

WomenTechWomenYes - MTA Turnstile Traffic Analysis

By: Jesse Reonardy

Abstract

The goal of this project is to determine when and where to strategically locate street teams at the entrances of New York's subway station to collect the most email addresses for WomenTechWomenYes summer annual gala. I work with MTA turnstile data and big tech companies location to accomplish an accurate result. After processing the data, I created bar charts to communicate my results.

Design

The client for this is WomenTechWomenYes (WTWY). The data is obtained from the MTA turnstile database and big tech footprint chart from Yahoo! Finance article. I use the pandas module for data manipulation and analysis.

Data

- MTA turnstile data for 12 weeks. This data duration is chosen to have a consistent result among the quarters in the year.
- May 20, 2021 through June 11, 2021 date is chosen for the 12 weeks data because this is the quarter that is leading to the annual gala that will happen at the beginning of the summer each year.
- Since most new yorkers exit the subway through the exit door instead of the turnstile, I decided that the exit data is inaccurate and I solely rely on the entries data for this analysis.
- Big tech footprint in NYC article from Yahoo! Finance. See Figure 2.

Algorithms

First, based on the MTA database, analysis is performed to determine the top 10 stations by total traffic volume only. See figure 1. Then, out of these top 10 stations, I truncate the list to top 6 stations based on stations that are near big tech companies. See figure 2 and 3. This is done so that we can focus the effort on someone who has more tendencies to attend the gala and contribute to the cause. Furthermore, the data is analyzed per daily traffic to determine which days are the best for our goal. See Figure 4. Finally, the data is further analyzed per hourly traffic to determine which hours are the best for our goal. See Figure 5 and 6.

In conclusion, the top 6 stations to focus our efforts are 34th Street-Penn Station, 34th Street-Herald Square, Grand Central-42nd Street, 14th Street-Union Square, 23rd Street Station, 42nd Street-Port Authority Terminal. Also, the best time to focus our efforts are on weekdays from 4:00pm through 9:00pm for all stations

Tools

- Numpy and Pandas for data manipulation
- Matplotlib for plotting

Communications

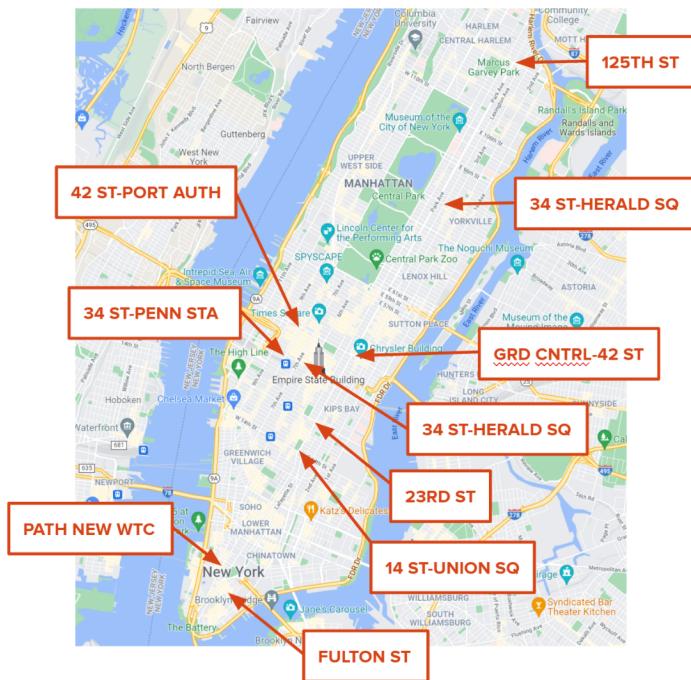


Figure 1 - Top 10 stations by traffic volume only



Figure 2 - Big tech footprint in NYC. Source: Why America's tech giants are flocking to one part of NYC article by Sarah Paynter and Krystal Hu Yahoo! Finance - January 5, 2020.



Figure 3 - Top 6 stations by traffic volume and Big Tech Companies location



Figure 4 - Average daily traffic in top 6 stations across 12 weeks



Figure 5 - Average hourly traffic in top 6 stations based on weekdays traffic (Monday-Friday) of a random week

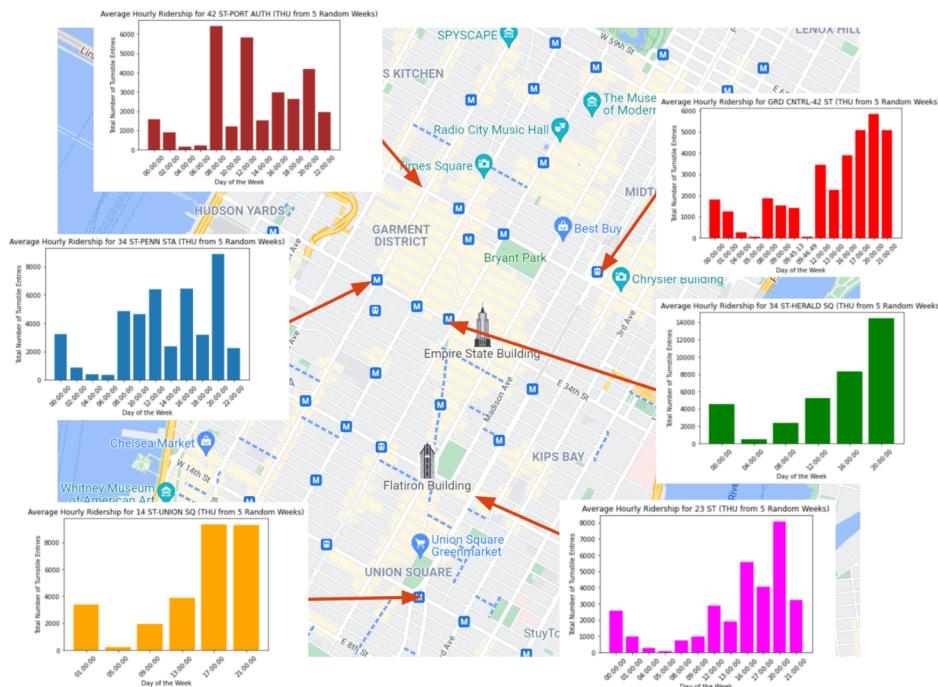


Figure 6 - Average hourly traffic in top 6 stations based on Thursday traffic from 5 random weeks