Exploratory Data Analysis
Project Write-up
Fri., April 2, 2021

Abstract:

An independent film company will be shooting a film called *Home & Away* in the Bronx, NYC beginning in August 2021. The primary filming location is Yankee Stadium, however five MTA subway stations are needed as secondary locations. Budget constraints are a concern and data analysis is needed to make sure the most cost efficient stations are used. I worked with data published weekly by the MTA summarizing entries and exits in the system in order to provide the necessary results and conclusions.

Data Description:

The data is the MTA turnstile published to the public weekly. Specifically the 13 weeks between July 27, 2019 and October 26, 2019 were used. Primary data fields included in the analysis were:

- Station
- Linename
- Date
- Cumulative Entries
- Cumulative Exits

Algorithms:

To my knowledge, no algorithms were used in this project. I expect this to change with future projects.

Tools:

SQLite for initial data storage and inspection
SQAlchemy to help load data into a Pandas Dataframe
Pandas and Numpy for data manipulation
Matplotlib and Seaborn for plotting

Communication:

I will be presenting results to the Exploratory Data Analysis cohort on Friday, April 2, 2021. A PDF of the presentation is also included on my personal GitHub repository.