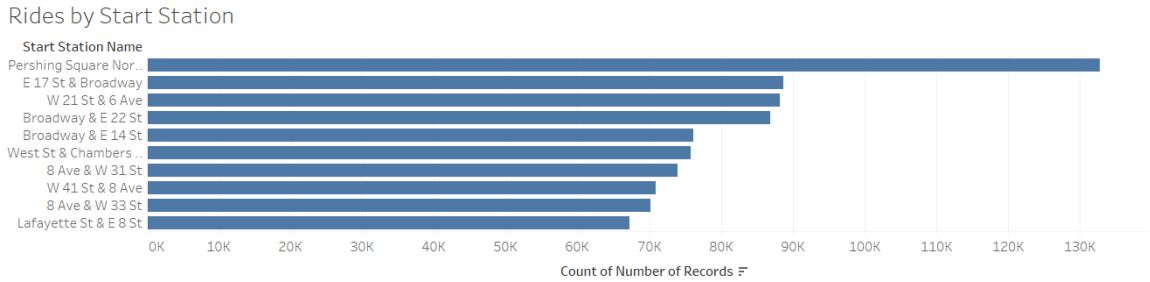


Rides by End Location

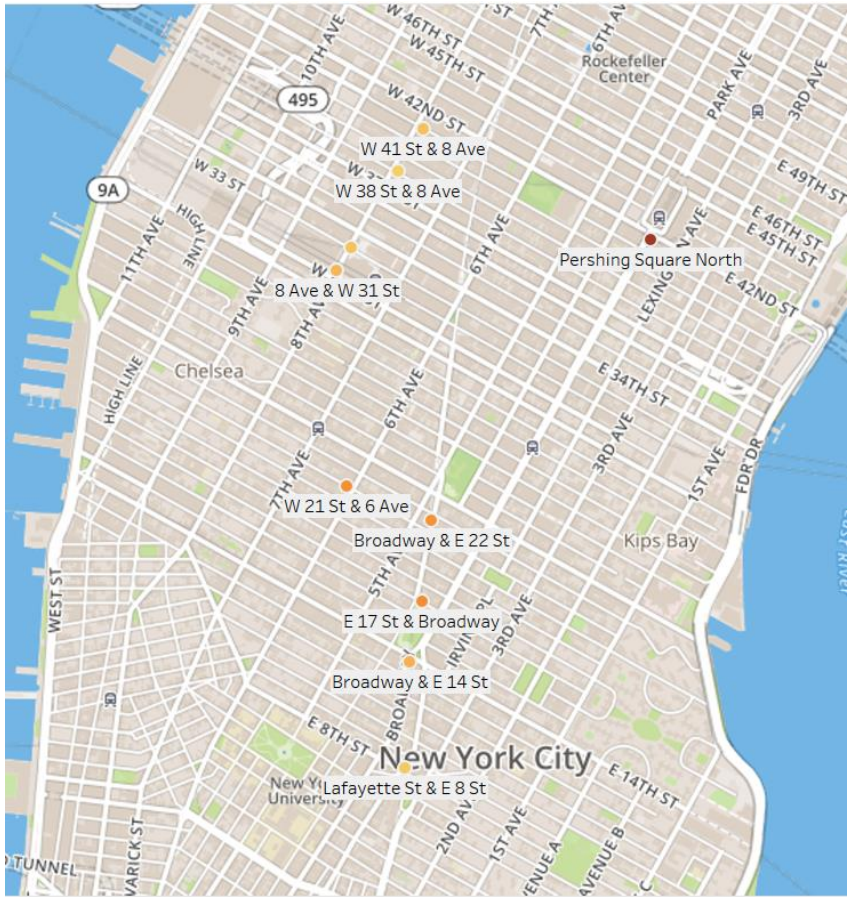


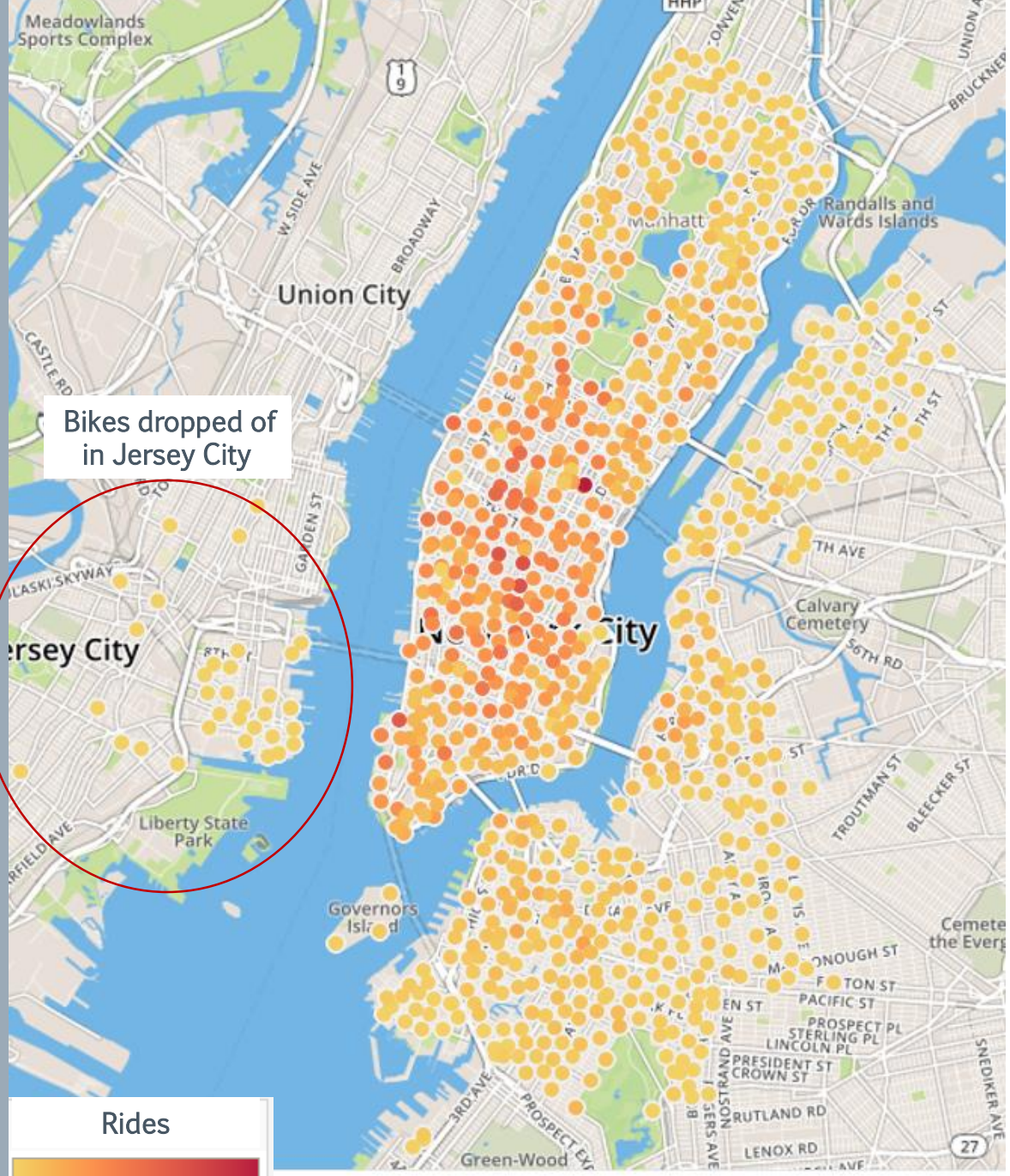


Ridership AUG 2017 to MAY 2018
Start Stations

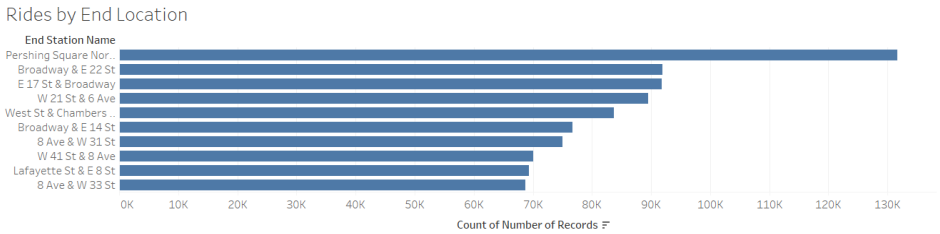


Top Ten Start Stations AUG 2017 to MAY 2018

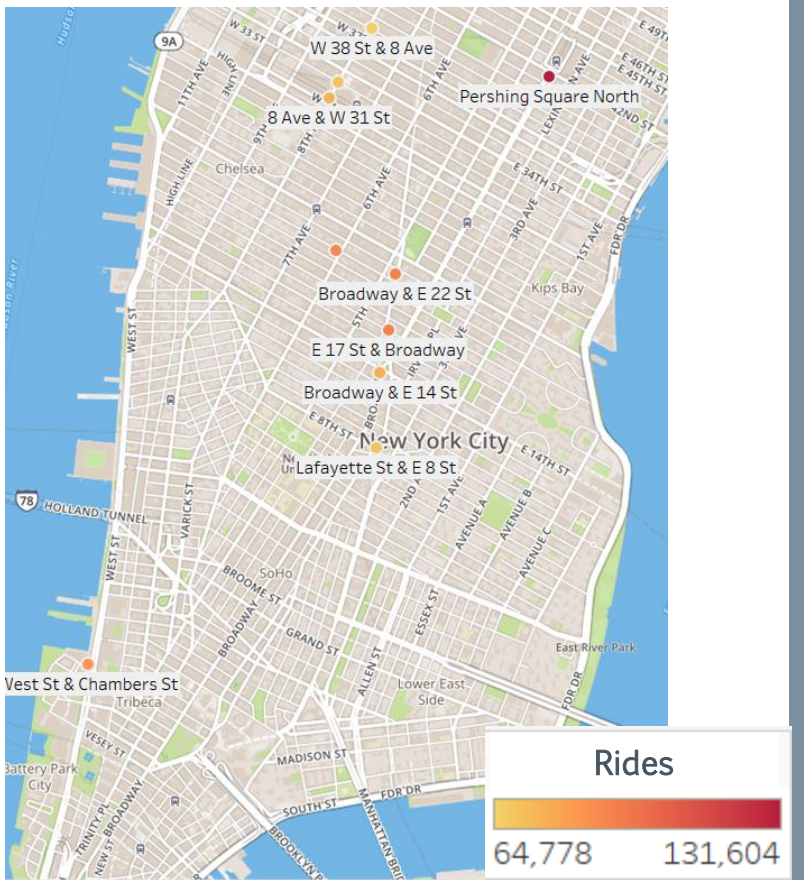




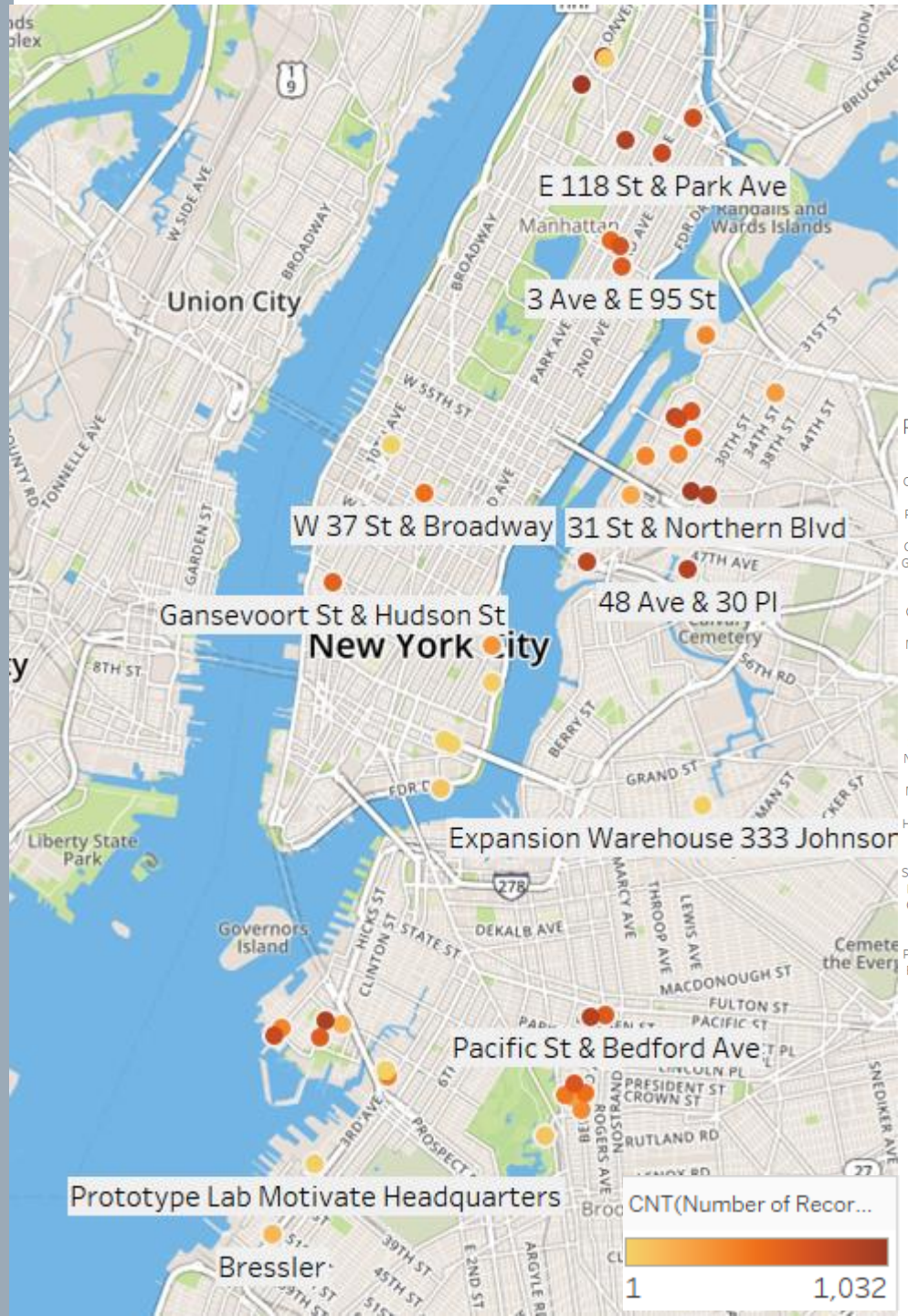
Ridership AUG 2017 to MAY 2018
End Stations

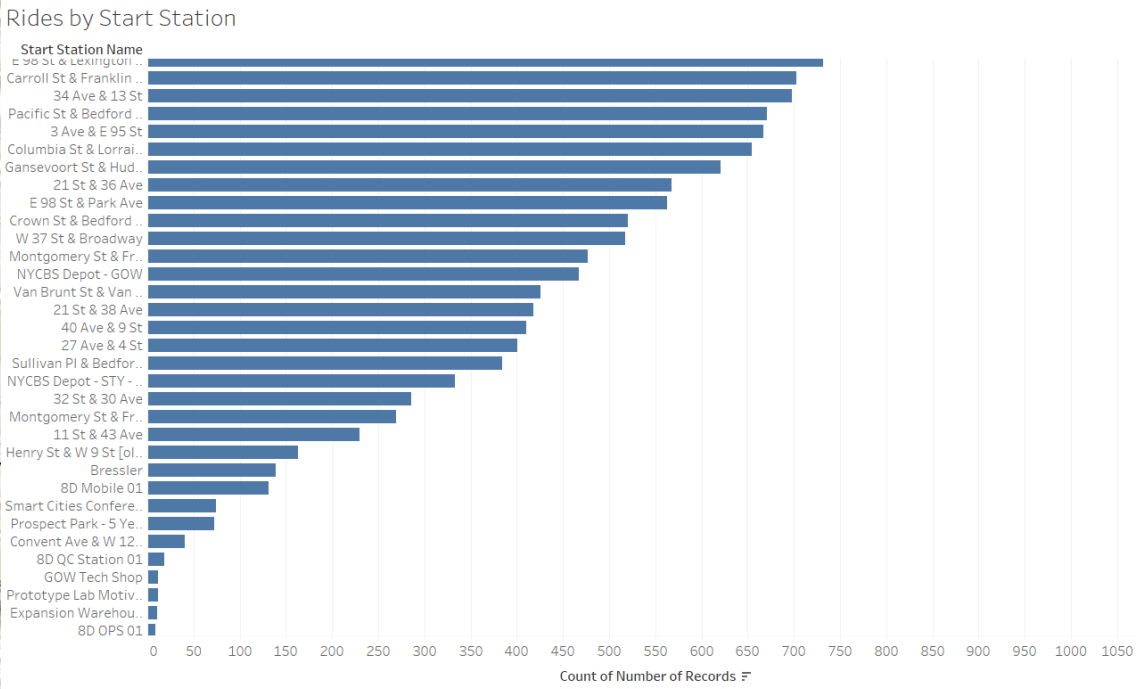


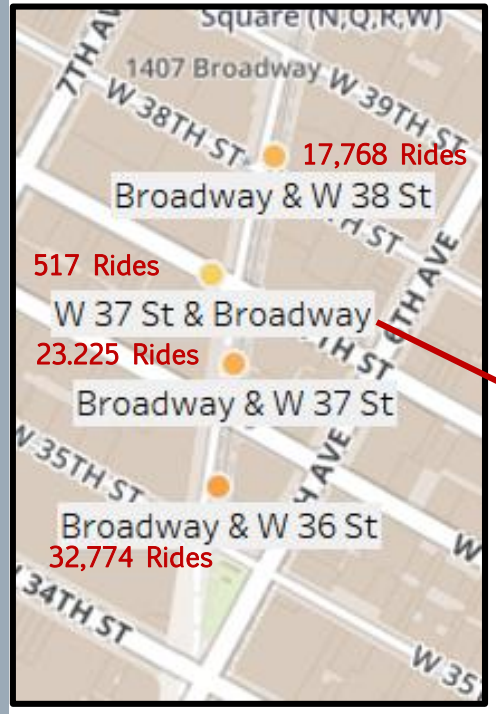
Top Ten End Stations AUG 2017 to MAY 2018

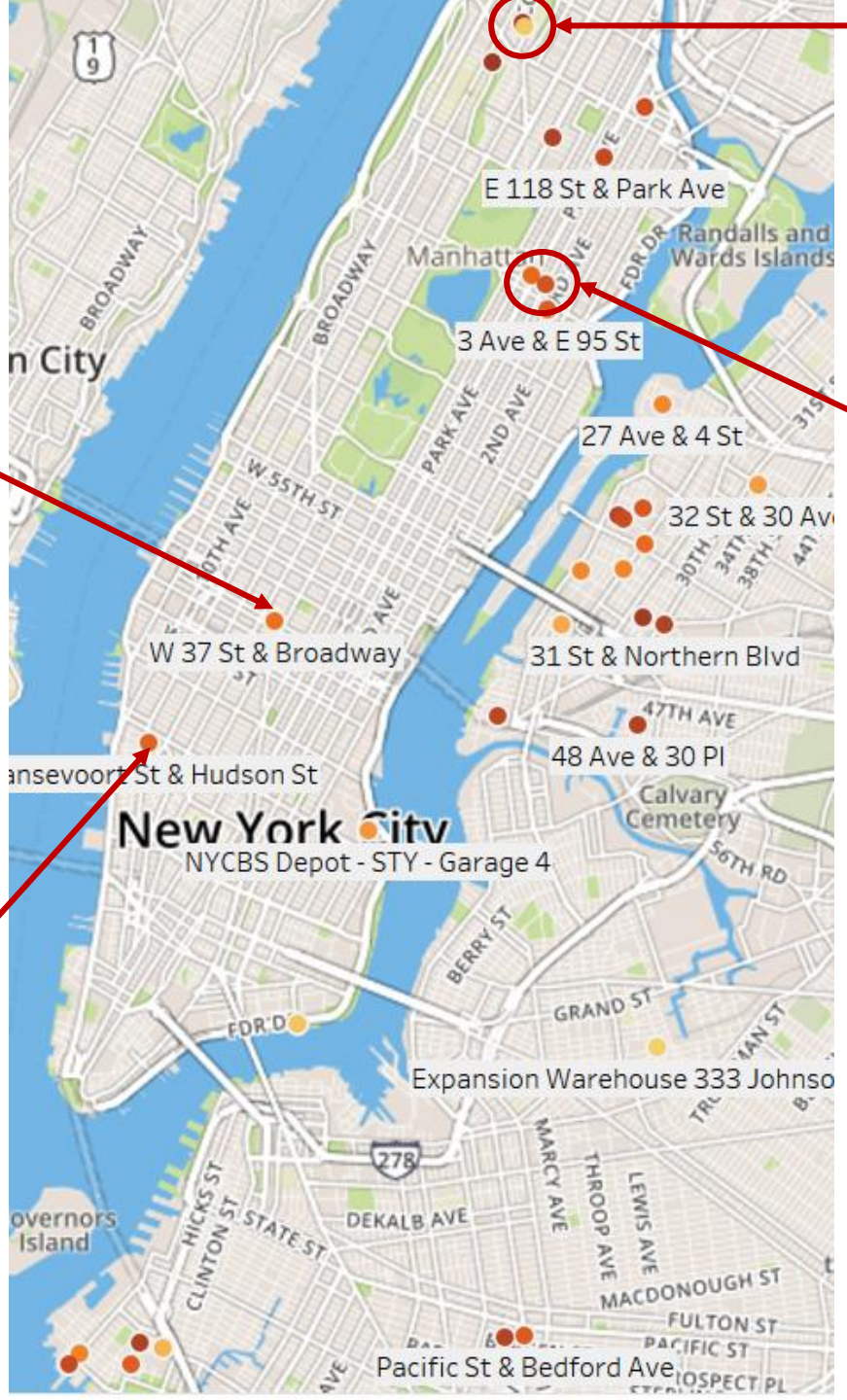


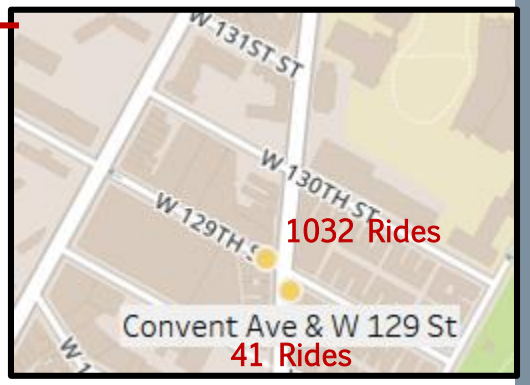
Bottom Performing Stations

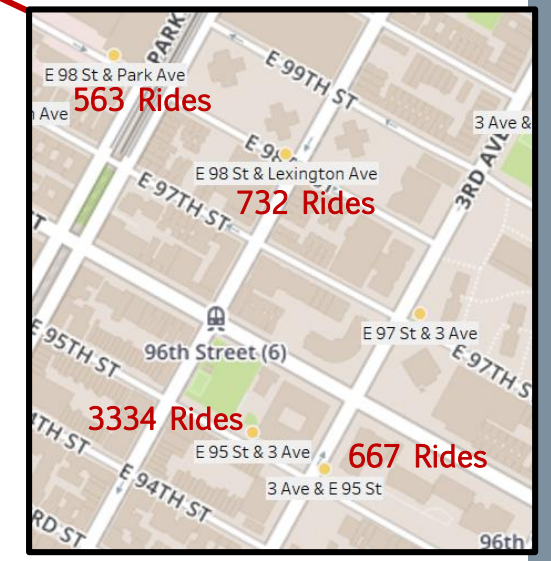


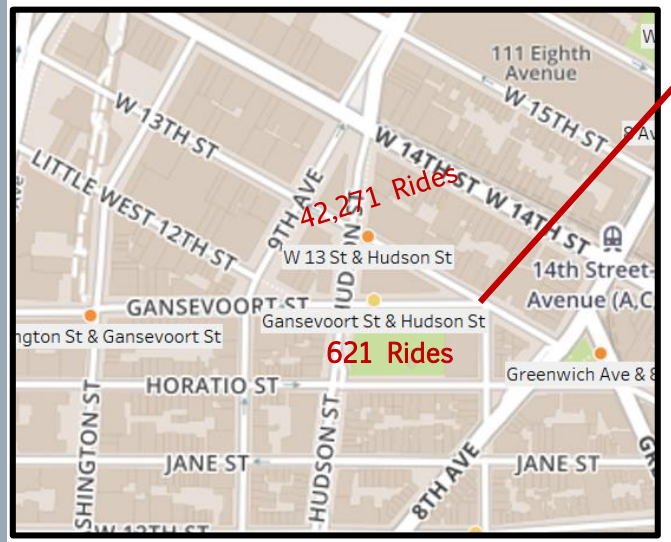






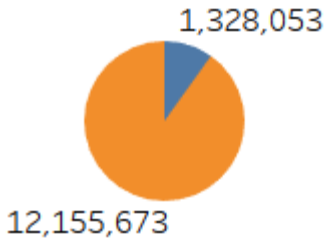






Proximity of Bottom Performing Stations to High Performing Stations

User Type



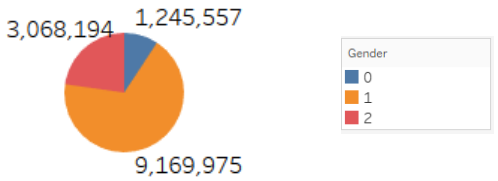
Usertype
- Customer
- Subscriber

Ridership by Gender

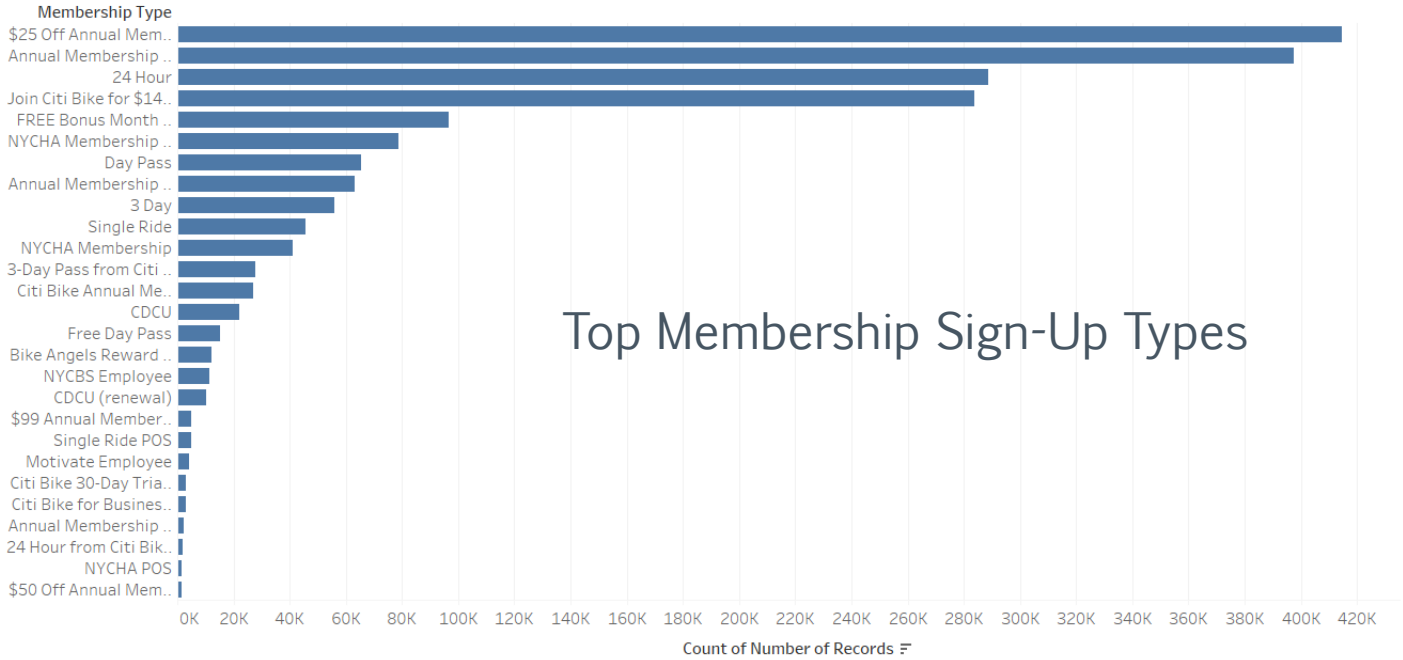
0 = Unknown

1 = Female

2 = Male

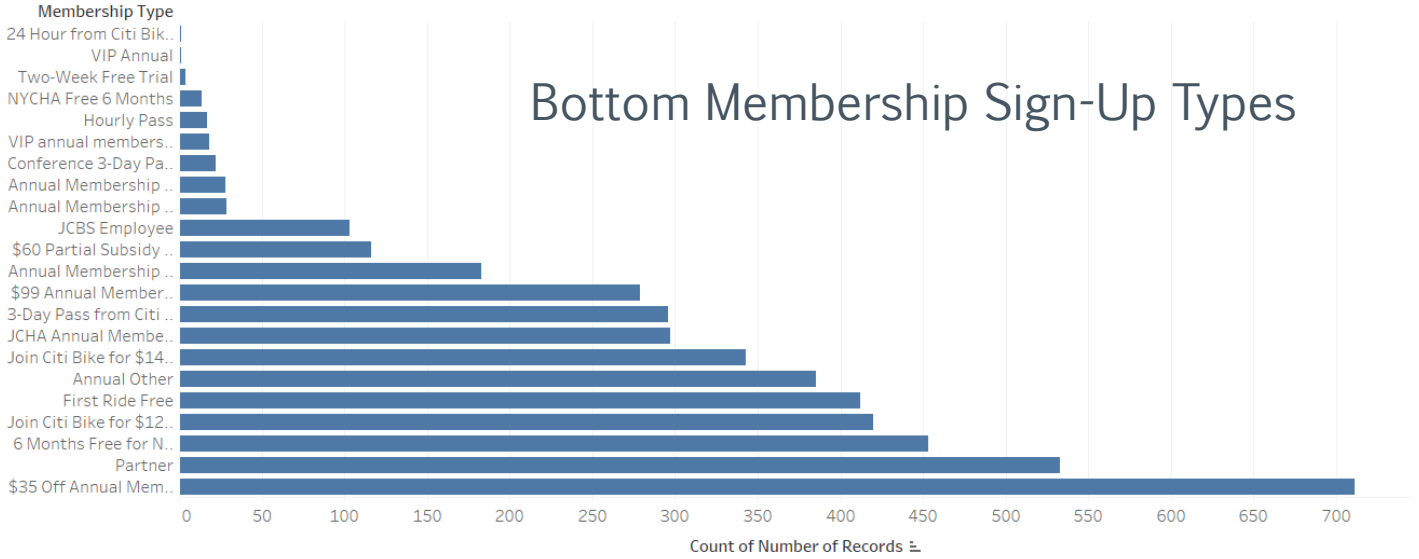


Type of Membership Sign-Up



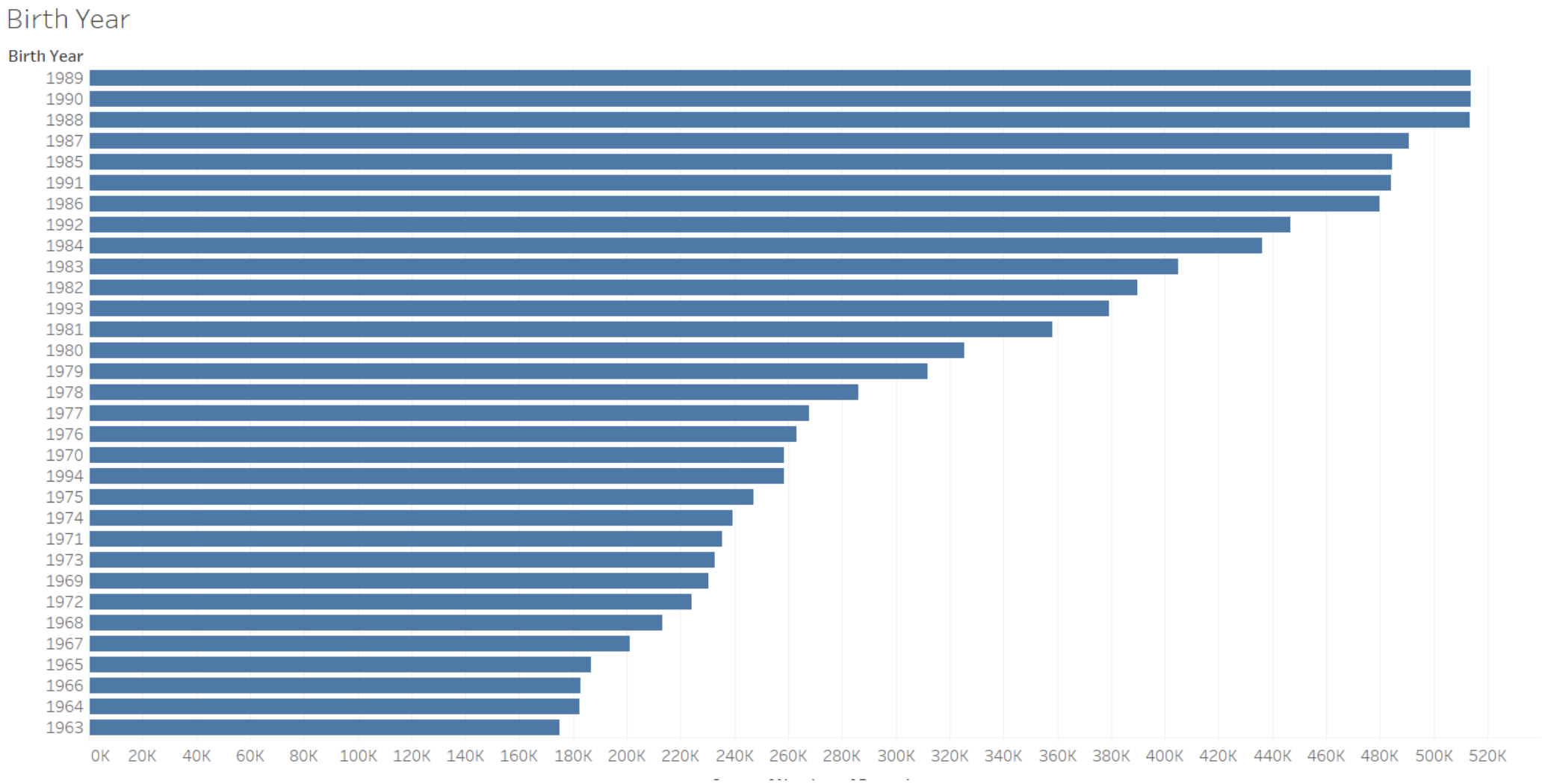
Top Membership Sign-Up Types

Type of Membership Sign-Up

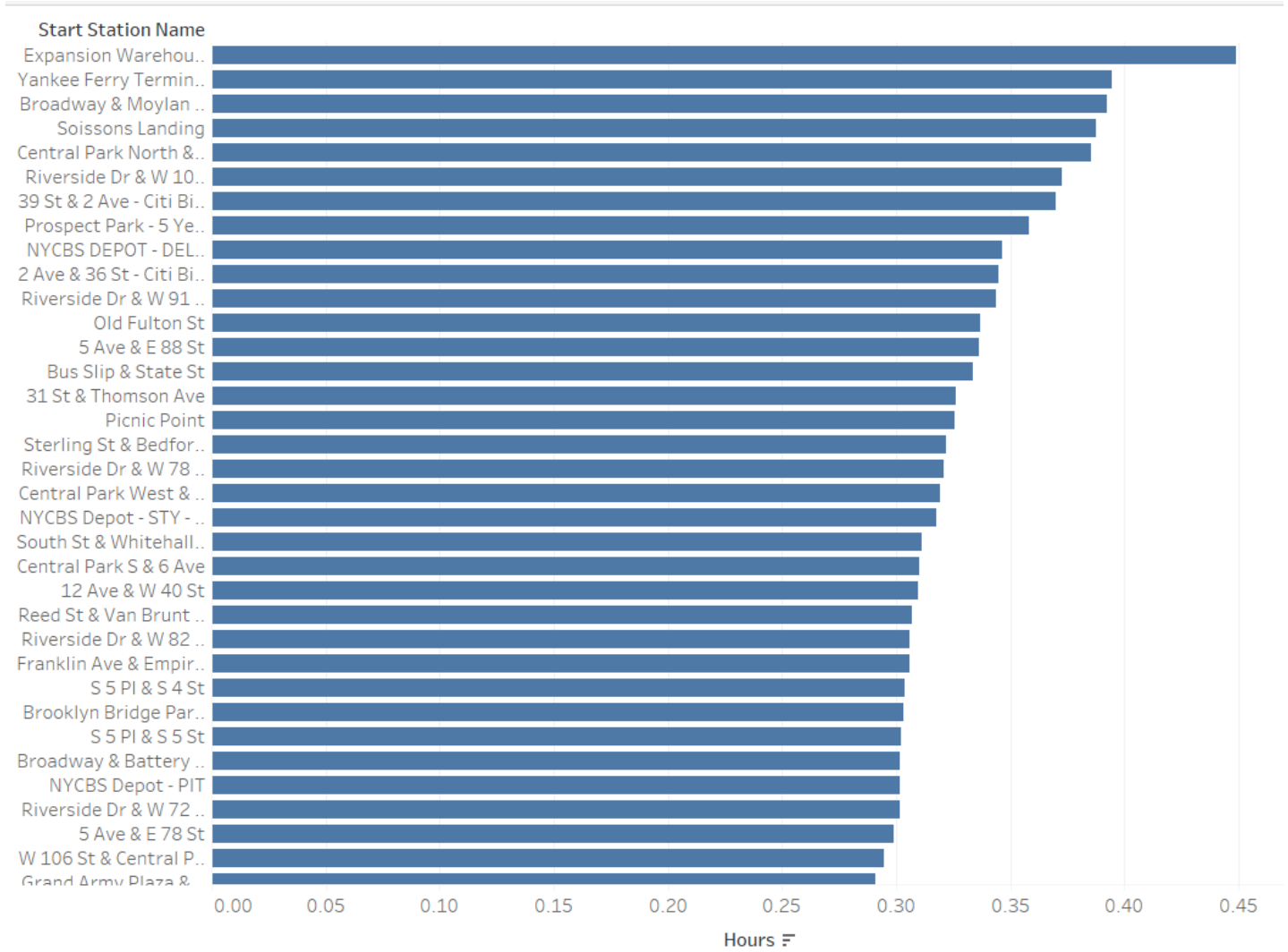


Bottom Membership Sign-Up Types

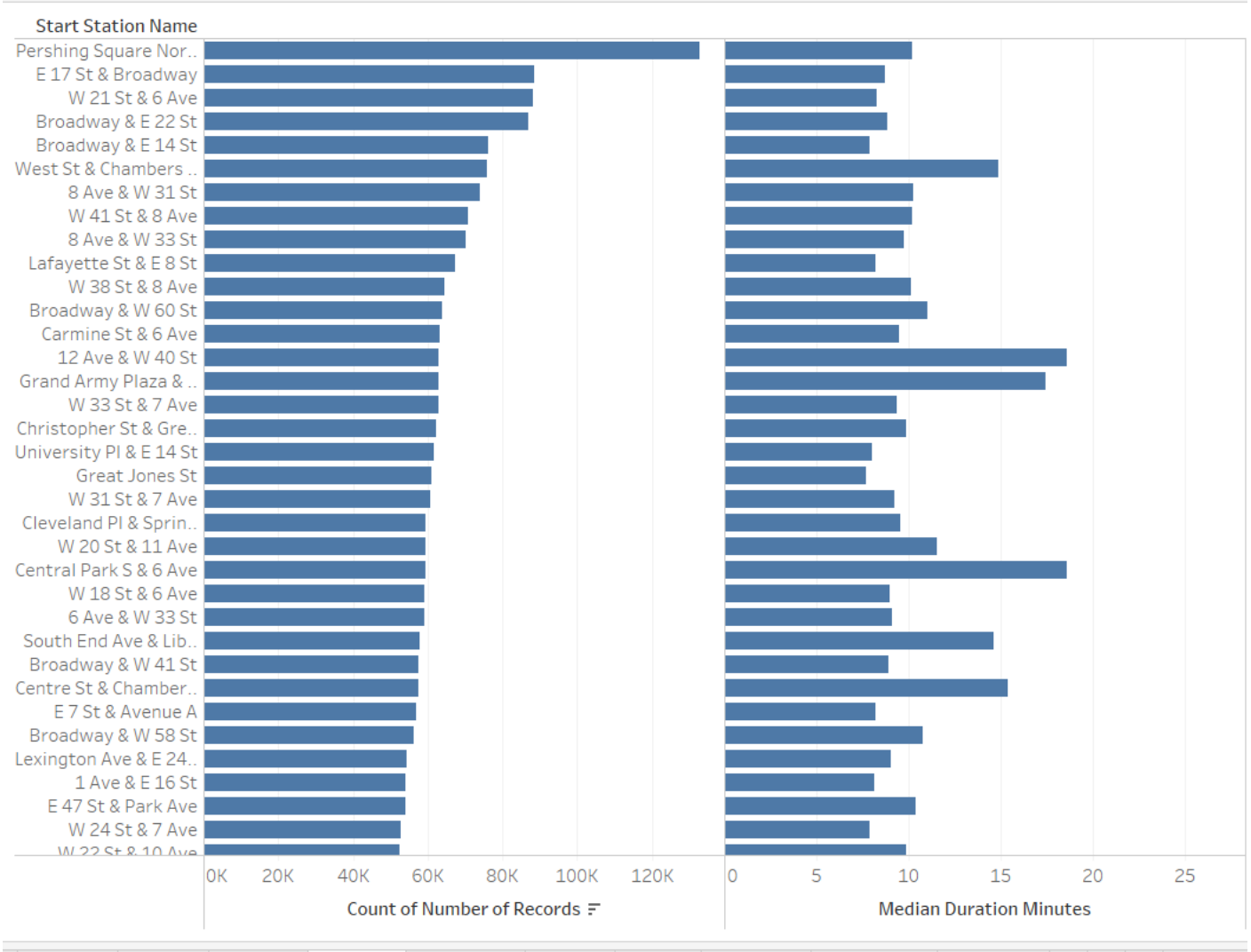
Age of Riders



Stations with the Highest Duration Trips



Trip Duration: Highest Performing Stations



Observations

- Top stations used to start trips are also the top stations for ending trips. Median time of trips less than 10 mins.
- Men outnumber women riders by approximately 3 to 1
- Most riders subscribe to the service (12 to 1)
- The ages of most riders are late twenties and early thirties
- The $25 discount for the annual membership is most popular among riders
- Proximity between stations may be contributing to low utilization of some of the worst performing stations
- Utilization decreases as the distance from the center of the city increases.