

MTA Turnstile Data Analysis

By Team JSP: Jhonsen, Sharmila, and Po-Yan

Project Goal

WomenTechWomenYes(WTWY) has a **annual gala** every summer and would like to leverage data analytics to optimize the **placement of their street teams** to maximize data collection from the contacts.

Methodology

Data Sources

- Downloaded **MTA Turnstile** and Income data from relevant sources* for 2018
- Grabbed the Median Income for Zip Codes in NYC from **Census data** for 2017
- Get a list of NYC MTA Subway Stations with names and corresponding geocodes
- Used **Google API** to find Zip Codes for the corresponding geocodes

* web.mta.info, American Census data and NYC Transportation Data

Methodology (Cont'd)

Data Cleaning

- Formatting
 - Renamed the columns; striped leading and trailing whitespaces from the column names
 - Formatted **date** and **time**
- Used *Unit* and *SCP* variables as the key to identify each MTA turnstile data; removed duplicates; sorted the records based on Unit, SCP, Date and Time; computed the delta between the cumulative Entries and Exits
- Outliers
 - Eliminated device counts over 5000 for each turnstile
 - Removed zeros and negative values from the Entries and Exit counts

Tools

Data Analysis	Data Visualization
Pandas	Seaborn and Matplotlib



Results

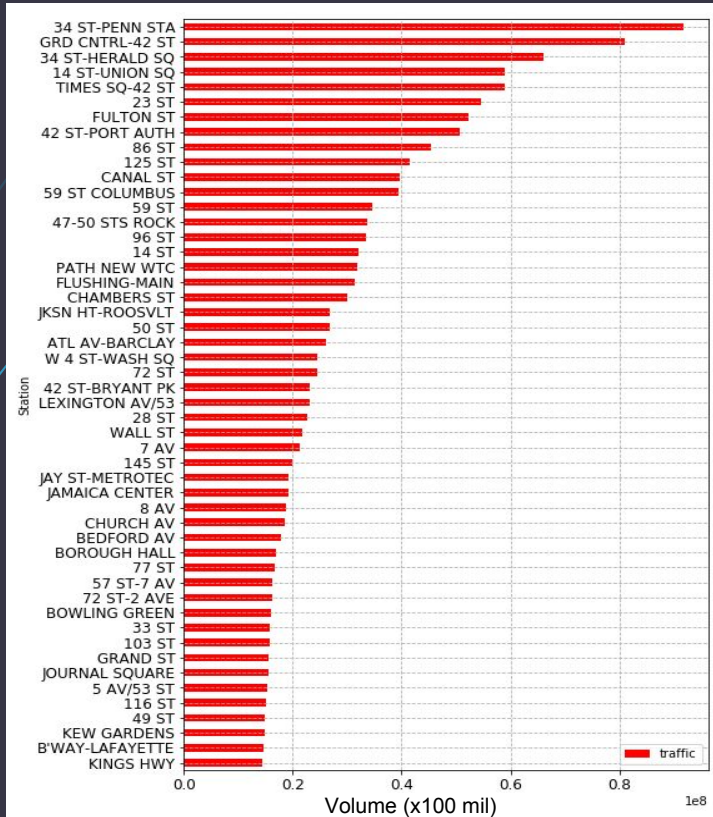
- Top 10 Subway Stations by annual foot traffic volume for 2018
 - Monthly pattern (which months are best)
 - Weekly pattern (which day of the week)
 - Daily pattern (time)

Selecting Best Subway Locations

- **Initial** Dataset
(379 stations)

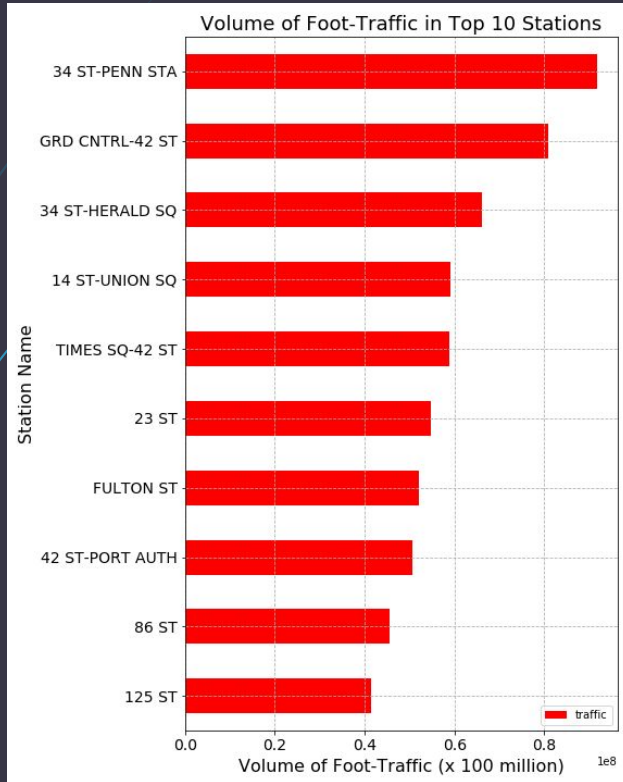


Selecting Best Subway Locations



- Based on Entries and Exit Counts
“Foot-Traffic”

Top 10 Most Busiest Subways



1. 34 St- Penn Station *
2. Grand Central- 42 Station *
3. 34 St- Herald Station
4. 14 St- Union Square *
5. Times Square- 42 Station
6. 23 ST *
7. Fulton ST *
8. 42 St- Port Authority
9. 86 St *
10. 125 St

* Median Income > \$100K

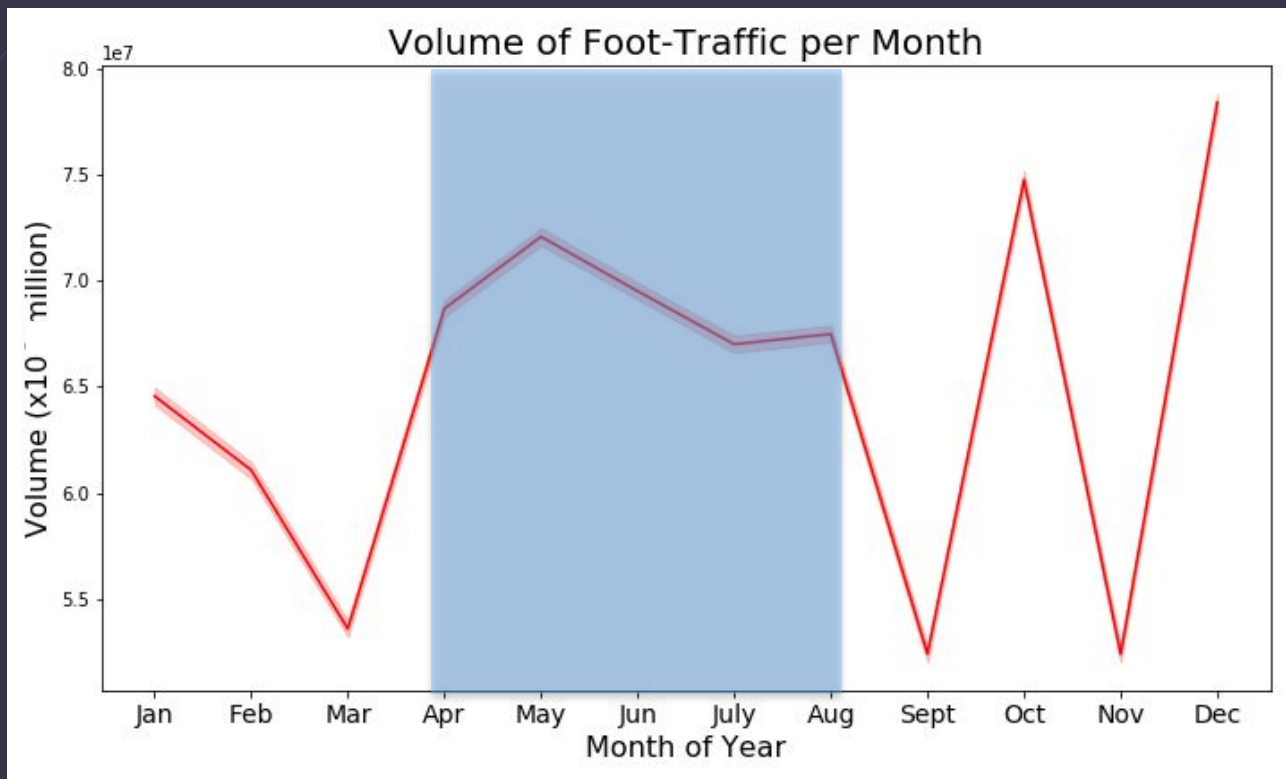
Top 10 Most Busiest Subways

- Dispersed around different boroughs

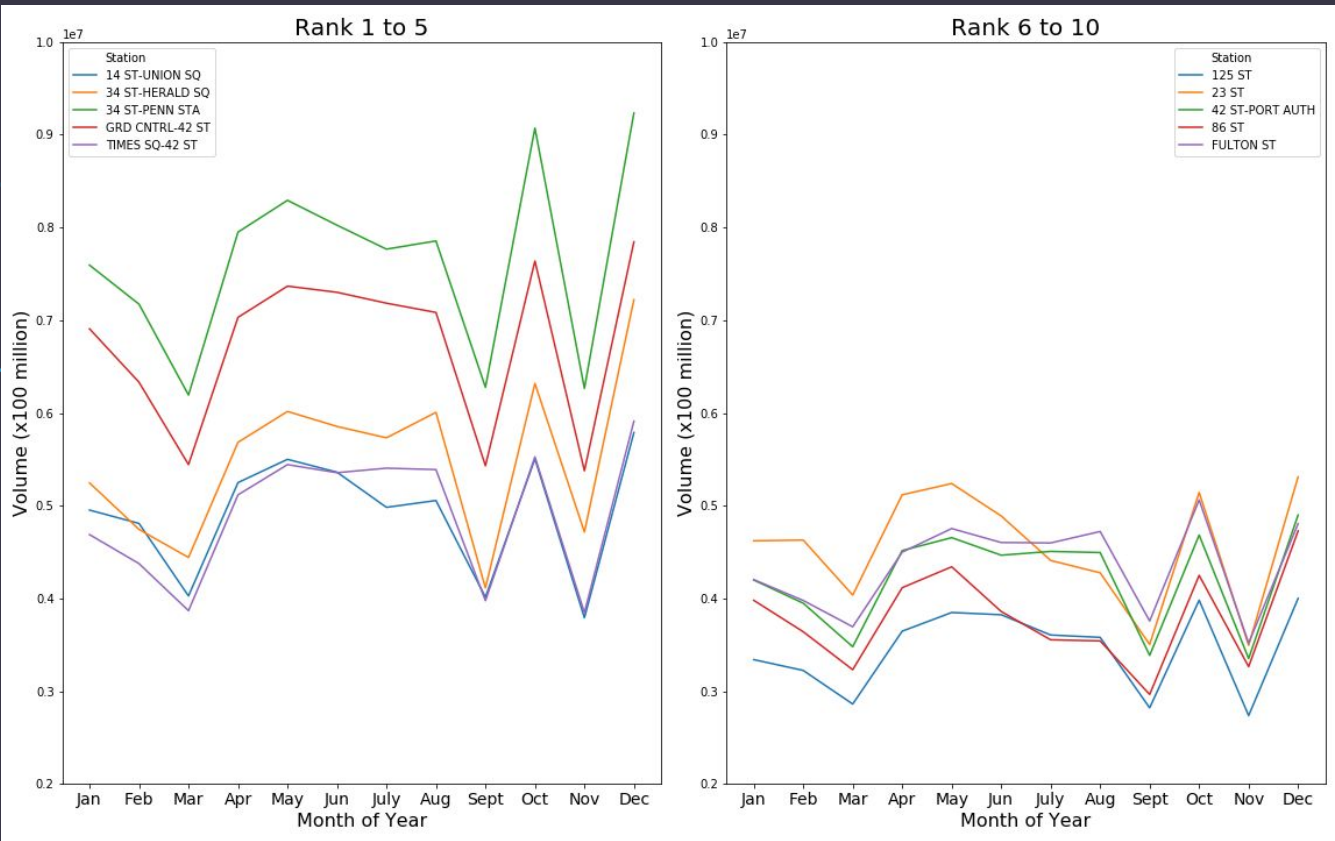
Subway Stations in NYC



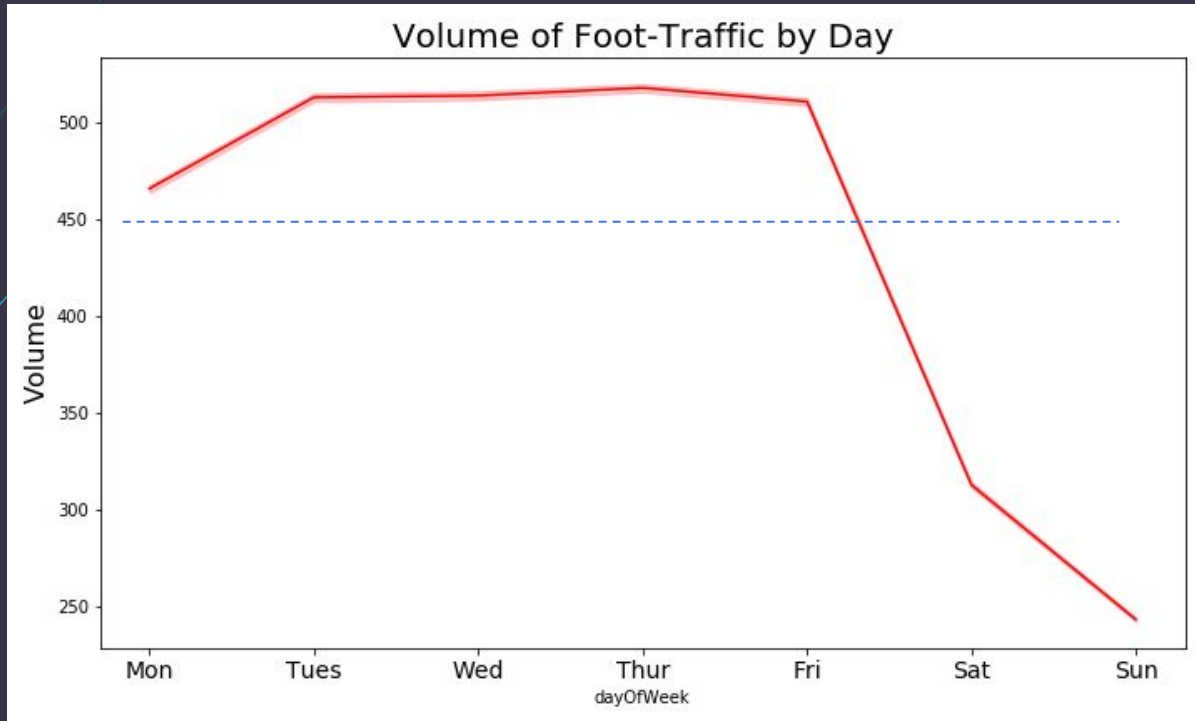
Top 10 by Monthly Foot-Traffic



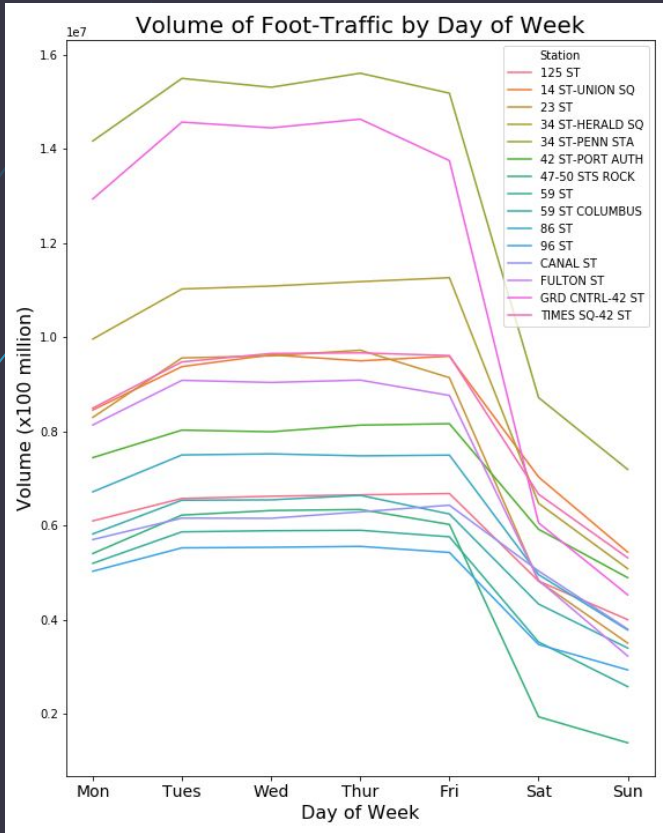
Top 10 by Monthly Foot-Traffic



Top 10 by Weekly Foot Traffic

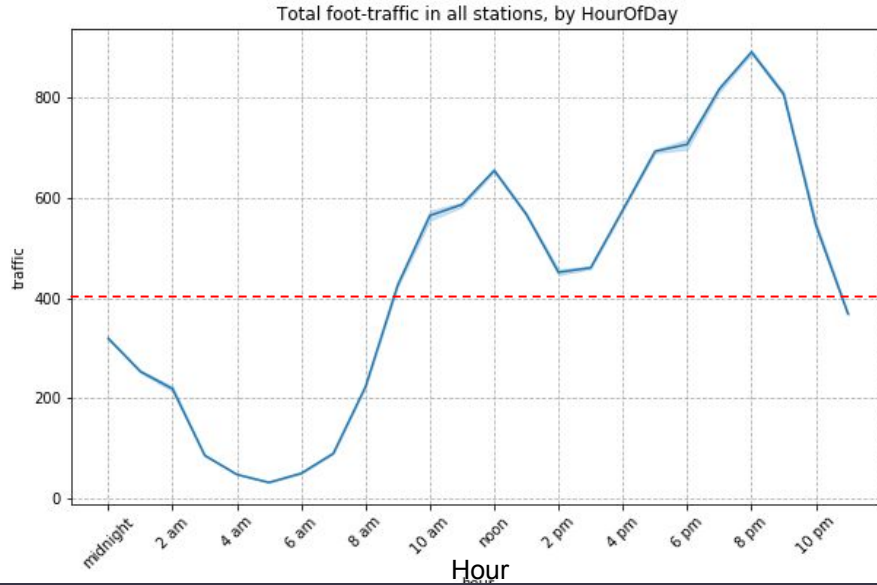


Top 10 by Weekly Foot Traffic

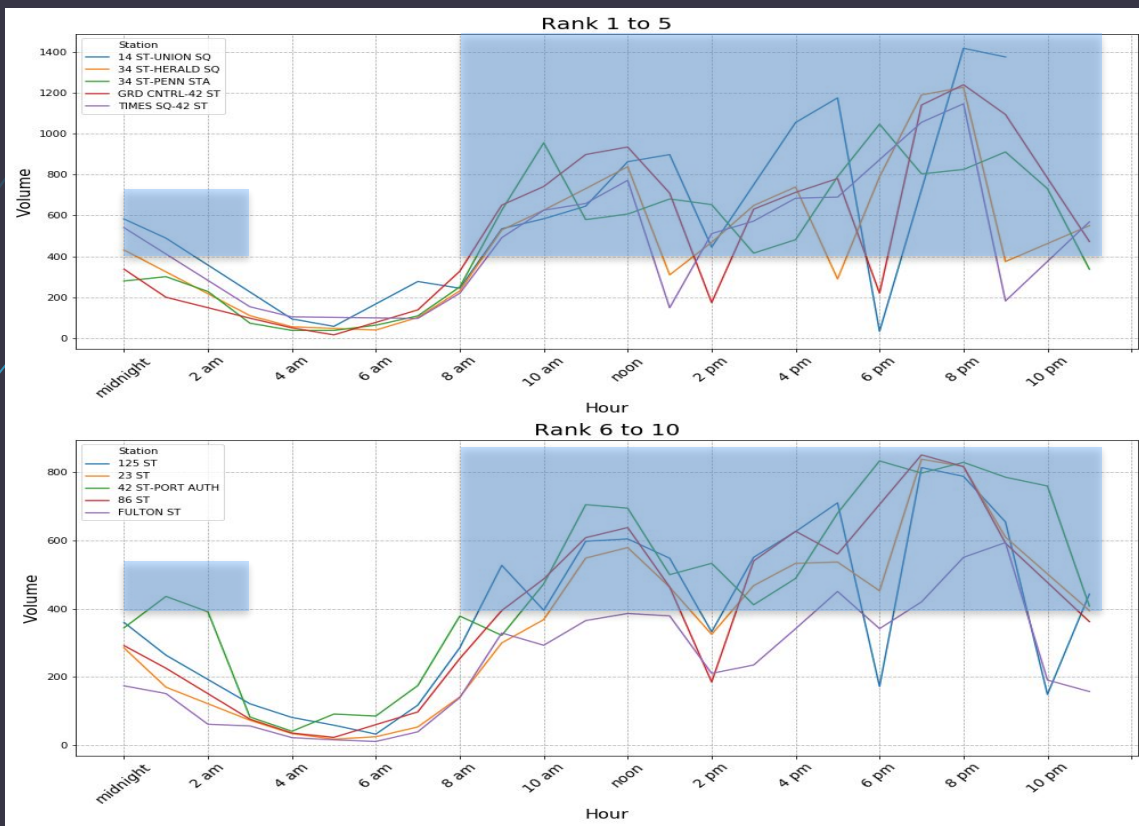


- Weekdays are best!

Top 10 by Daily Foot Traffic and Hour



Top 10 by Daily Foot Traffic and Hour



Recommendation

Station Name	Month	Week	Day/Time
34 St- Penn Station *	April May June	Weekdays	8am- 11pm
Grand Central- 42 Station *			8am- 11pm
34 St- Herald Station			8am- 11pm
14 St- Union Square *			8am- 1am
Times Square- 42 Station			8am- 1am
23 ST *			10am-11pm
Fulton ST *			4pm- 9pm
42 St- Port Authority			8am- 2am
86 St *			9am- 11pm
125 St			8am - 9pm

* Median Income > \$100K



Future Work

- More refined data cleaning e.g. better outlier identification per subway station
- Expand the MTA data set to include more years into analysis to account for variances
- Extra data sources
 - Consider demographics data such as Education attainment, Industry etc.
 - Consider Technology companies geographical locations
 - Align with events targeting Technology and Women



Q & A

Appendix

Data Sources

- NYC MTA Turnstile Data - <https://www.data.gov>
- Census Data - <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>
- Subway Station Names and their Geocodes - <https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49>