Project Benson

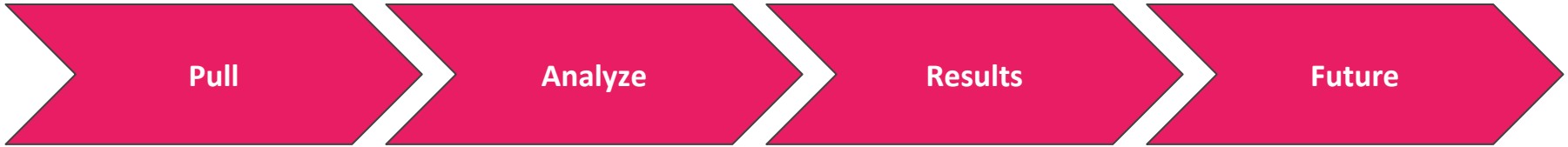
By: An, Evgeniy, Priyanka, Steven

Purpose

To recommend the top MTA stations best suited for soliciting and creating awareness about WTWY Gala by analyzing MTA turnstile data.

Approach

Investigate 52-weeks of MTA subway data and incorporate census information to consult on placement of street teams at various subway entrances with the goal of reaching out to target audience and maximizing the number of participants at WTWY's gala.



Data sources:
- MTA data
- Geocoding
- Census data

Capture MTA Activity based on:
- Entry Frequency
- Geography
- Income

Recommendations based on analysis

Phase 2 of MTA data analysis

Data Sources, Tools

Data Sources

MTA

Google Maps Platform

United States™ Census Bureau

Tools

python

jupyter

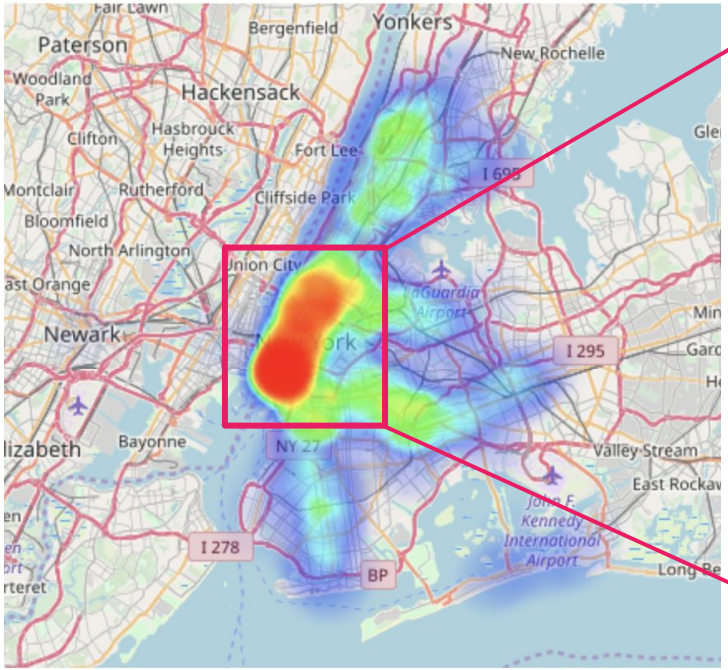
pandas

$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$

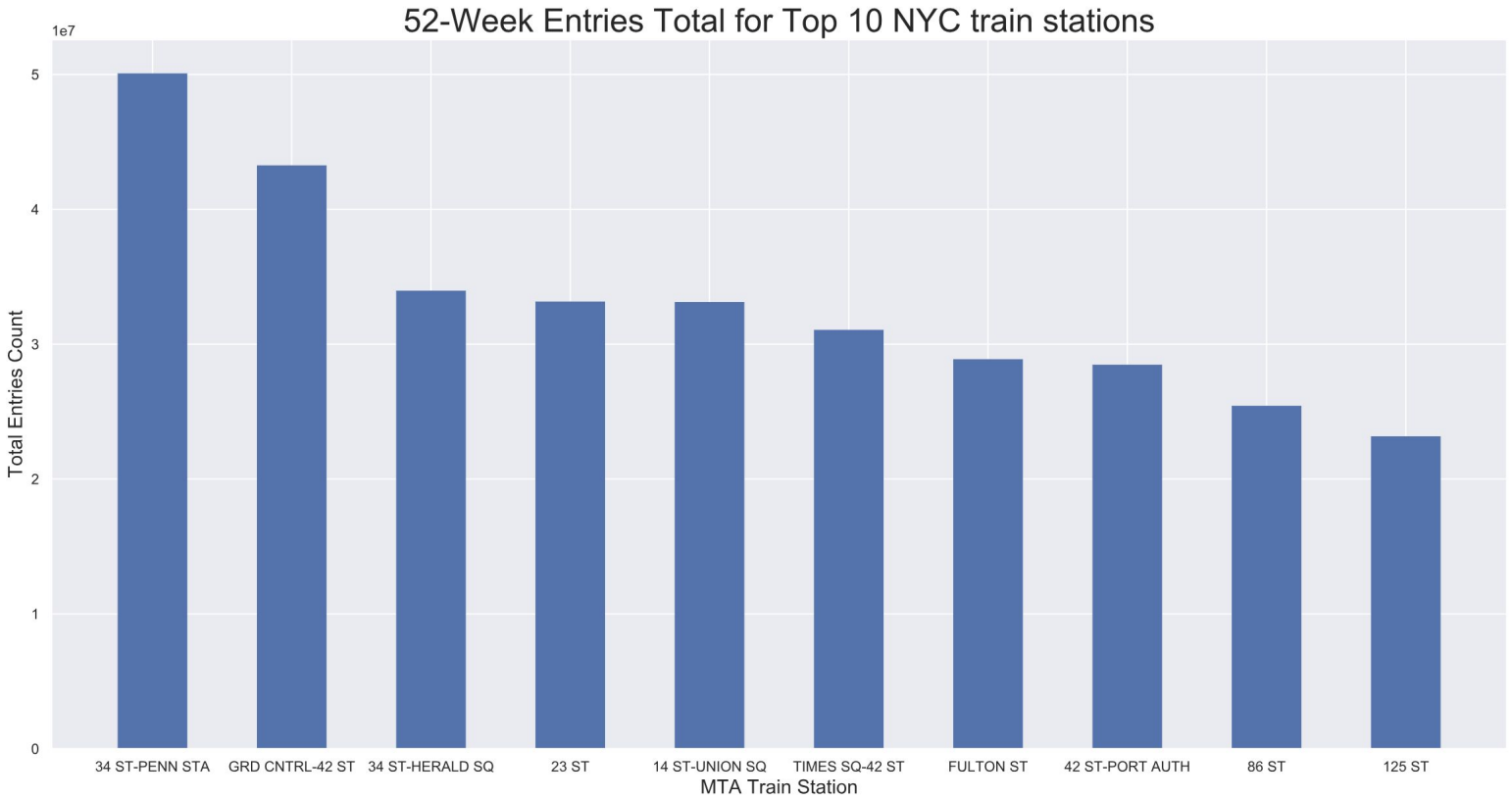
matplotlib

Seaborn

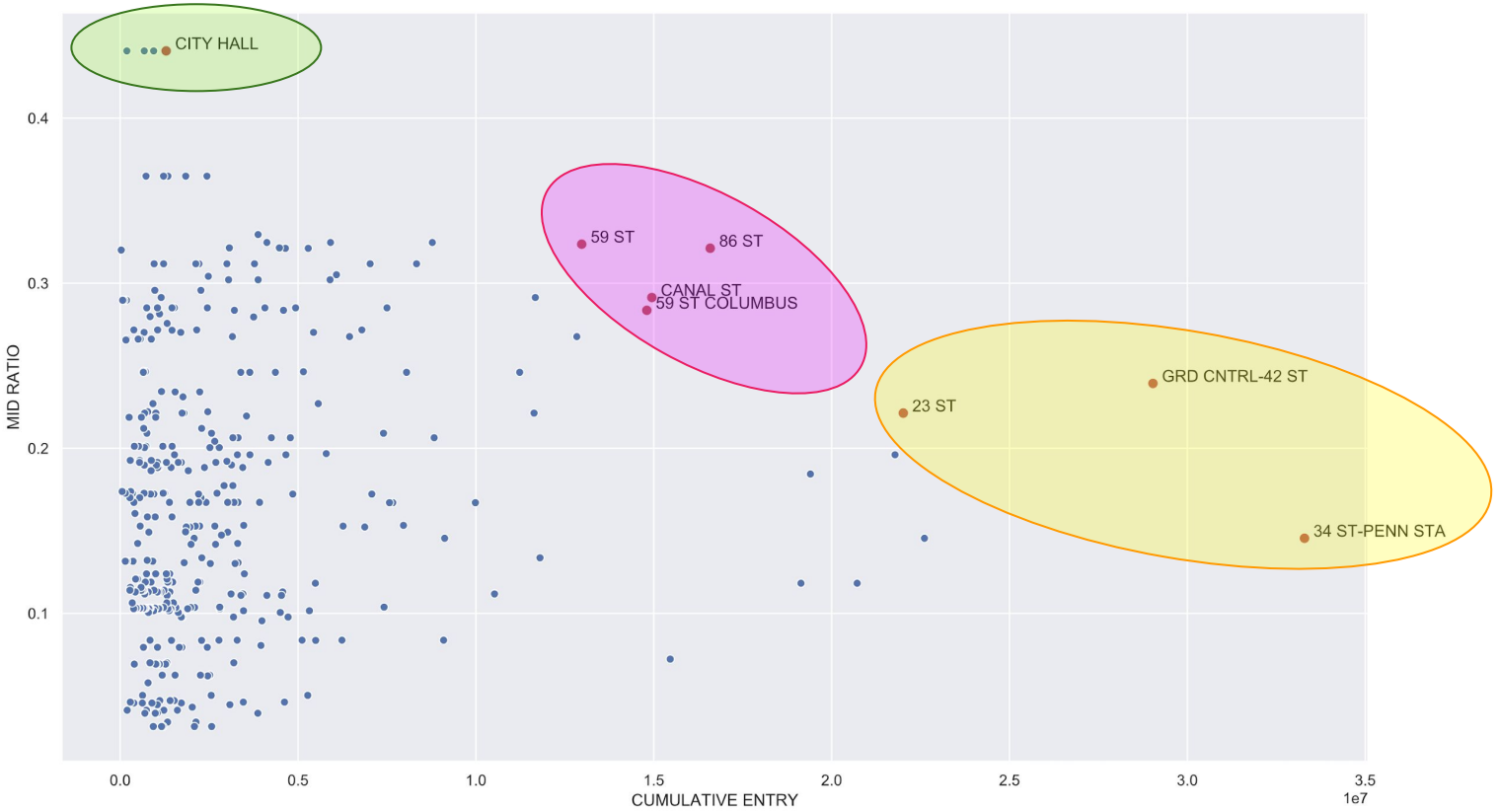
Entries Heatmap



Top 10 Stations



Middle Income Ratio vs. Entry

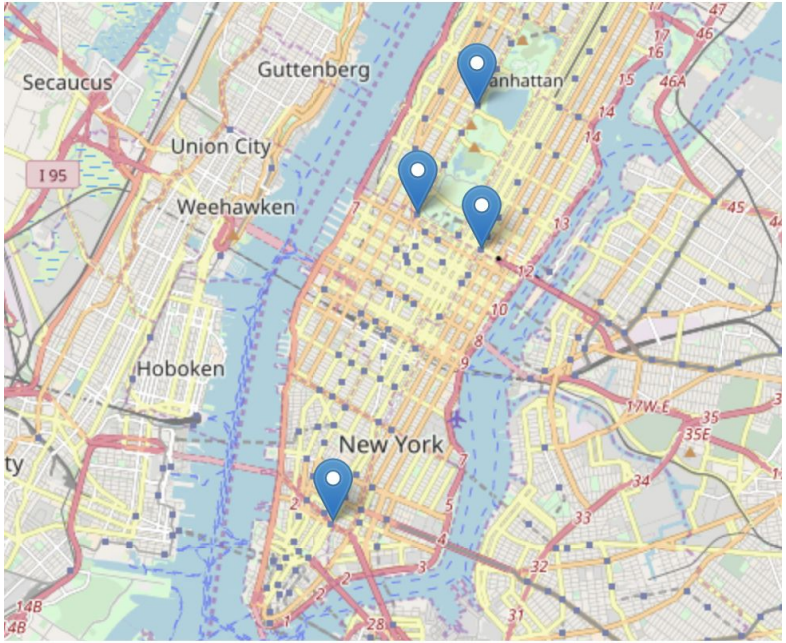


*Mid ratio is the proportion of $100k-200k households for the area

Recommendation

Per results, we recommend street team placements at:

- **86th Street**
- **59th Street**
- **Canal Street**
- **59th Street - Columbus Circle**
- 23rd Street
- Grand Central - 42nd Street
- 34th Street - Penn Station
- City Hall



Next Steps

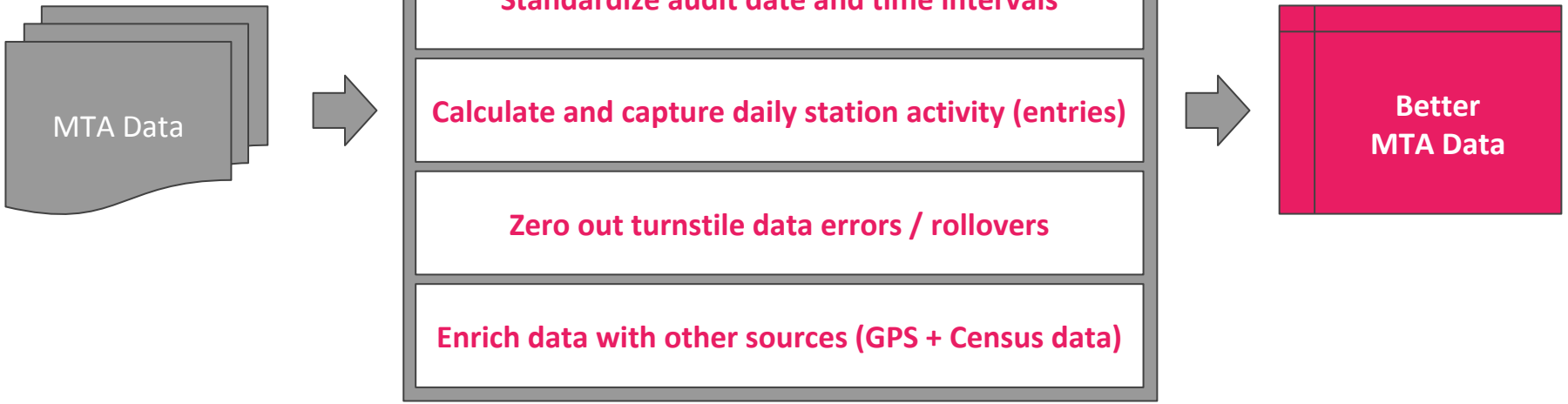
- Optimal street team sizes and placements relative to resources
- Total station exits
- Identify Tech-centric locations
- Demographic statistics (gender)
- Longitudinal MTA activity

Questions ???

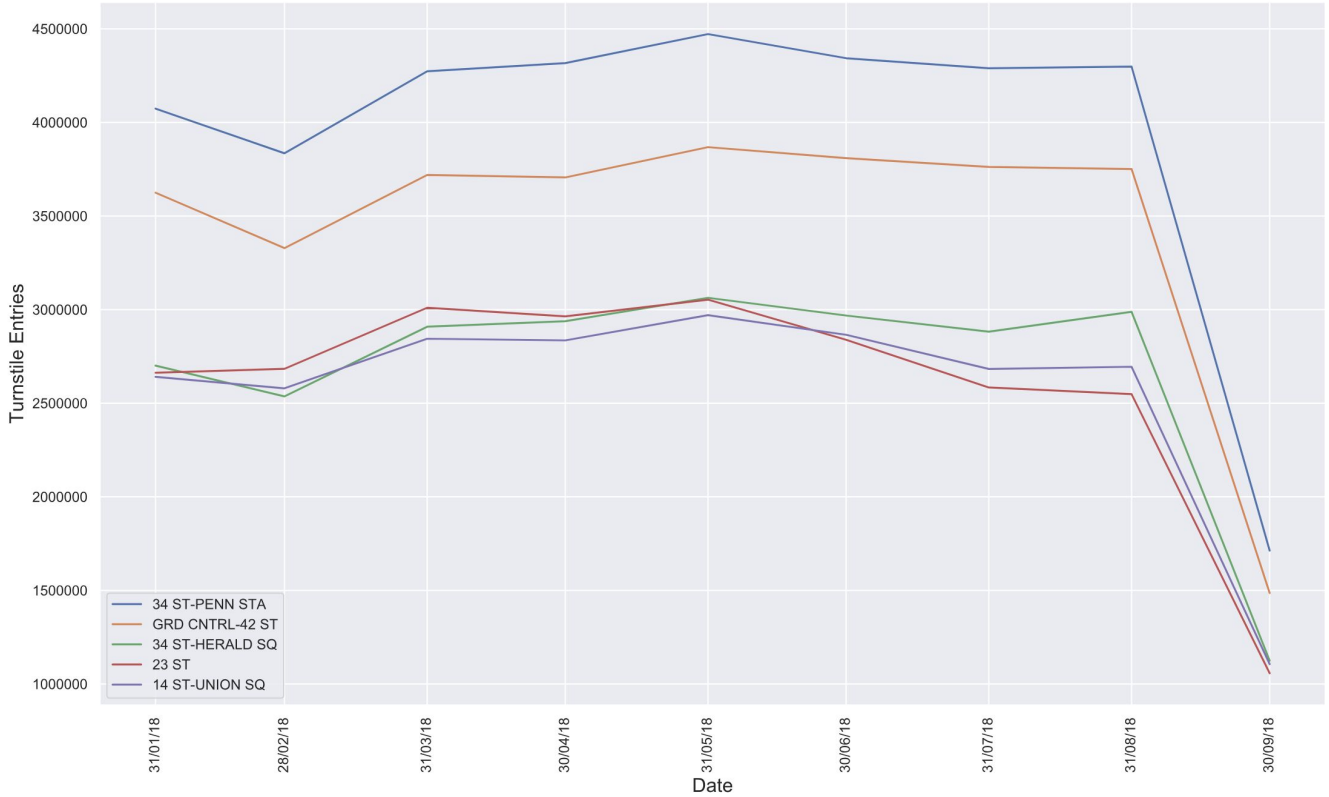
Appendix

Data Cleaning & Wrangling

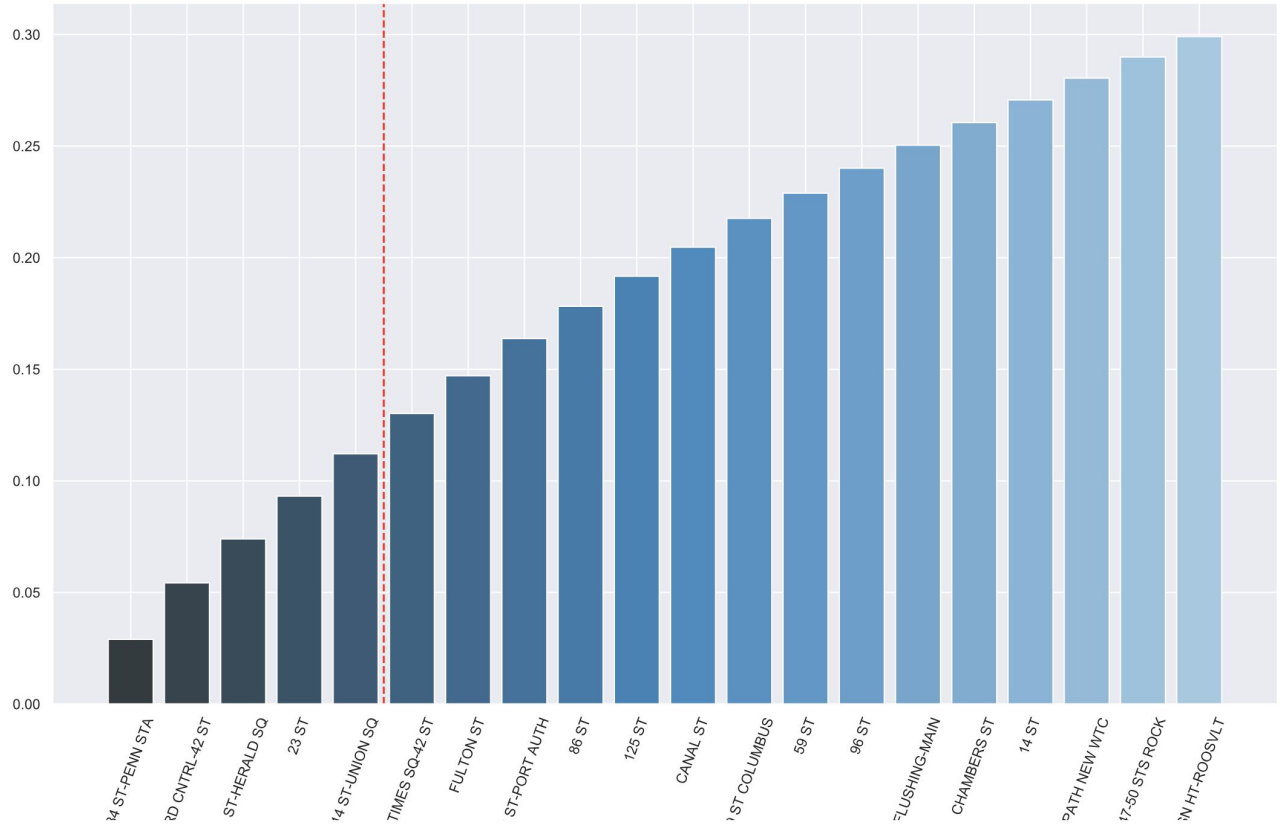
Data Cleansing & Enrichment Activities



Monthly Trend of busiest stations



Cumulative


0.30
0.25
0.20
0.15
0.10
0.05
0.00
4 ST-PENN STA
D CNTRL-42 ST
ST-HERALD SQ
23 ST
4 ST-UNION SQ
TIMES SQ-42 ST
FULTON ST
ST-PORT AUTH
86 ST
125 ST
CANAL ST
ST COLUMBUS
59 ST
96 ST
FLUSHING-MAIN
CHAMBERS ST
14 ST
PATH NEW WTC
7-50 STS ROCK
N HT-ROOSVLT

Data Dictionary: MTA Turnstile Usage Data

Turnstile Usage Data
Data Dictionary

Data Label	DataType	Data Description	Examples
C/A	TEXT	Control Area name/Booth name. This is the internal identification of a booth at a given station.	**A002** 4 to 5 characters.
UNIT	TEXT	Remote unit ID of station	**R001** 4 characters. A numeric designation preceded by 'R'
SCP	TEXT	Subunit/Channel/position represents a specific address for a given device	**01-00-01** Normally six characters in groups of 2 separated by a dash.

Link to Data Dictionary

Link to Github Repo

github.com/RedGeryon/MTA-subway