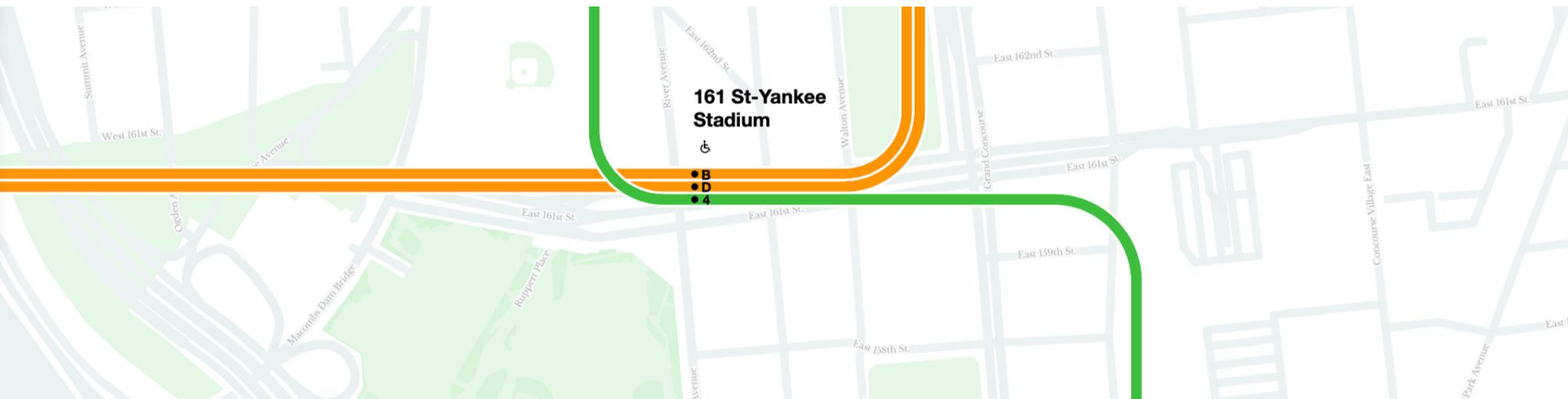


Exploratory Data Analysis of Yankees Home Games and MTA Subway Stations



Motivation

COVID-19 restrictions are easing, and entertainment events are returning to normal frequencies and capacities. How can travel via subways remain safe and accessible to all groups, whilst still allowing for events to take place?

Objective: Carry out an EDA for the frequent commuters of Yankee Stadium Station on how games at the stadium correlate with the foot traffic of surrounding subway stations.

Goal: Provide solutions on how to maintain safe subway travel for all groups (perhaps alternate stations or arrival/departure times).



Methodology

Datasets: 1. MTA Turnstile Data 2. Yankees 2019 Season Schedule Data

1 Data ingestion of datasets via SQLAlchemy and Pandas

2 Data cleaning and aggregation for both datasets

Station 'complex' formulation:

unit	linename	station
R014	2345ACJZ	FULTON ST
R028	2345ACJZ	FULTON ST
R318	G	FULTON ST



unit	complex
R014	R028,R014
R028	R028,R014
R318	R318

Removing duplicate unit-complex associations:

unit	complex
R010	R010
R010	R010,R011
R033	R032,R033
R033	R033
R163	R163
R163	R163,R105



unit	complex
R010	R010,R011
R033	R032,R033
R163	R163,R105

Methodology

3 More cleaning and aggregation

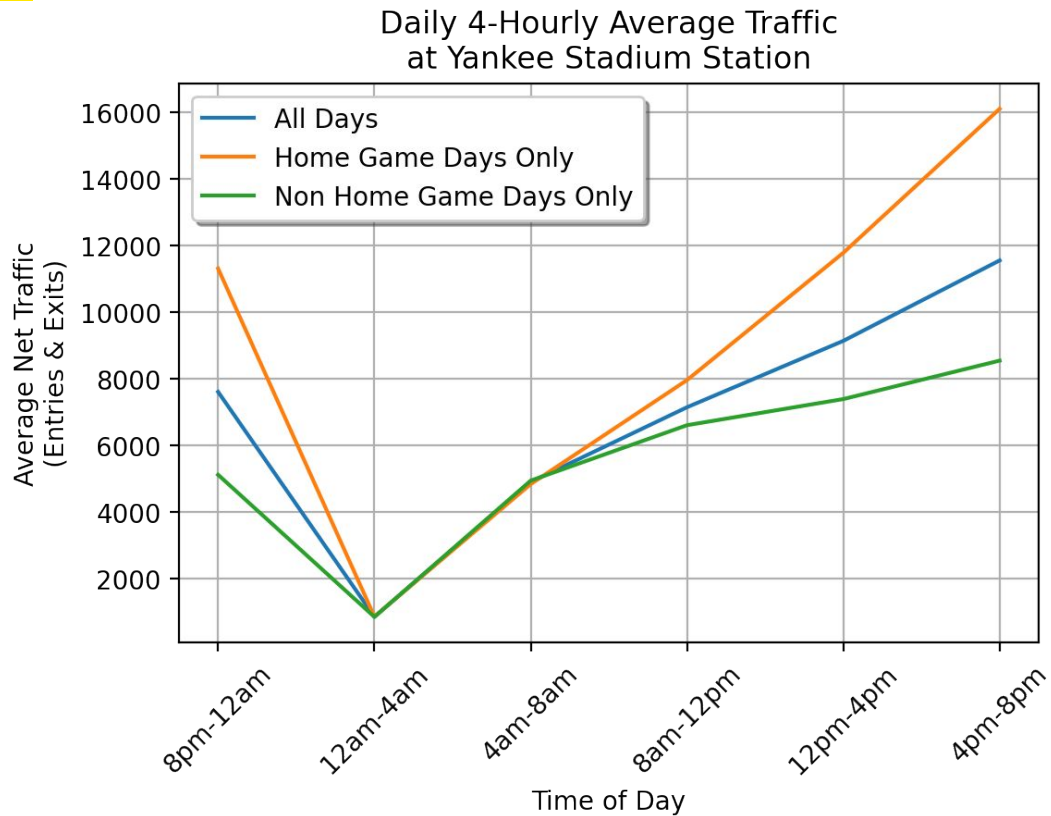
- Dropping invalid/unneeded rows and columns
- Datetime/type conversions
- Aggregation by both time interval and date
- Calculating net_entries and net_exits
- Formulation of net_traffic

4 (Left) Joining both datasets on date

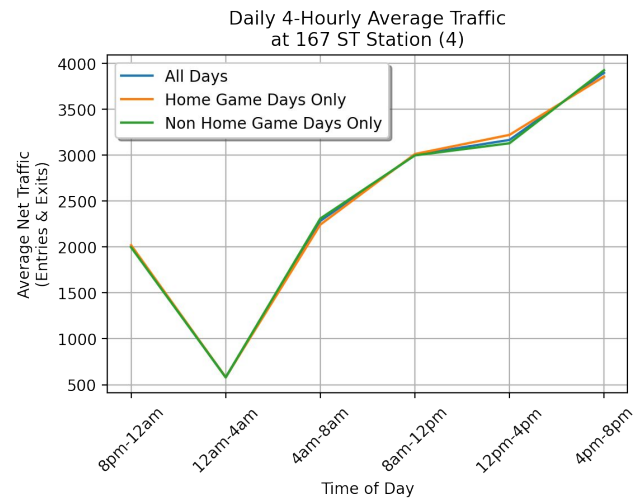
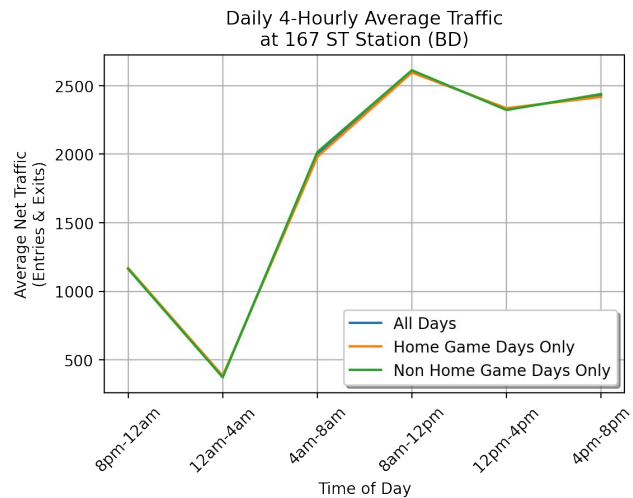
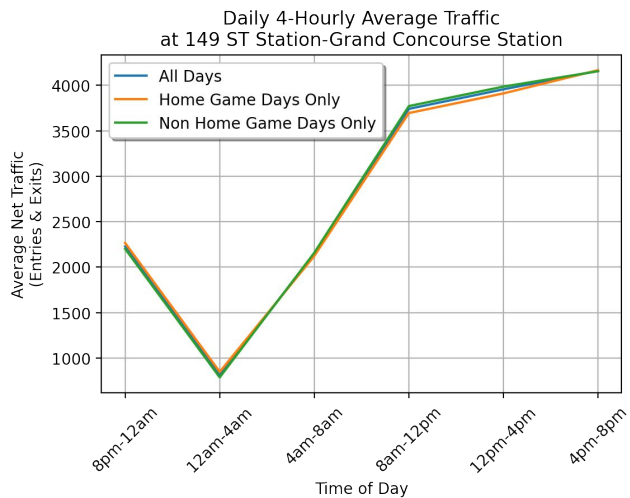
All daily traffic aggregations can be associated with a game and all it's data or no game at all.

4 Visualised findings with matplotlib and seaborn

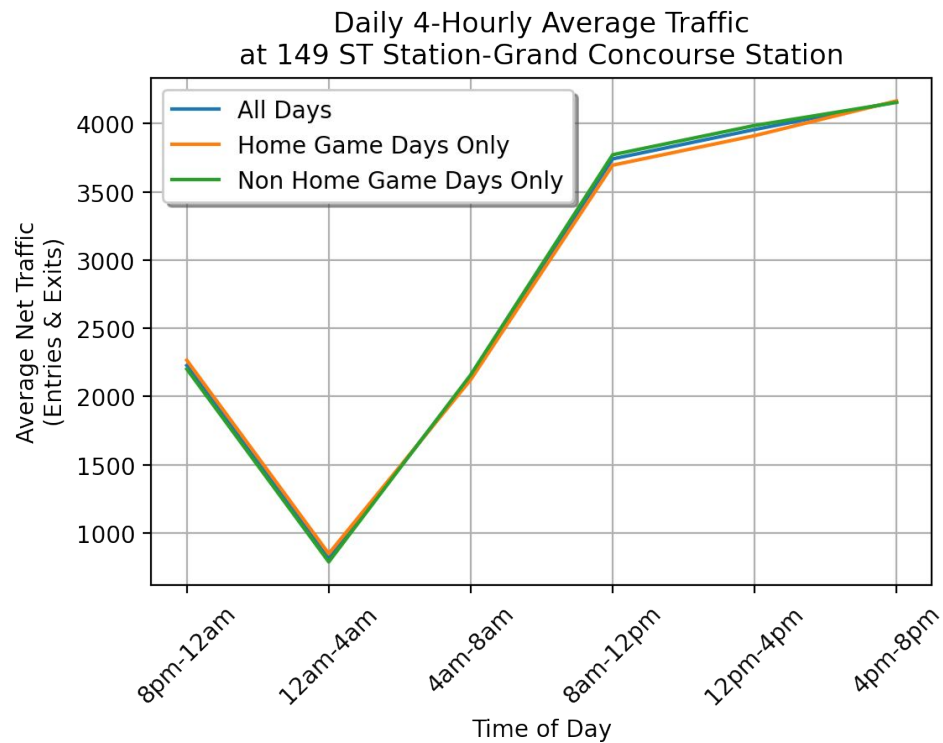
Results



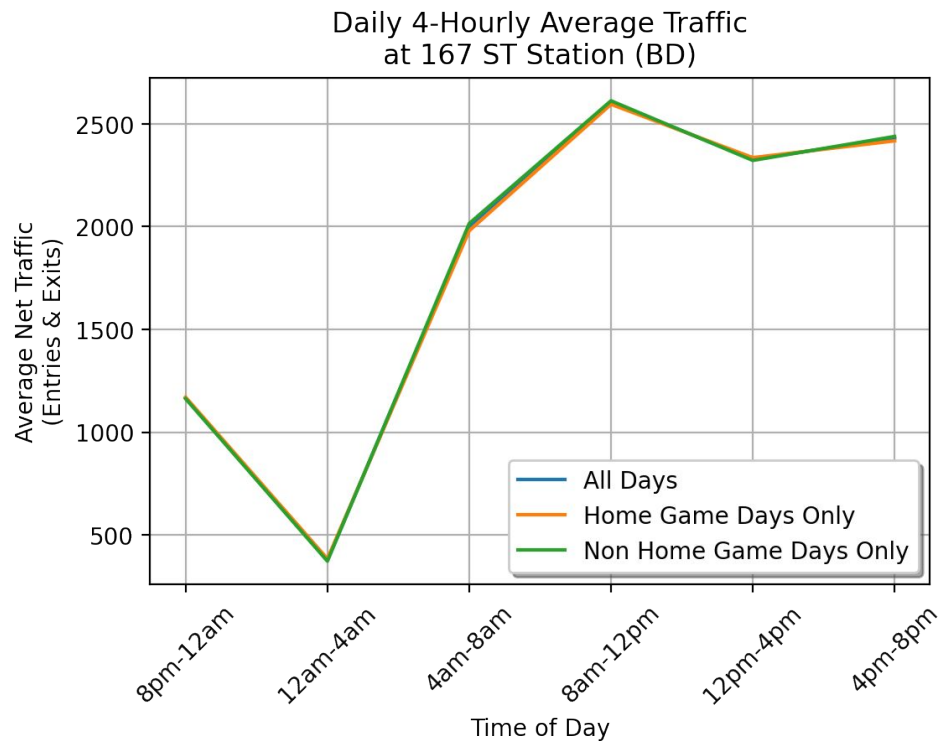
Results



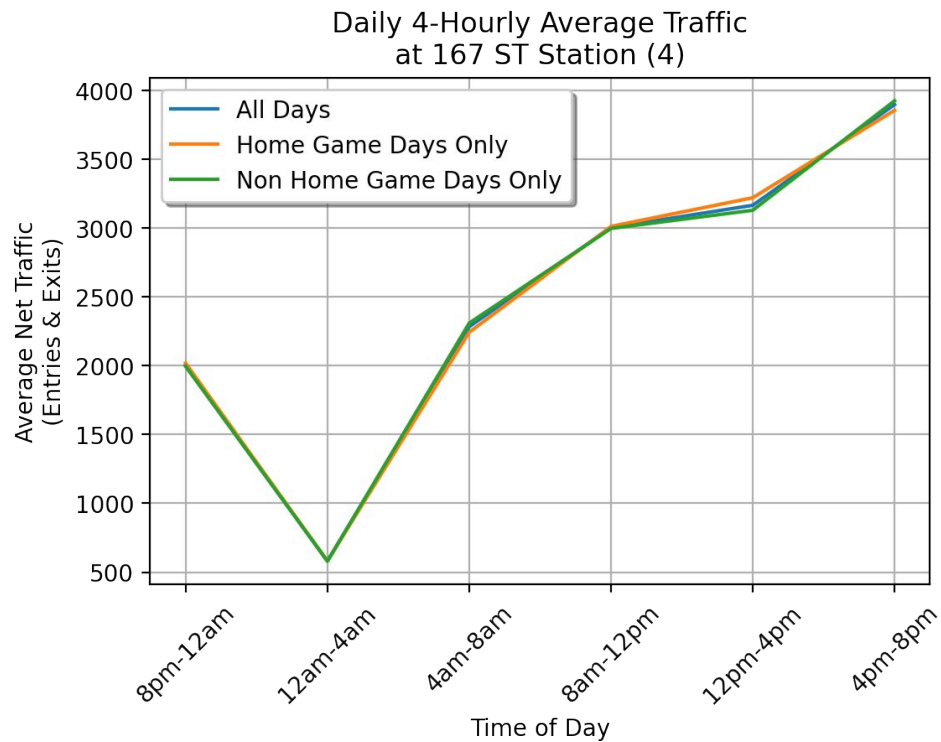
Results



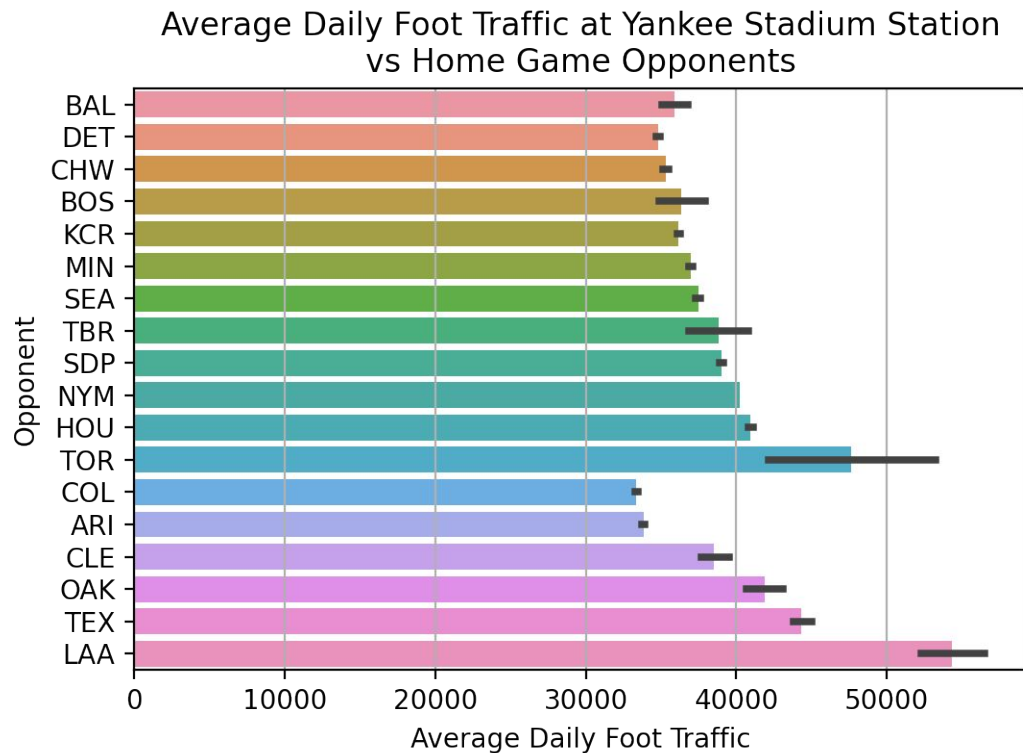
Results



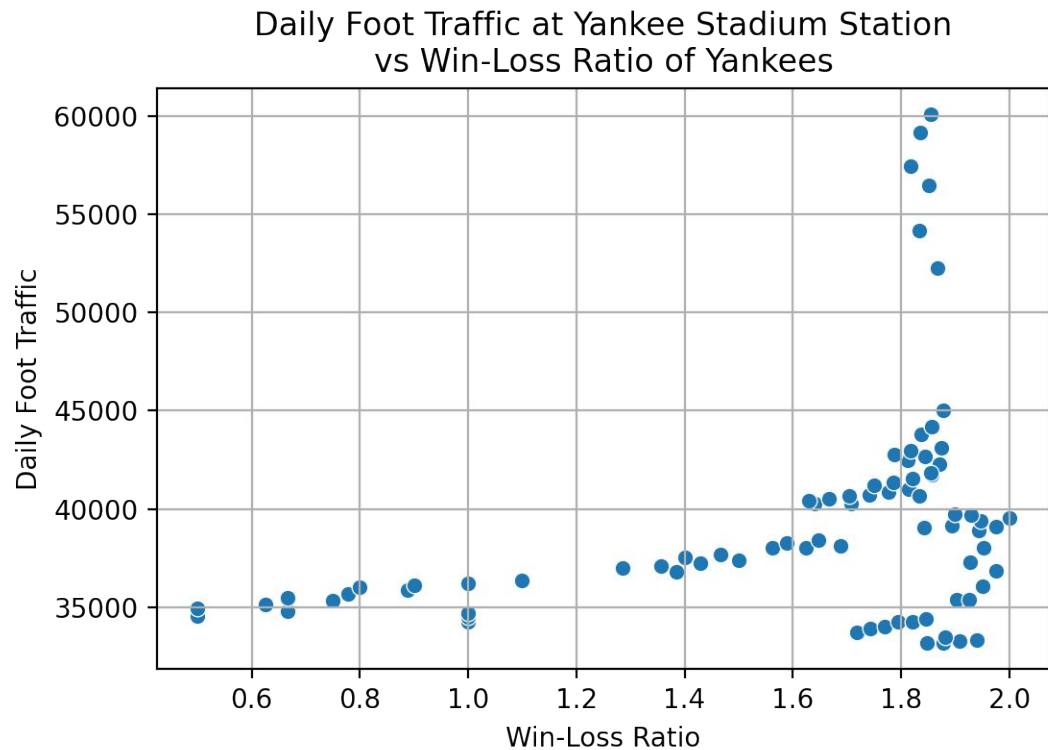
Results



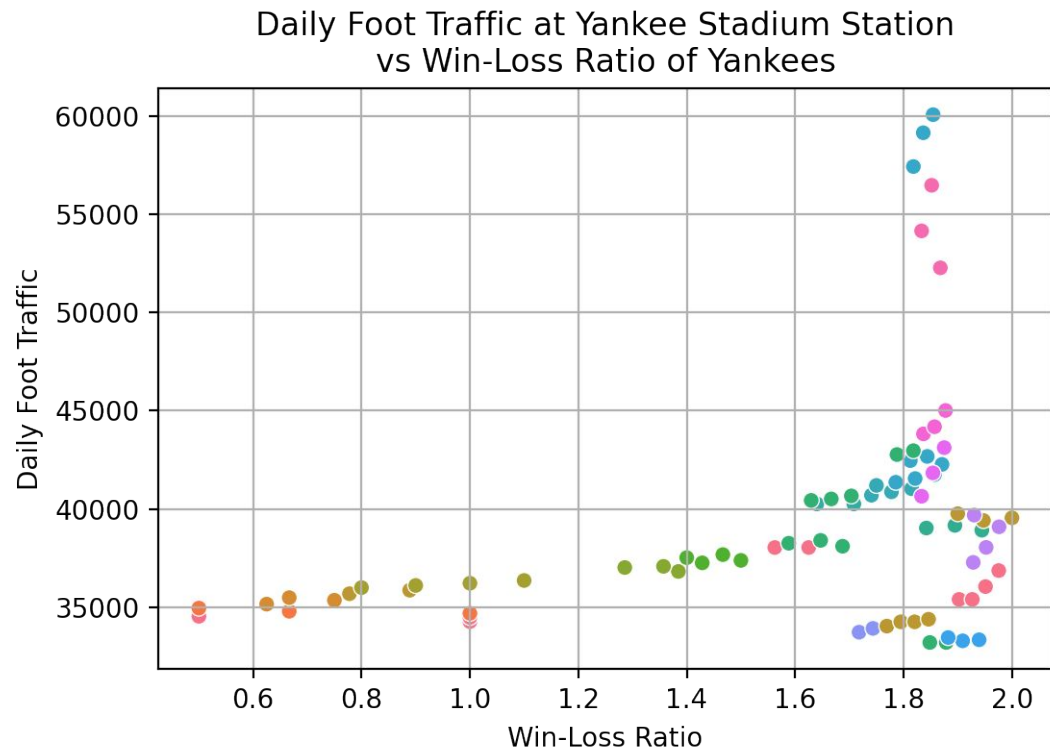
Results



Results



Results



Conclusions & Future Work

- 1 The Yankees playing a home game increases traffic at the Yankees station
 - 2 Nearby stations are a great alternative to avoid crowds
 - 3 Who the Yankees are playing against may influence how large the crowds are
 - 2 The win-loss ratio of the Yankees may also influence how large the crowds are
-
- 1 Explore further stations and see how distance from the stadium relates to traffic
 - 2 Explore other game attributes such as win/lose streak, day/night games.