MTA Turnstile Traffic Exploratory Data Analysis

Prepared for WTWY

Lucy Abbot
Julian Cheng
Michael Green
Solomon Klein

The Challenge

WomenTechWomenYes (WTWY) wants to optimize the placement of their street teams using MTA data, in order to gather the largest possible amount of attendees for their gala event.

Our Approach

Selecting Data

- Five weeks preceding event
- 2019, not 2020

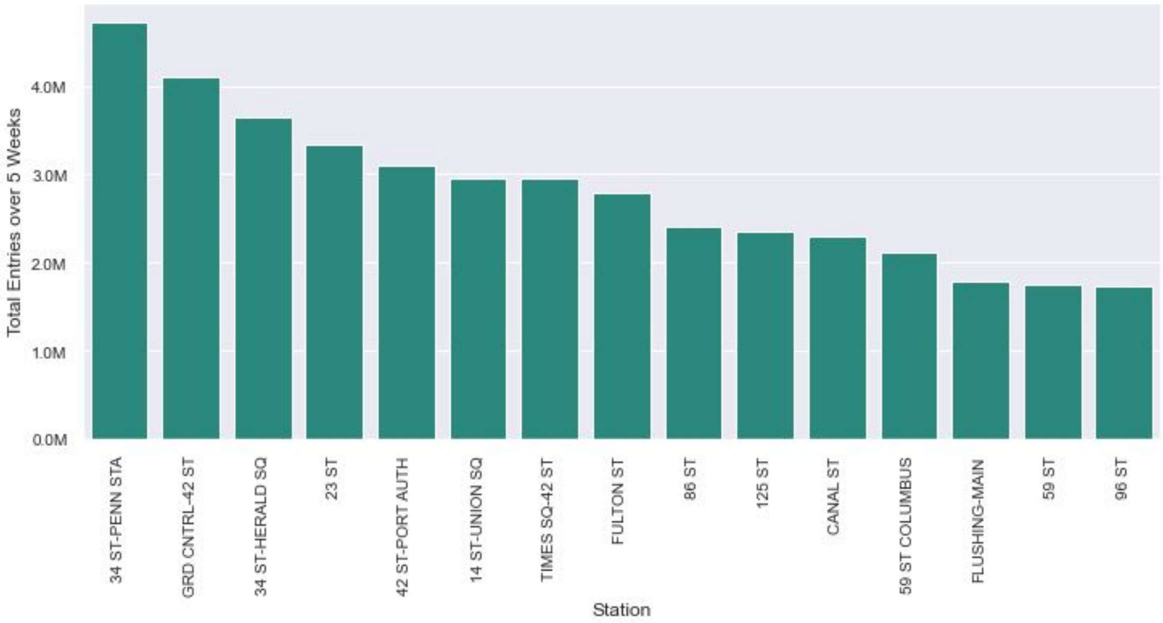
Cleaning Data

- Removed duplicates
- Reversed instances of turnstiles that count backwards
- Removed instances of turnstile resets

Grouping Parameters

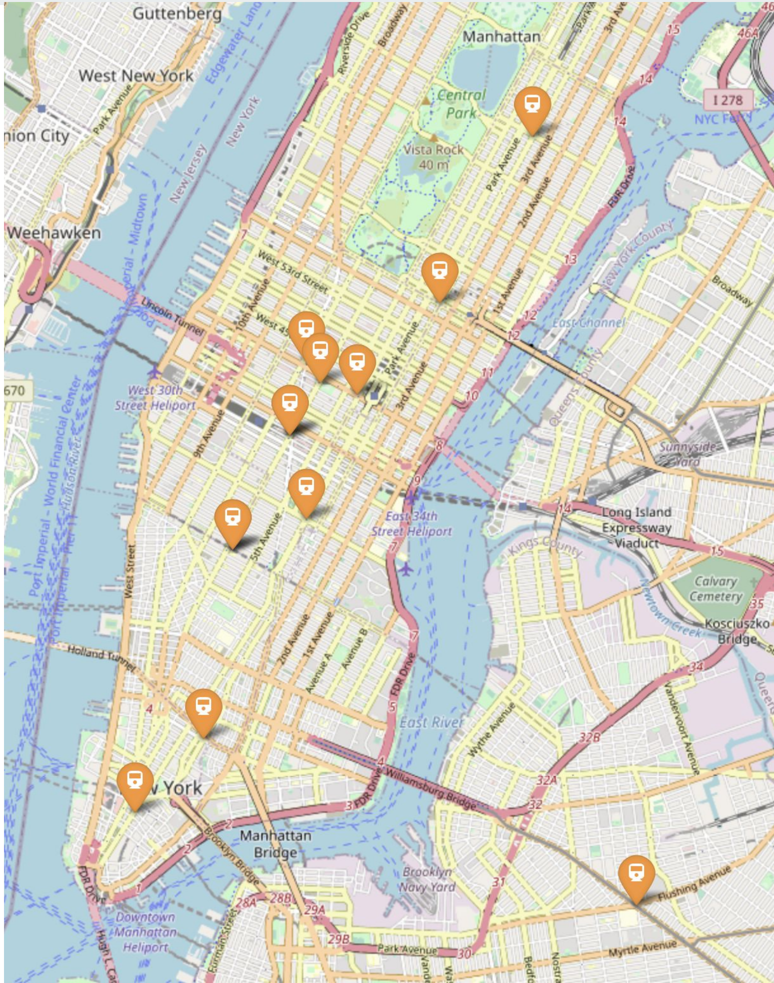
- Turnstiles -> Stations
- 4 Hours -> 24 Hours

Which stations had the most traffic during the period observed?



Location of High-Volume Stations

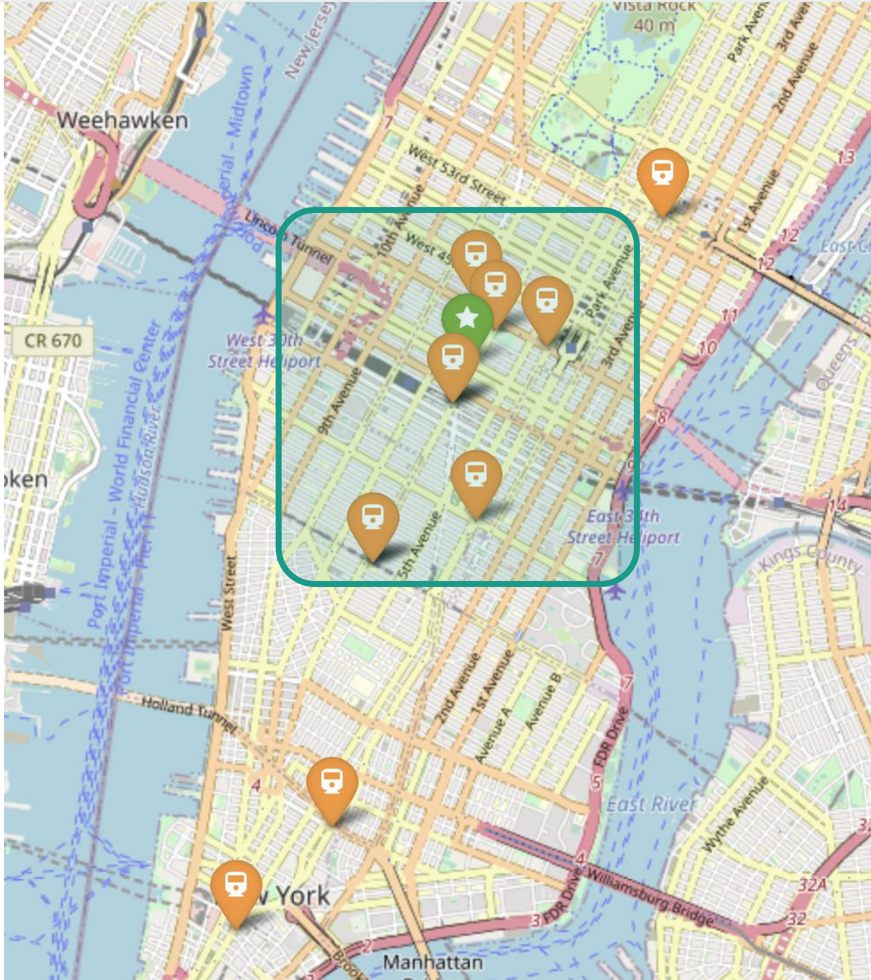
These are the 15 highest-volume stations mapped.



Location of High-Volume Stations

We recommend prioritizing outreach across the six stations located closest to the gala location, Gotham Hall.

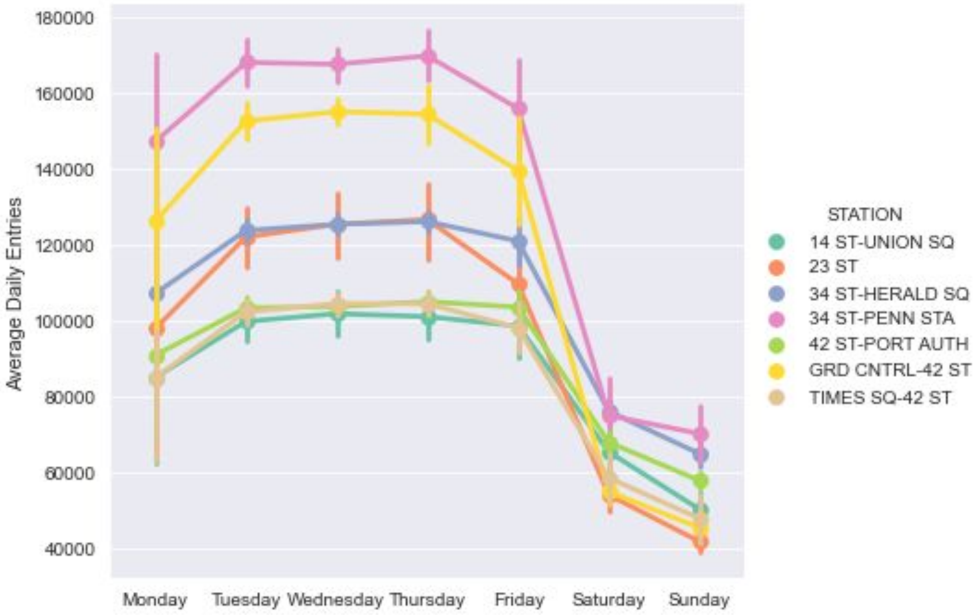
We anticipate higher conversion rates among MTA-riders who already spend time near the event space, since they won't need to go far out of their way to attend.



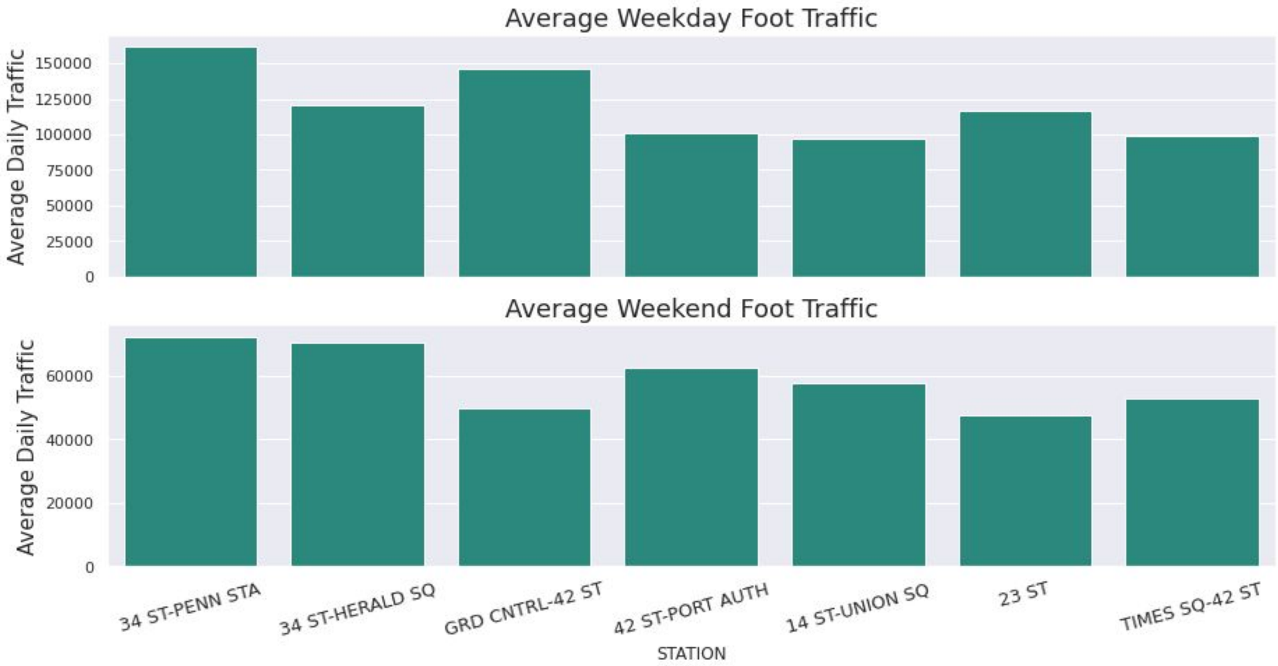
Weekday Variability among Target Stations

Across each of the six target stations, traffic declines significantly on the weekends.

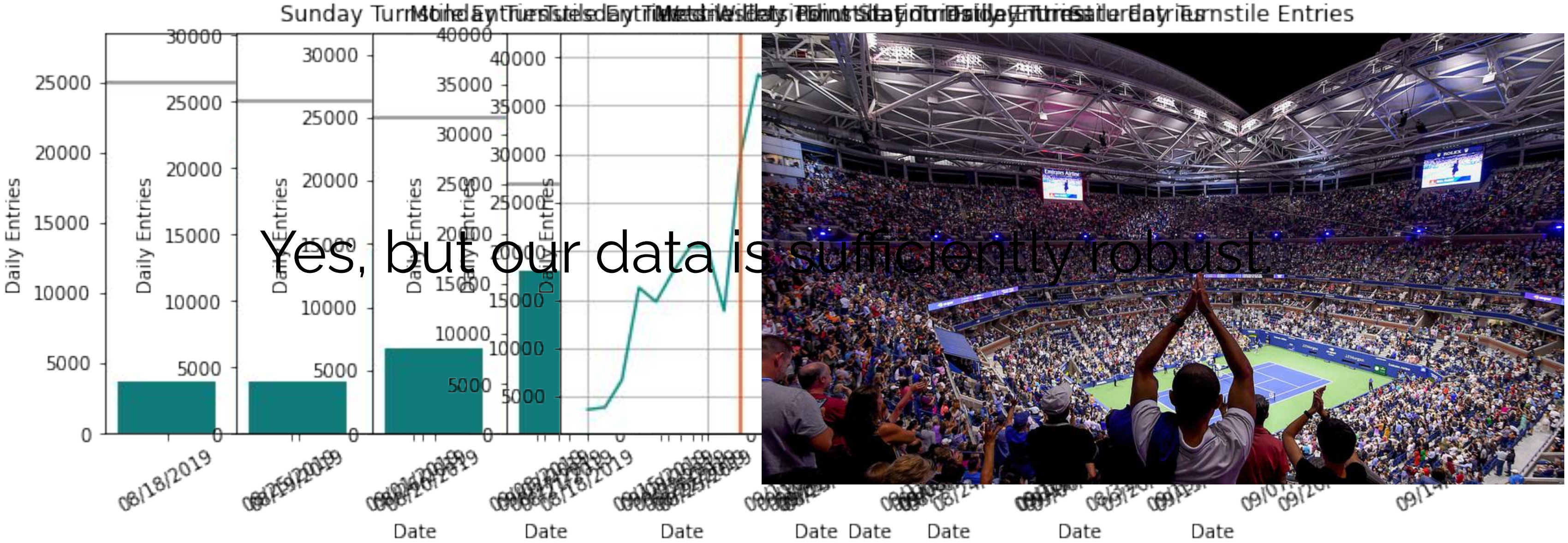
23 St and Fulton St stations decline most significantly on the weekends.



Are there “Weekend Stations”?



Do significant events in NYC affect MTA statistics?



Yes, but our data is sufficiently robust.

Next Steps to Consider

- Build a dashboard that updates regularly, allowing street teams to respond to weekly trends.
- Use neighborhood demographics data to identify stations in key locations for the event.
- Ask street team members to report daily results to augment MTA data with local information.

APPENDIX

Removing Bad Data

01

Turnstiles counting down

- Reverse direction of entries to account for design variations in turnstiles

02

Massive changes in volume

- A turnstile cannot take more than one person per second over an entire day.

03

Resetting counts

- If a turnstile suddenly drops to a very low (<10000) cumulative entry count from a much higher value, it is assumed that its counter had been reset during the day.

Sources

MTA Turnstile Data. MTA, http://web.mta.info/developers/turnstile.html.

NYC Transit Subway Entrance And Exit Data. https://data.ny.gov/Transportation/NYC-Transit-Subway-Entrance-And-Exit-Data/i9wp-a4ja

Dao, Dan. “Best Things to Do NYC August.” *Forbes*, July 2019, https://www.forbes.com/sites/dandao/2019/07/31/best-things-to-do-nyc-august-2019/#228f3ee961fd.

The New York City Subway System. https://www.ny.com/transportation/subways/. Accessed 24 Sept. 2020.

Questions:

- Which station(s) get the most traffic? And on which day(s) do they have the highest number of entries?
  - E.g. seek max per weekday? OR max on weekdays and max on weekends
- For the selected stations, which units get the highest volume? In general, how much variability is there across turnstiles within the same station
- Are there times of day which see higher traffic than others? (Are there stations that see higher traffic even during off hours?)
- For each day, which station has the highest traffic on that day?
- Which stations(s) might have higher concentrations of the target demographic for this event?

Style guide

- Seaborn:
  - Darkgrid
  - color/palette based off: #1a9988ff
- Slides with data:
  - Should include chart/table, and *maybe* a brief takeaway in dark gray text (#434343)

Requirements

Repo <- read.me, .ipynb file

Own repo <- also blog

Introduction + Restate the problem [15 sec], high-level (how we will approach the problem, what concerns need to be addressed), the nitty gritty (dataframe, etc.)