

Exploratory Data Analysis

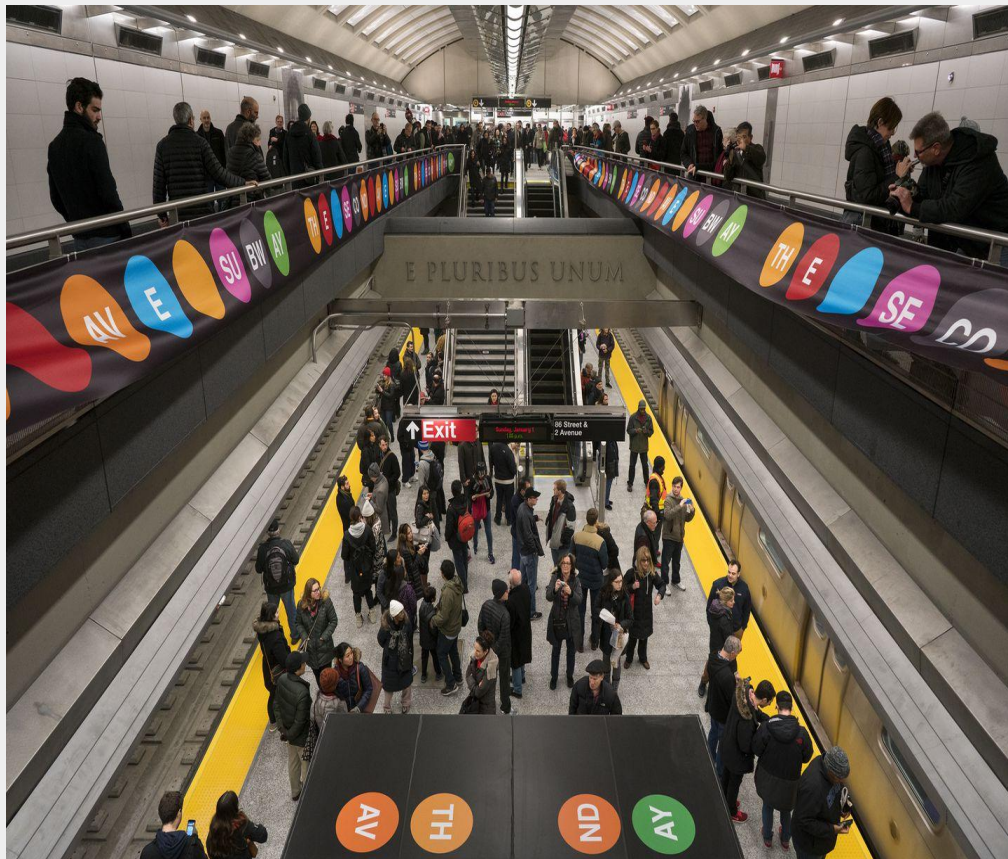
MTA Turnstile Data

Introduction

- Motivations:
 - Expand ZogSports softball league into Manhattan.
- Objectives and Goals:
 - Find the best time to send team members into the field near Yankee Stadium and 86th Street stations.
 - Provide specific time frames with the most traffic.



Methodology



Assumptions:

- More foot traffic will result in more participants.
- People who attend baseball games are more likely to participate in a softball league.
- People that ride enter and exit near Central Park are more likely to join a league located in Central Park.

Data:

- MTA Turnstile data(Yankee Stadium and 86th St stations)
- Apr 2019 - June 2019
- Yankees 2019 Home Schedule (Apr-June)

Tools:

- SQL/ SQLAlchemy
- Python (Pandas, Numpy, Datetime, etc...)
- Matplotlib (Plots & Bar Graphs)

Methodology

Exploration:

- Gathered turnstile data from the MTA Website
- Created SQL database to store MTA data
- Imported database into a Python engine using SQLAlchemy

Preparation:

- Clean data
- Obtain Yankees' Home Schedule through ESPN Website

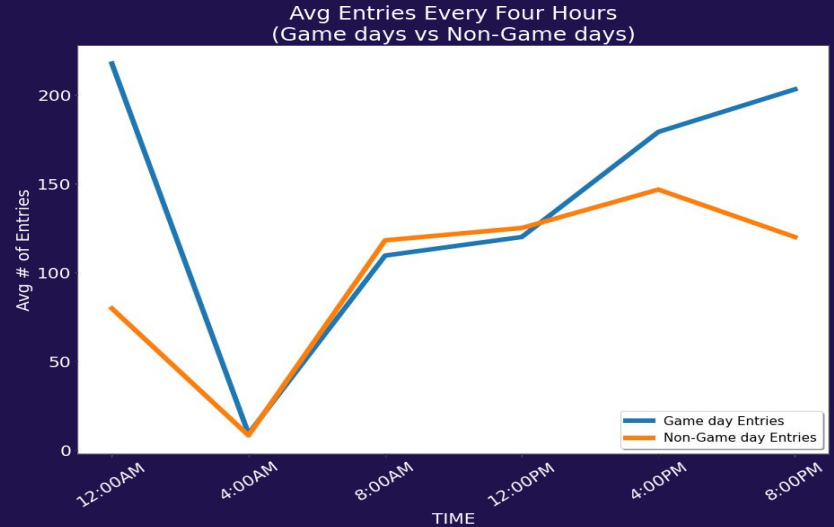
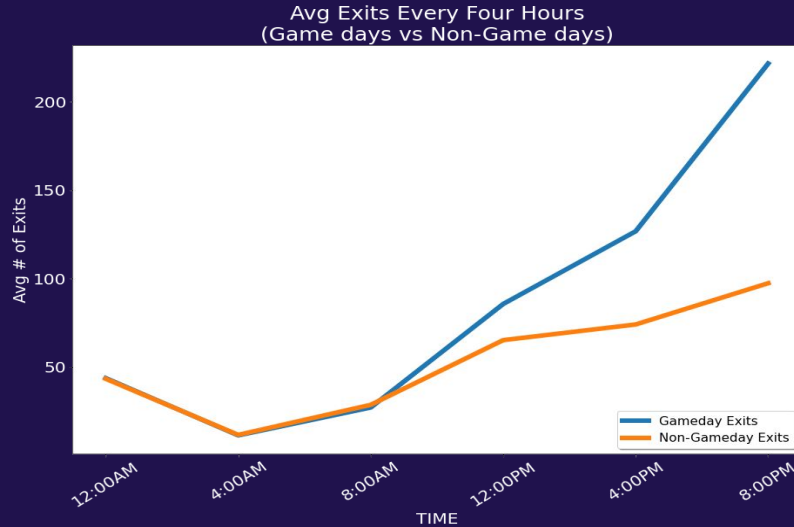
Visualization:

- Create line and bar graphs

Results: Yankee Stadium



Yankee Stadium: Before or After the Game?



Results:

- Game days produce a significant increase in ridership.
- There is a rapid increase in ridership throughout the afternoon and evening.
- Exits skyrocket on game days more so than entries.

Results: 86th Street Station



86th Street: Busiest day of the Week

Avg Entries vs Exits Each Day
(86th Street)



Results:

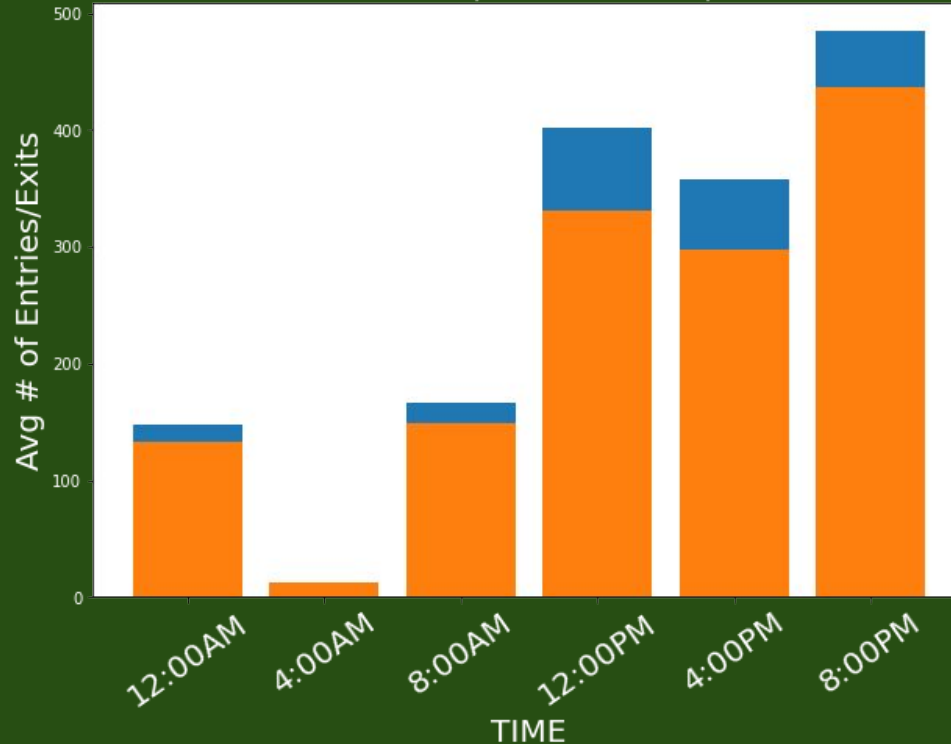
- Middle of the week (T,W,R) have the highest ridership.

However:

- What part of the day is best?

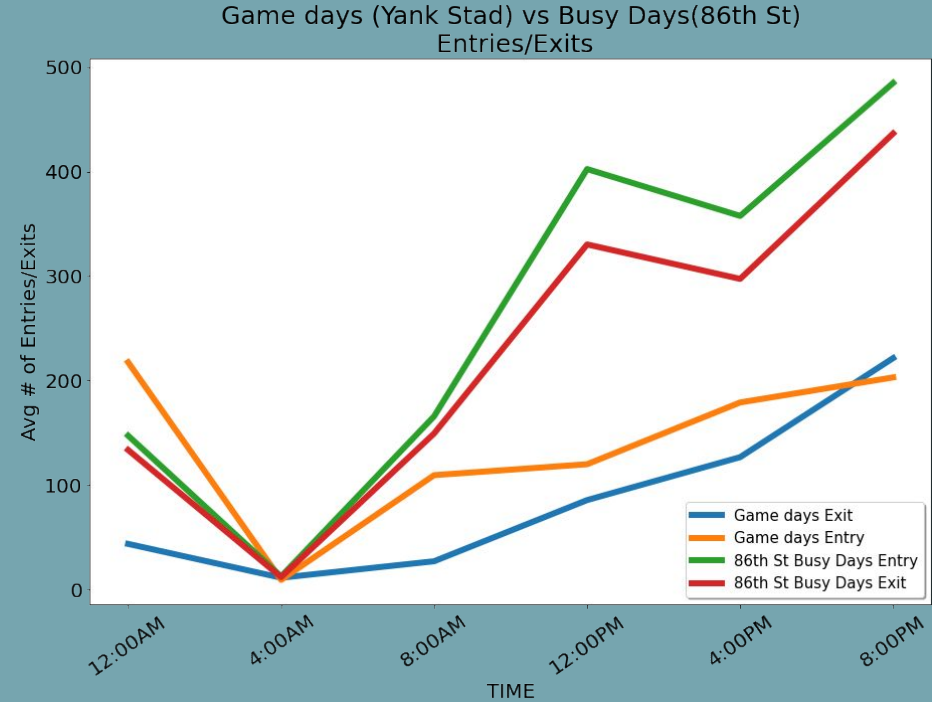
86th Street: A Deeper Look

Avg Entries/Exits by Hour
T,W,R (86th Street)



Which is better?

- If we find ourselves in a situation with limited resources, where should our focus be?
- 86th Street gets almost twice as much traffic during peak hours.



What Now?

- Yankee Stadium
 - A large majority of riders are coming between 12:00 PM to 8:00 PM (similar to a typical start time to games).
 - 2-3 hours prior to game time would be the ideal time for members to go out in the field.
- 86th Street
 - Tuesday, Wednesday and Thursday are the busiest days.
 - Between 12:00 PM and 4:00 PM are the times with the most traffic.
 - If resources are limited we should put our focus here.

Future Work

- This analysis focused on only two stations. There are plenty of other viable options.
- Check how commuters affect our assumption (heavy traffic means more participants). How many people are signing up on their commute?
- Utilize Yankees home game attendance data in order to find out which games would have the biggest impact.



Thank you!