Project 1 Presentation

MTA Exploratory Data Analysis
Winter 2021 Data Science Cohort

Sturdy Goggles
Gabriel Equitz (Ridge)
Patrick Norman (Ridge)
Anubhav Pareek (Ridge)

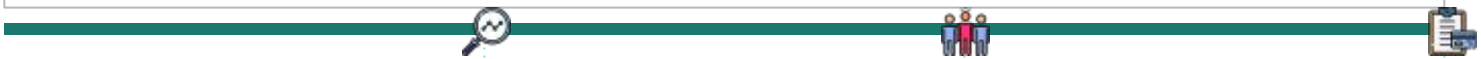
Presented January 8, 2021

Problem Statement

Our team, **Sturdy Goggles (SG)** has collaborated with WomenTechWomenYes (WTWY) to explore and seek recommendations on optimizing deployment of their street team to collect email addresses of potential attendees for their upcoming gala event.

Assumption : WTWY Gala is in June. Our data analysis was focused on late April and May.

Dataset used: http://web.mta.info/developers/turnstile.html



MTA DATA ANALYSIS

- Data Cleaning:
  - Outliers Filtered (1% outliers removed)
  - Inconsistent Turnstile data removed
- Shortlist Top 5 stations based on Footfall Traffic
- Analysis Parameters:
  - Hourly Traffic
  - Daily Traffic
  - Monthly Traffic

DEMOGRAPHIC ANALYSIS

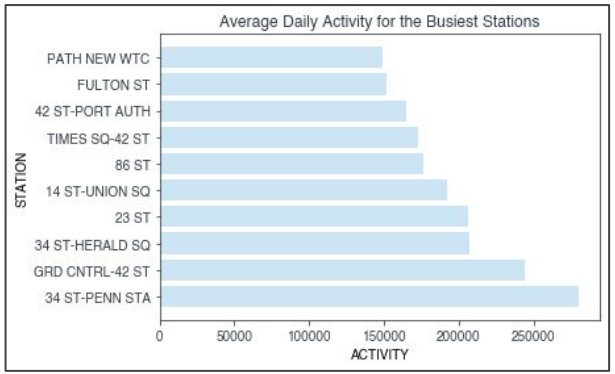
- Analyze demographics for top 5 stations
- Parameters looked at :
  - Household Income
  - Women Population
  - Public Transport Activity

SCORECARD

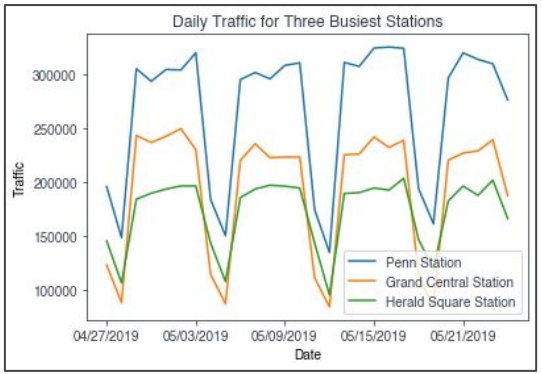
- Report snapshot based on the below parameters:
  - Daily Traffic
  - Women Population
  - Public Transport Activity
  - Household Income
  - Prox. to Tech Hubs
- Final Recommendation

MTA Data Analysis

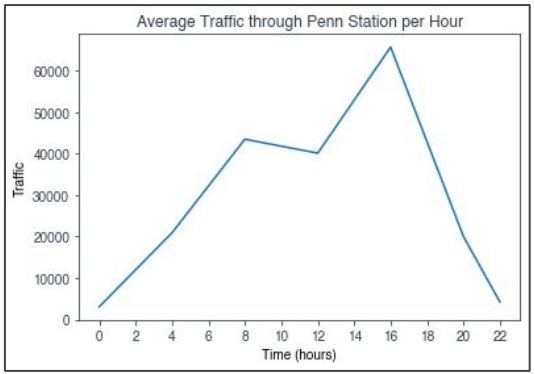
04/28/2019 - 05/24/2019



04/28/2019 - 05/24/2019



04/28/2019 - 05/24/2019

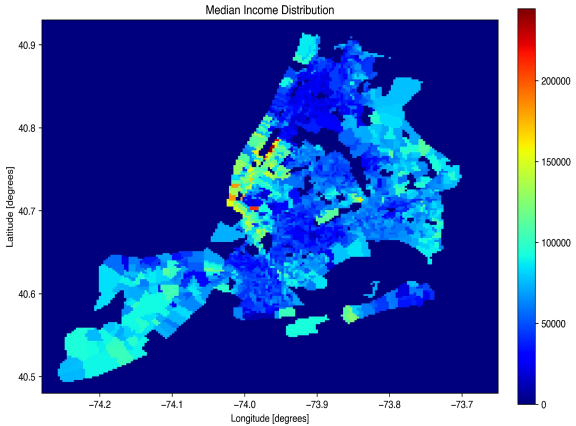


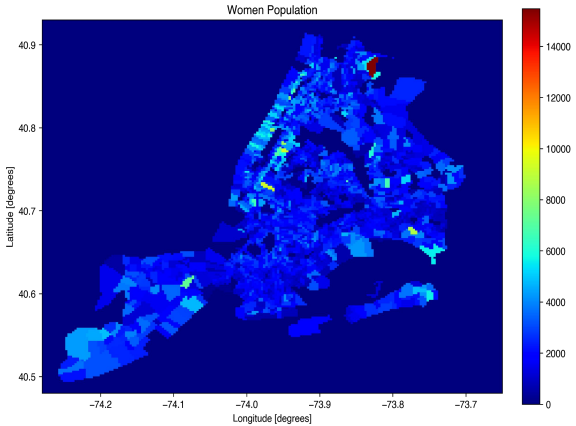
Abbreviation	Full Station Name	Location
34 ST-PENN STA	34th Street - Pennsylvania Station	Midtown Manhattan
GRD CNTRL-42 ST	Grand Central Station - 42nd Street	Midtown Manhattan
34 ST-HERALD SQ	34th Street - Herald Square	Herald Square, Midtown Manhattan
23 ST	23rd Street	Madison Square Park
14 ST-UNION SQ	14th Street - Union Square	Union Square

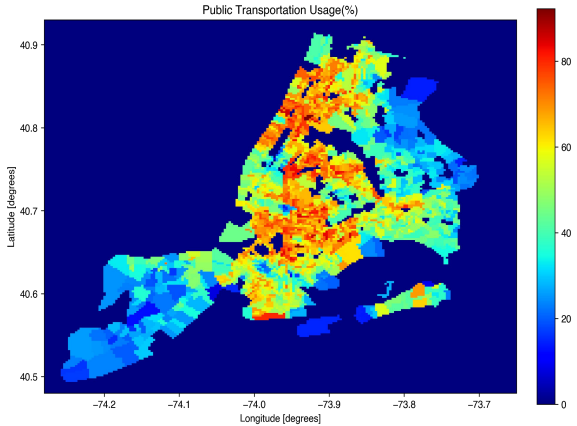
Key Observations :

- **Weekend Effect**
- **Afternoon Peak**

Demographic Analysis







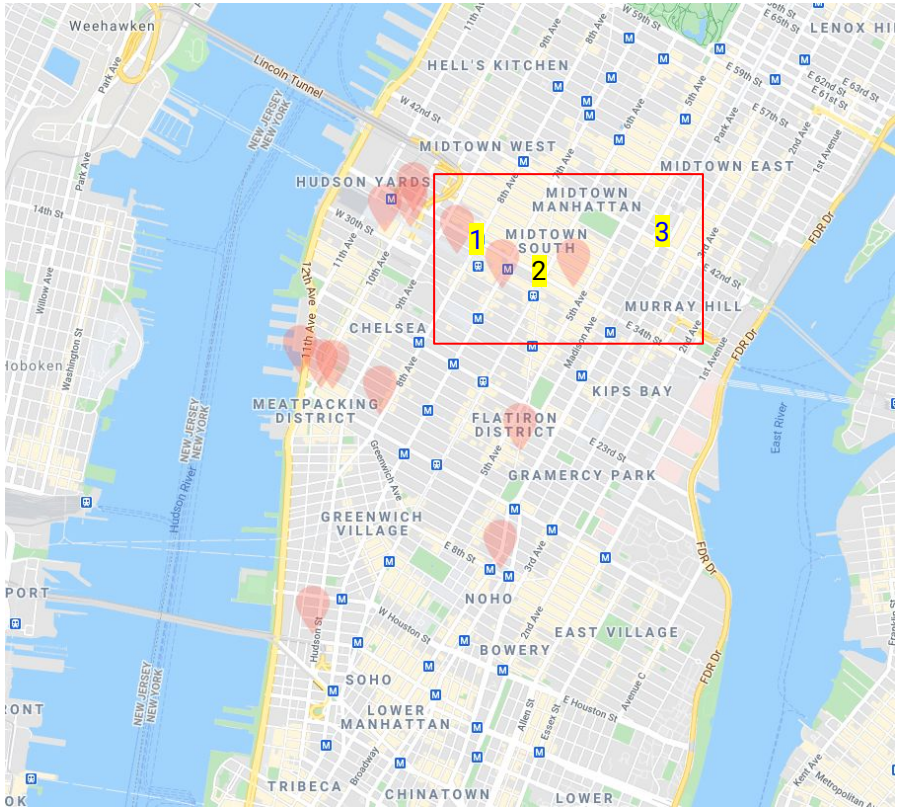
Key Observations:

- Subway Stations within **~1.5 mile radius**
- **Women population > 50%** : Areas(~0.5-1 mile radius) around 34 St-Herald Sq and 42nd St Grand Central stations
- **< 50% population** residing around the top 5 subway stations use public transit

Proximity To Tech Hubs

Key Observations

- **Station 1 - Penn Station,** is between Facebook and Apple
- **Station 2-Herald Square**, is between Apple and Amazon
- **Station 3 - Grand Central**, is distant from major tech campuses



Source: https://ny.curbed.com/maps/amazon-google-facebook-nyc-offices

Scorecard

Station Name	Traffic(Daily Avg)	Women(%)	Transit(%)	Median Household Income	Proximity to Tech Hubs	Score	Rank
Weight(%)	50%	15%	15%	5%	15%	Max - 3 Min - 1	
34th Street - Pennsylvania Station	275K	40%	40%	160K	3	**2.6/3**	**1**
Grand Central Station - 42nd Street	250K	52%	37%	120K	2	**2/3**	**2**
34th Street - Herald Square	200K	56%	37%	100K	3	**1.85/3**	**3**

Key Observations:

- Traffic Volume - Highest Weight(50%)
- **Weighted average** scores calculated on a scale of 3 points. Maximum - 3 points and Minimum - 1 point
- Median Household Income - 3 points awarded to lowest Median Income

Recommendation : 34th Street - Penn Station on Thursday/Friday (3pm - 5pm)

APPENDIX

Data Sources

MTA subway station data: http://web.mta.info/developers/turnstile.html
NYC census data: https://www.kaggle.com/muonneutrino/new-york-city-census-data?select=nyc_census_tracts.csv
Reference :
https://medium.com/@cipher813/analyzing-nyc-subway-and-demographic-data-to-optimize-street-team-deployment-2614522bd83e
NYC tech map: https://ny.curbed.com/maps/amazon-google-facebook-nyc-offices

Libraries used

- Pandas
- Numpy
- Scikit-learn
- Matplotlib

Thank You for Watching!