Data Analysis of MTA Turnstile Traffic

ELLIOT, WEI, LIAM

Introduction

Project background:

- WomenTechWomenYes (WTWY) → to increase the participation of women in technology
- NYC annual gala in summer to build awareness and reach.
- Place street teams at entrances to subway stations

Goals:

- Determine the optimal stations and time to place street teams
- Provide recommendations to WTWY
  - where and when to deploy their teams

Methods

Data wrangling:

- **Data gathering:**
  - MTA turnstile data (11/09/2019 ~12/28/2019)
  - US CENSUS data (2018)
  - Google geocode API
- **Data cleaning:**
  - Broke down the data into AM/PM
  - Calculated daily entries and exits as well as total daily traffic
  - Added zip code and corresponding adjusted gross income for each station
  - Added day name

Methods

Analysis :

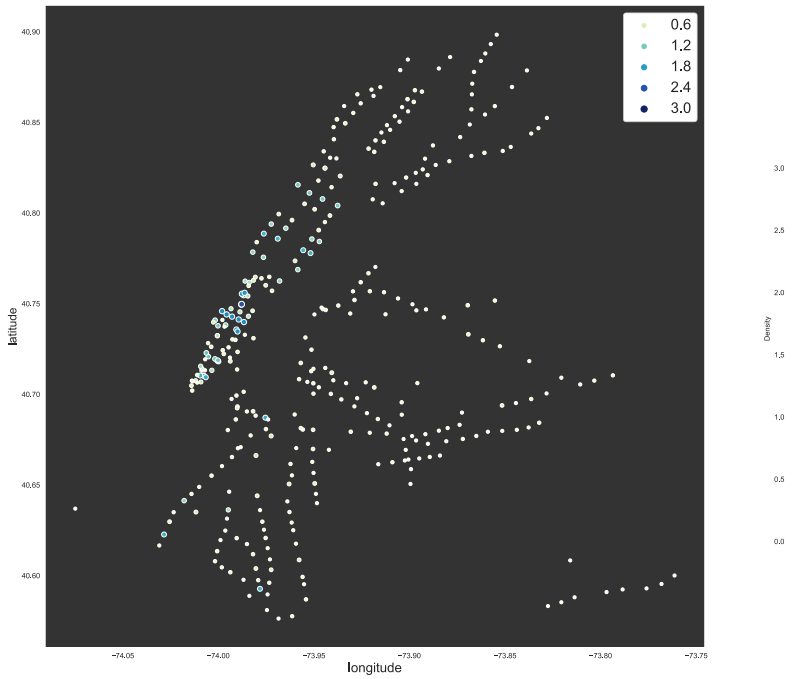
- **Software toolkits:**
  - pandas; numpy; matplotlib; seaborn
  - geopy; geopandas; json; requests module
- **Data analysis:**
  - Top 10 stations based on total daily traffic
    - Weekly traffic distribution
      - AM (00:00 AM~11:59AM)/PM (12:00 PM~11:59 PM)
      - weekdays and weekends
  - High-income areas
    - To prioritize certain stations by fundraising potential

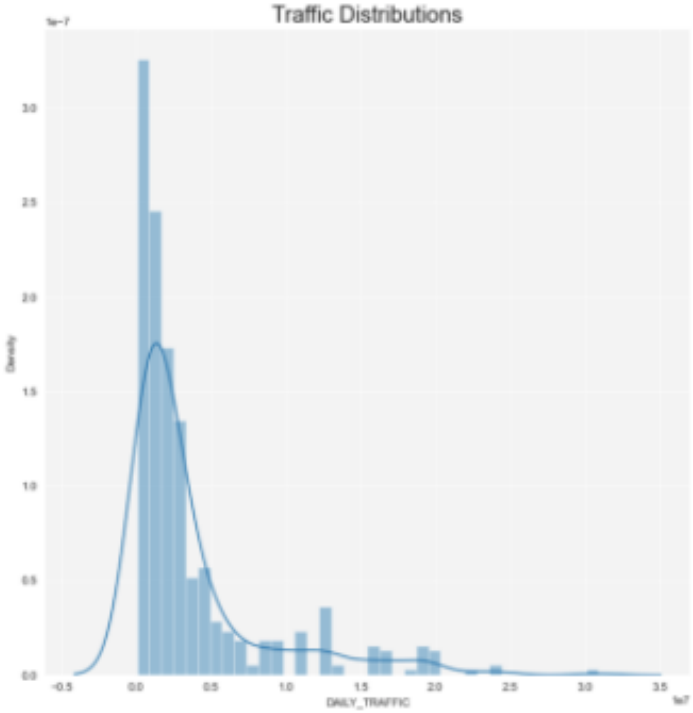
1st Idea: Go by where the people are

FIND OUT THE MOST CROWDED STATIONS, THE MOST TRAFFICKED, AND GO THERE.

Qualitative Picture

Manhattan by its Subways

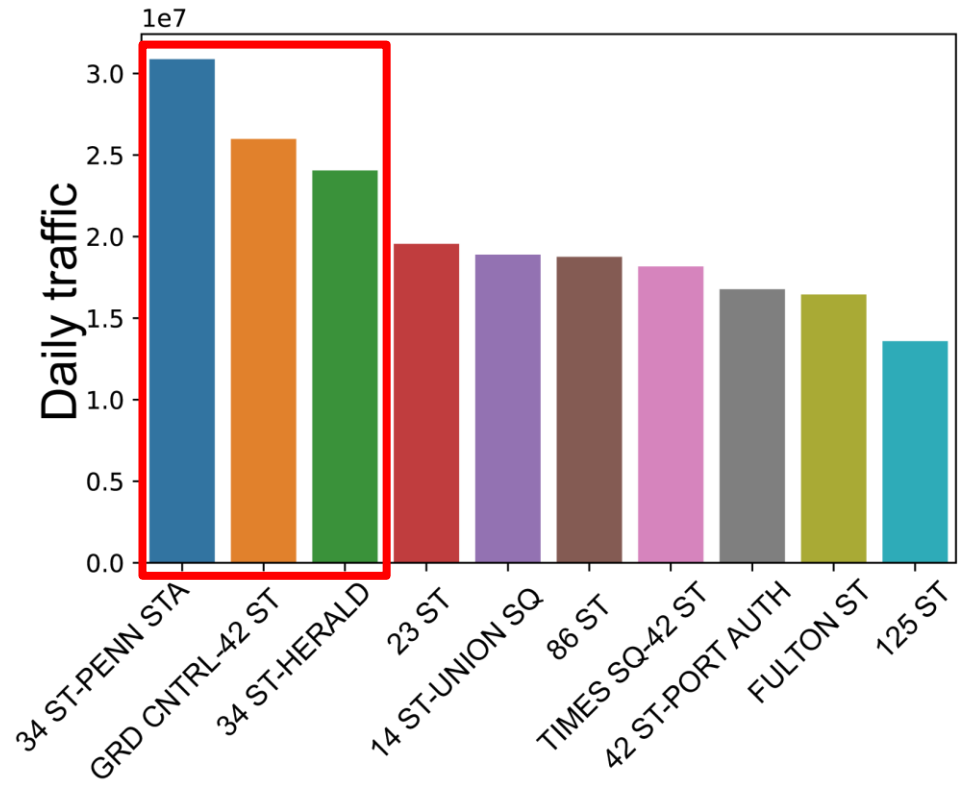




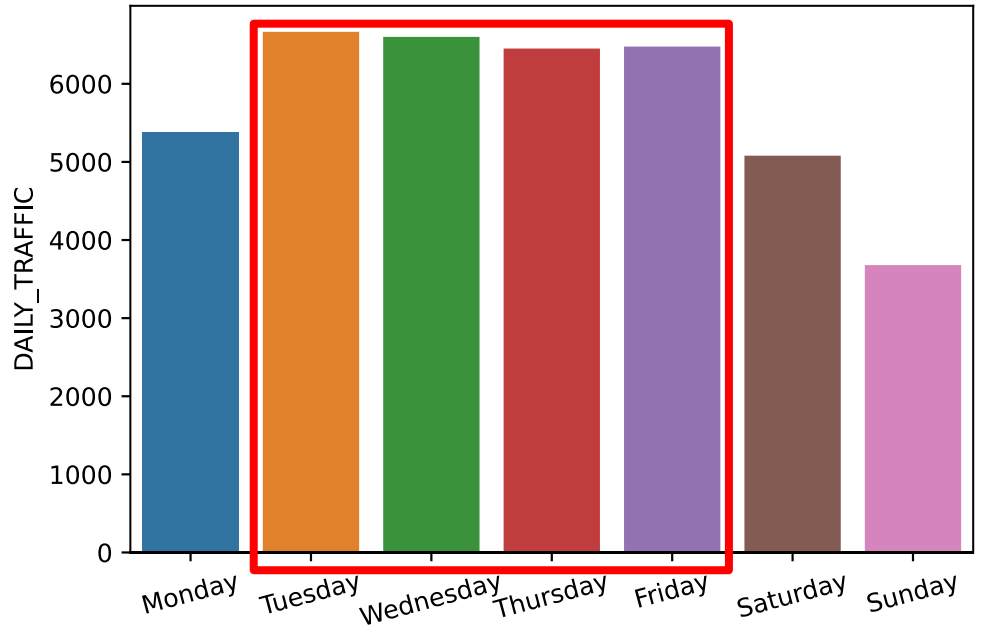
The top ten are statistically significant!

Quantitative List of Stations

Where

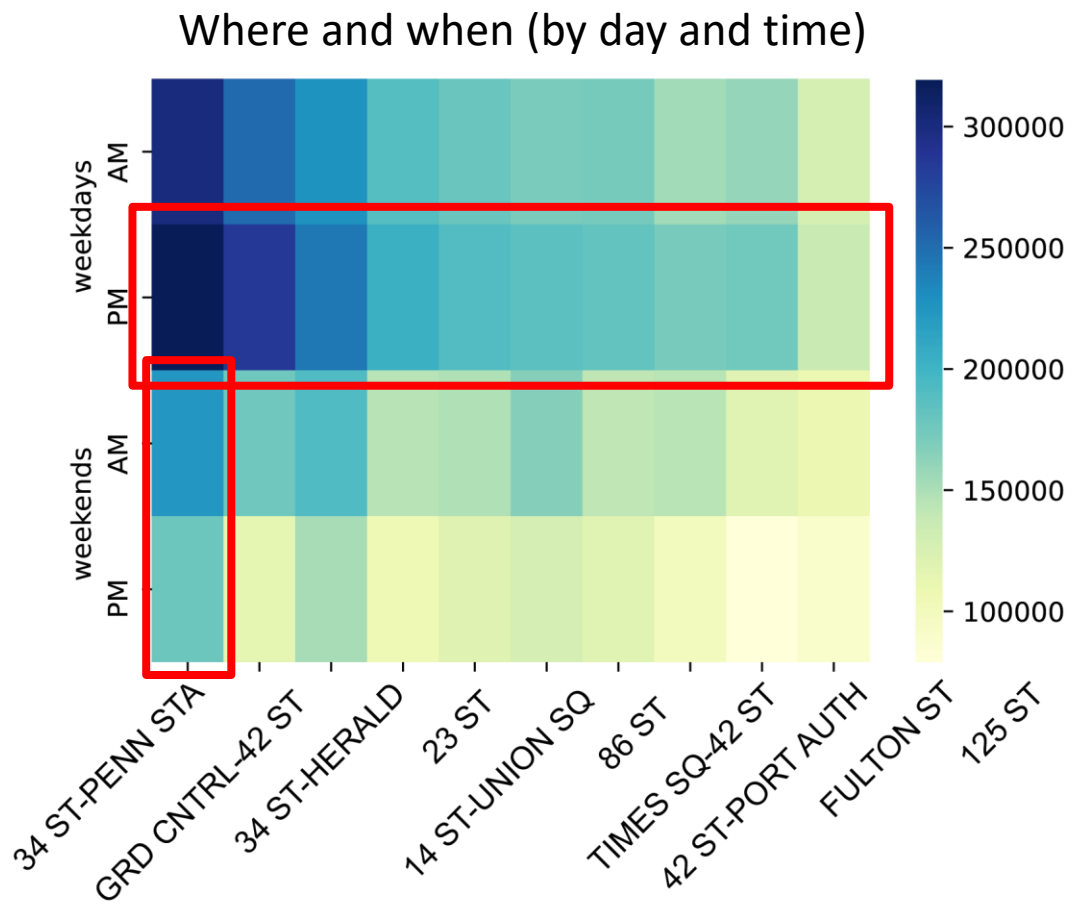


When (by day)



On what day, when

PM busier



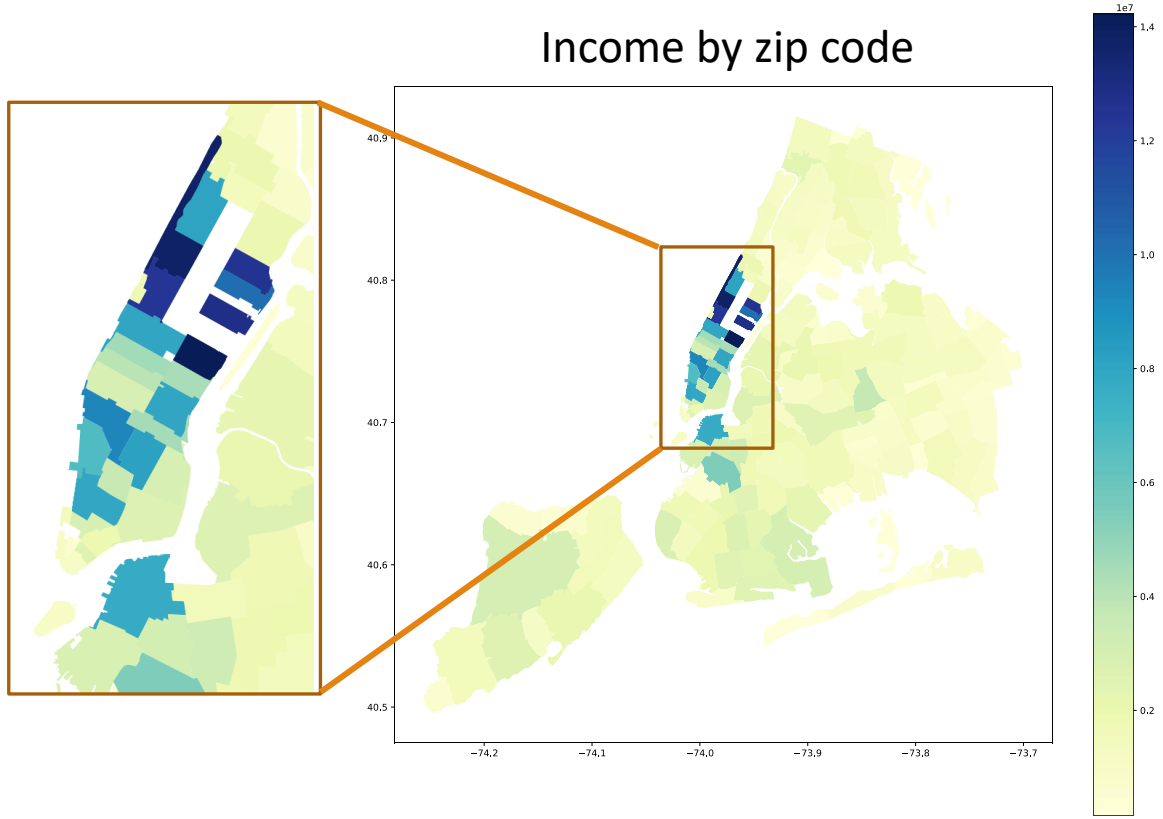
AM Shifts: 00:00 AM~11:59AM
PM Shifts: 12:00 PM~11:59 PM

2nd Idea: Go where the money is

FIND OUT THE HIGHEST INCOME BY ZIPCODE, AND GO TO THOSE STATIONS.

Qualitative Picture

Upper east side!



Conclusions

Place street teams at the top 10 stations

- Weekdays
  - Tuesday ~Friday
  - Afternoon time
    - More employees at ***top 3*** stations, less at others
- Weekends
  - Morning time at 34 ST-PENN STA

Future work

Find a way to efficiently obtain more MTA data to enhance our analysis.

- Identify monthly trends (holiday seasons) and year-over-year variances (big events)
- Estimate hourly entries (convert from 4 hour increments to 1 hr)
  - Hire hourly employees to ***save money***
- Work with WTWY to establish parameters for a schedule to be created
  - Depends on how many people there are, how long are the shifts
- Determine feasible locations for a booth based on street view
  - Avoid wasting resource