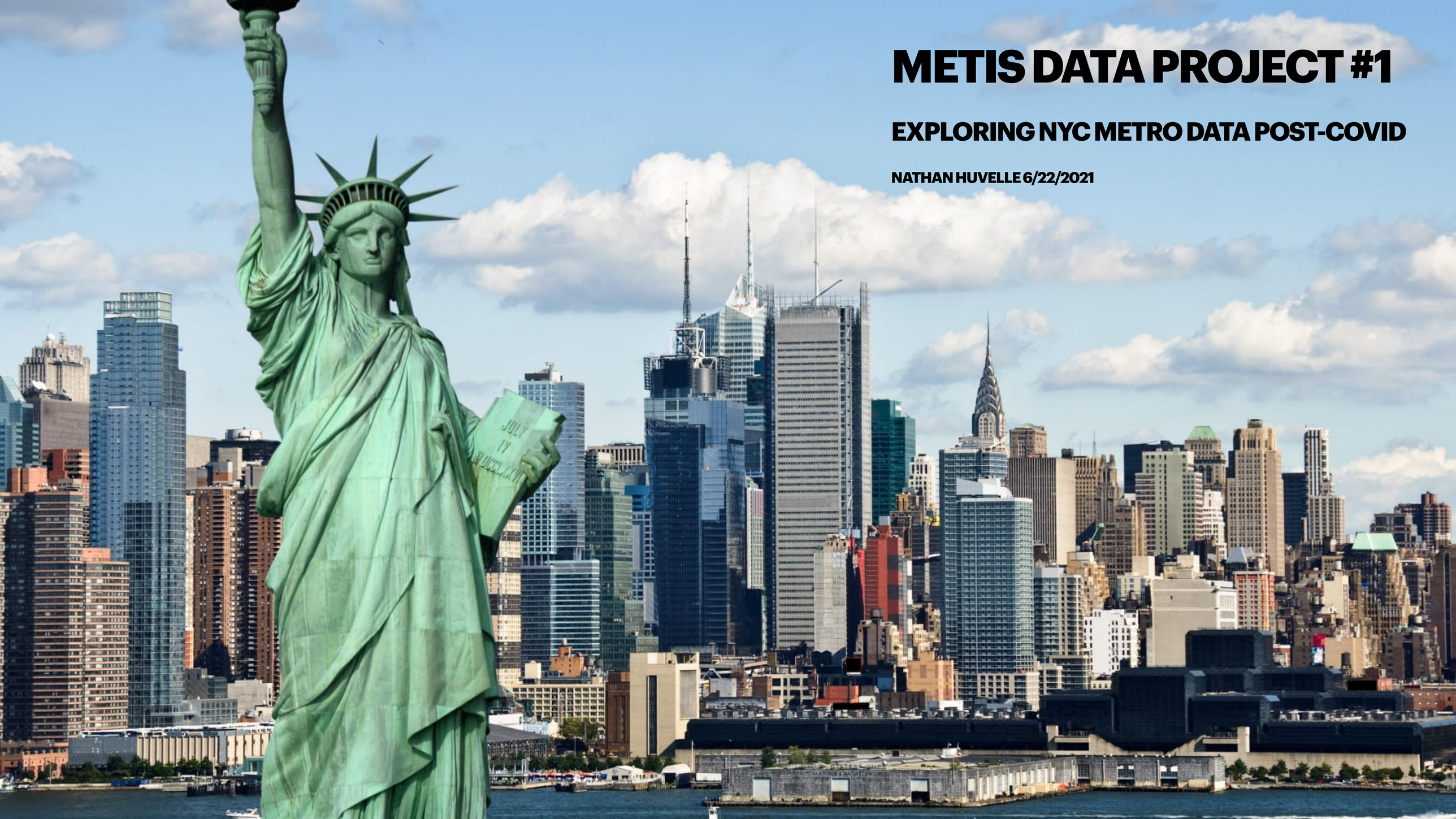


METIS DATA PROJECT #1

EXPLORING NYC METRO DATA POST-COVID

NATHAN HUVELLE 6/22/2021

Introduction

The goal of this project was to examine MTA turnstile data at several stations located near popular tourist attractions in NYC.

New York has been one of the slowest states to reopen post-Covid, with several key opening dates having just occurred in the past few months.

Metro data pulled from the MTA website was used to analyze the flow of travelers at these locations mainly for 2 reasons: To see if re-opening guidelines had a noticeable effect on the amount of tourists traveling to popular tourist spots and to give recommendations to small businesses who cater in the area of these tourist attractions. Initially, insights were sought related to times of day and days of the week, but as we will see this became problematic.

Methodology

MTA data for all of 2021 was pulled from the MTA website into a databse and then analyzed using a variety of data analytics tools including: SQL, SQLAlchemy, Python, Pandas, Matplotlib, and Seaborn.

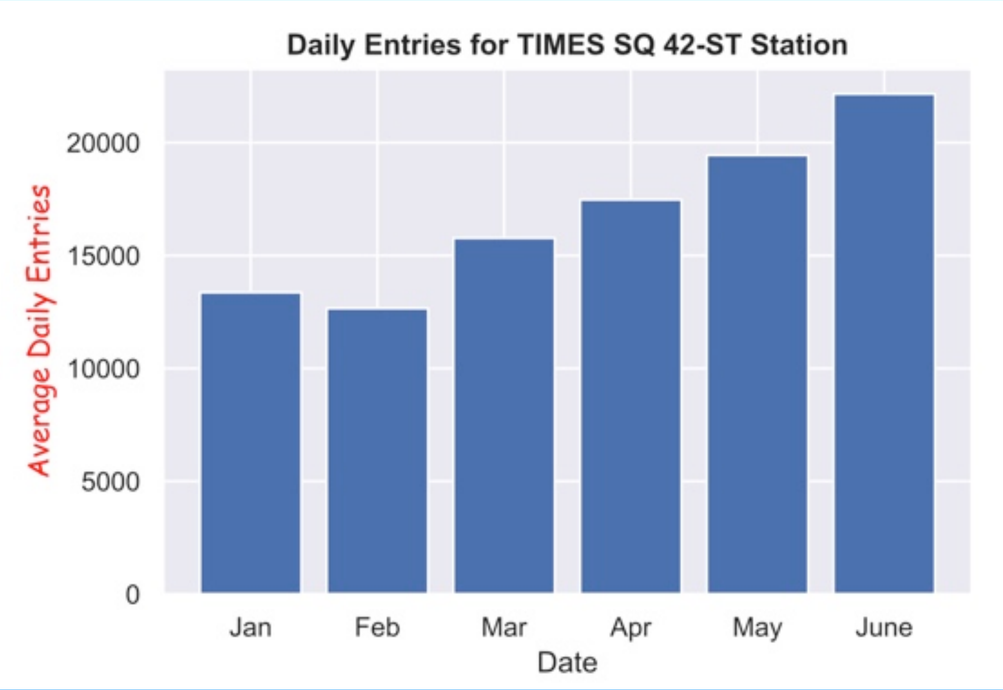
Data from 4 stations were chosen and analyzed based on their proximity to popular tourist attractions. The stations used in this study were:

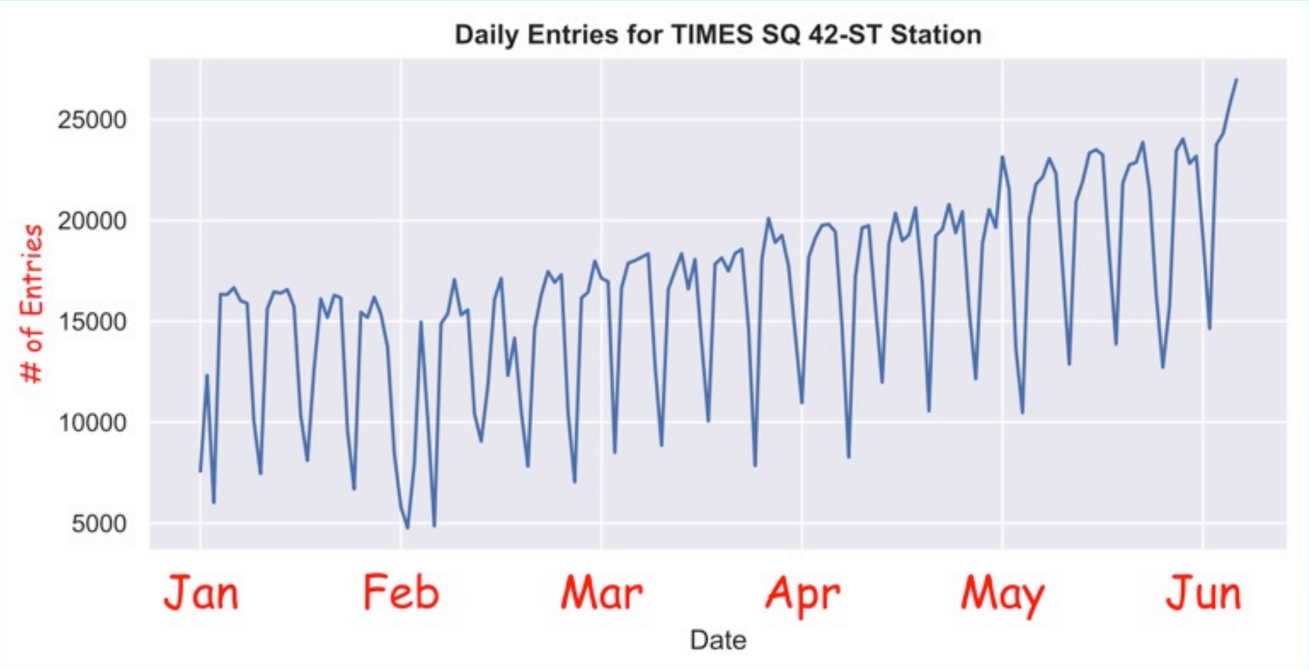
TIMES SQ-42 ST for it's proximity to Times Square.
81 ST-MUSEUM for it's proximity to the Metropolitan Museum of Art.
WHITEHALL S-FRY for it's proximity to the Statue of Liberty.
WORLD TRADE CTR for it's proximity to the new World Trade Center buildings and memorial.

The data was obtained using SQL and SQLAlchemy. Then the data was cleaned and analyzed using Pandas. Descriptive statistics were gathered using Pandas. Finally, the data from each station was plotted onto charts using Matplotlib and Seaborn.

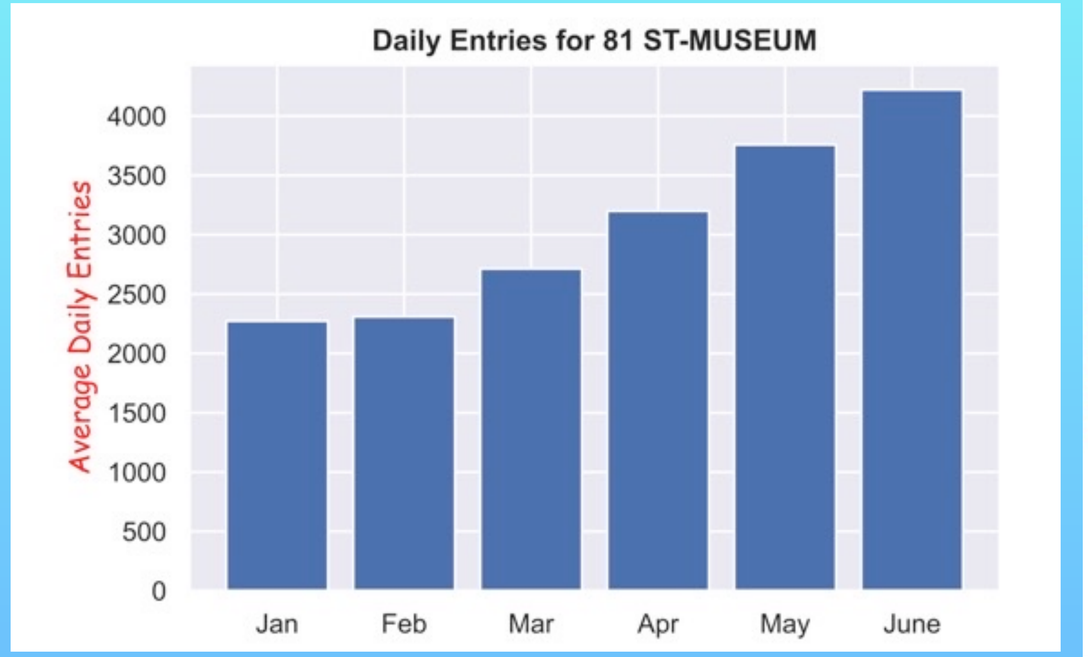
No statistical models were used for this project.

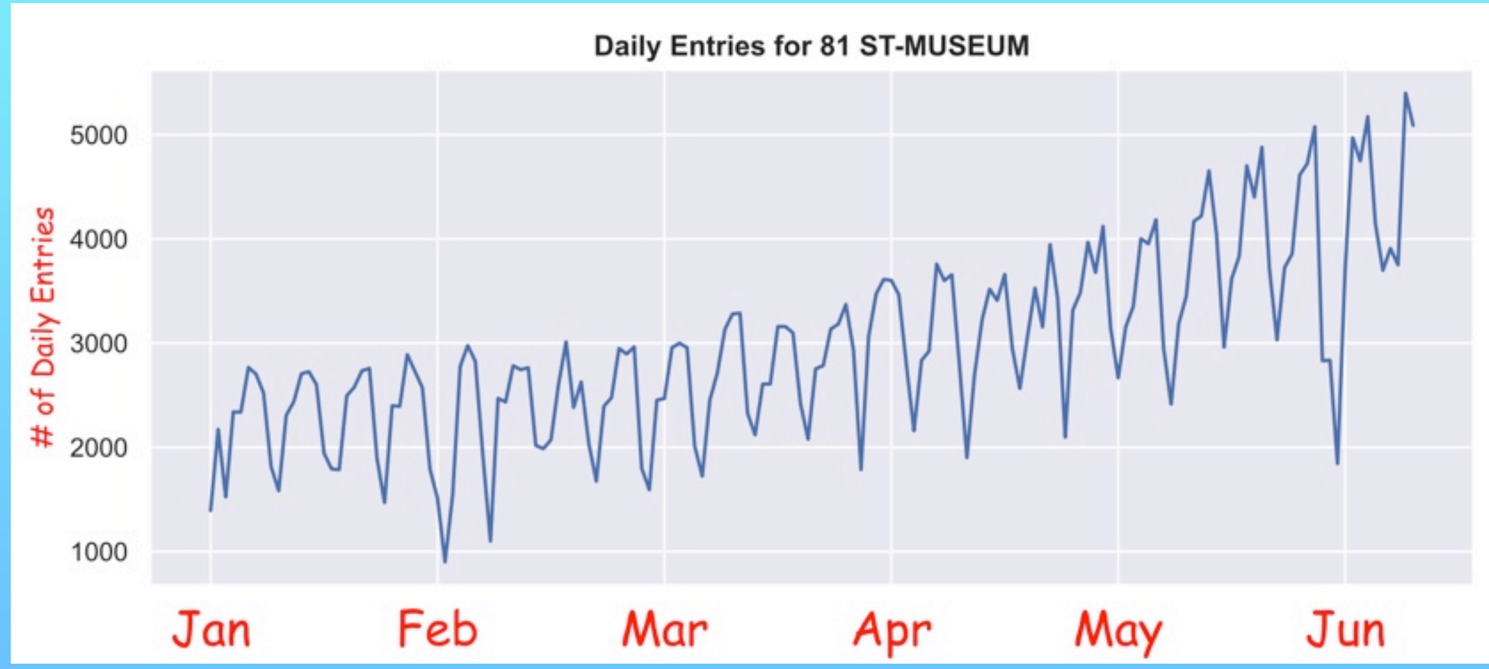
Times Square



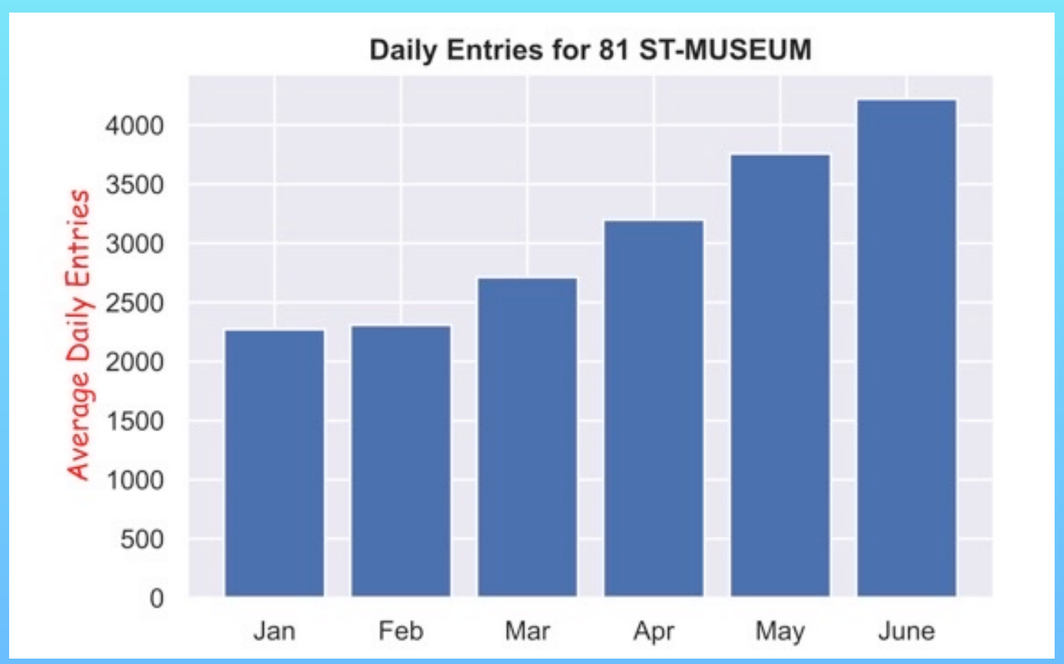


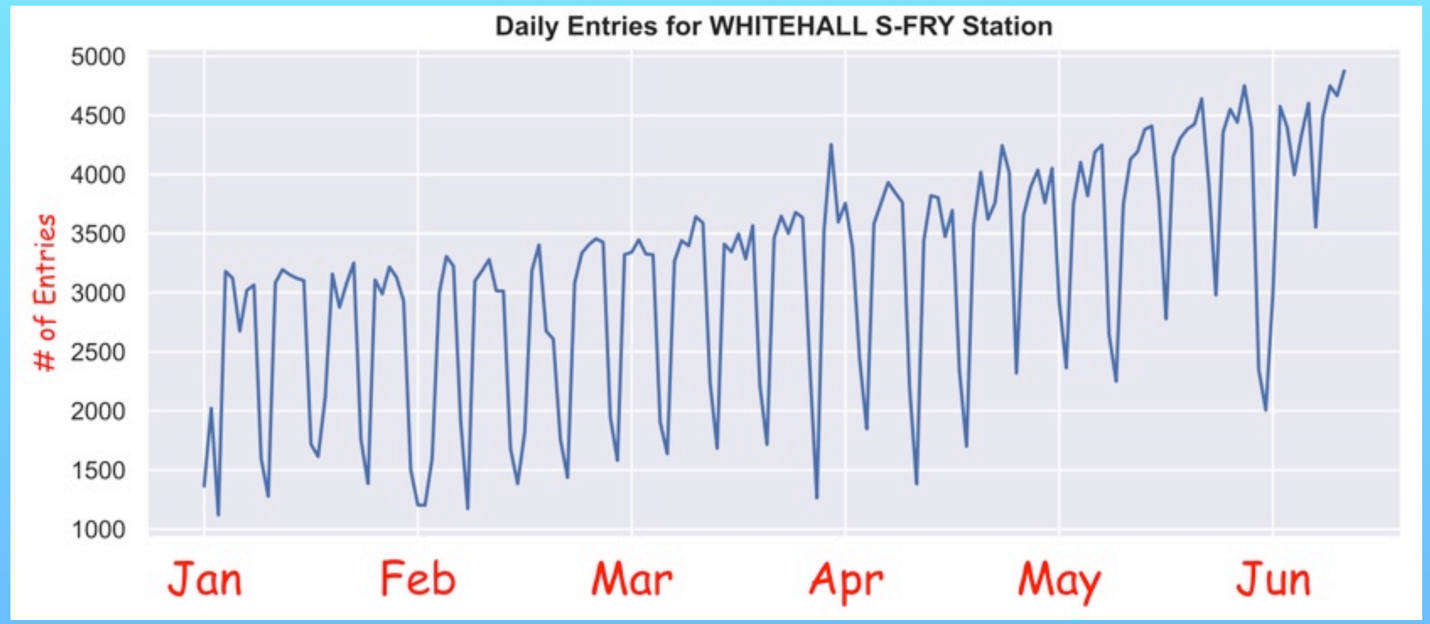
81 ST-MUSEUM


Daily Entries for 81 ST-MUSEUM
Average Daily Entries
4000
3500
3000
2500
2000
1500
1000
500
0
Jan
Feb
Mar
Apr
May
June

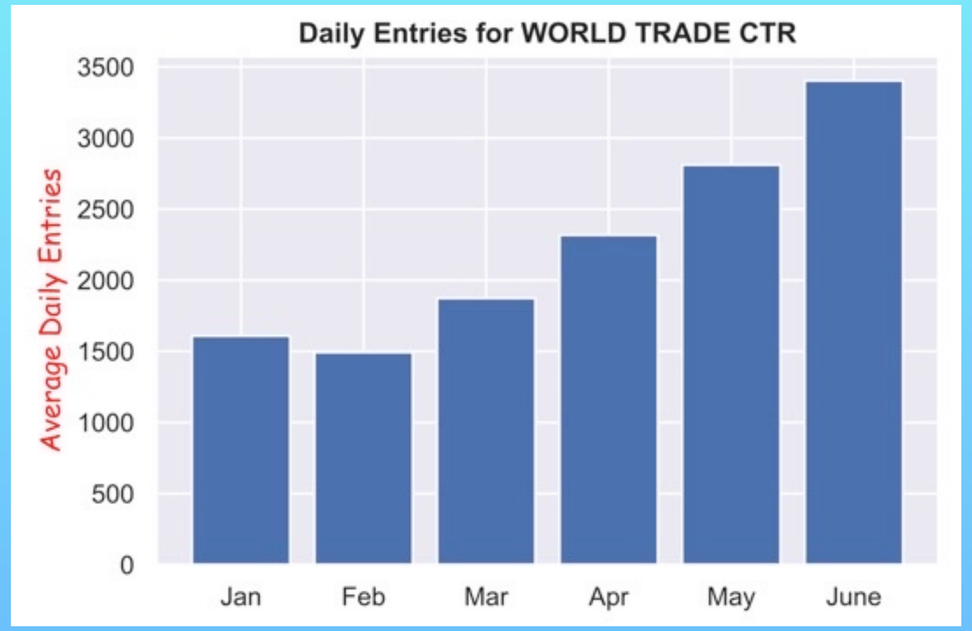

Daily Entries for 81 ST-MUSEUM
# of Daily Entries
5000
4000
3000
2000
1000
Jan
Feb
Mar
Apr
May
Jun

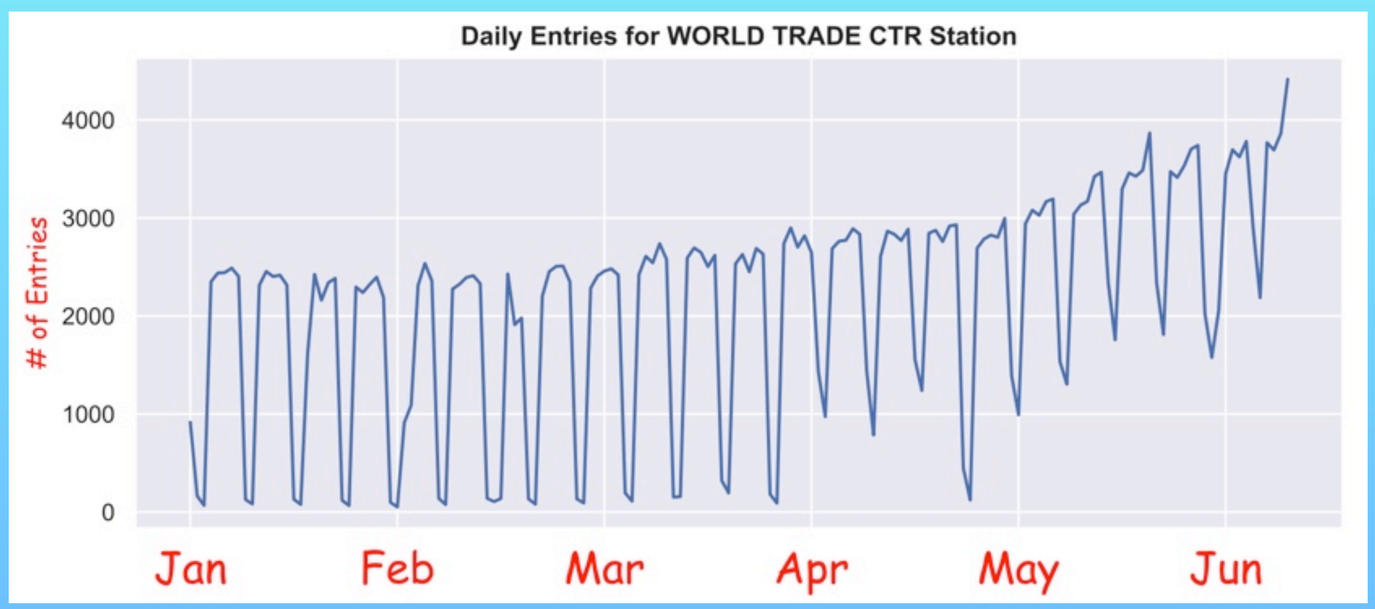
Whitehall S-Ferry





World Trade Center





Conclusions and Future Work

The main takeaway that relates to the initial goal of the project that was gleaned from the data is that MTA traffic is steadily increasing over the months of 2021. Unfortunately, I do not feel confident enough in the results to recommend actionable advice to businesses working at these tourist locations.

The most interesting finding is related to the periodicity of the daily average charts. This is perhaps a bold conjecture, but I believe the almost algorithmic looking periodicity observed in the data does not reflect a periodicity in the pattern of MTA users. I think it is far more likely that it reflects a periodicity in data entry. This hypothesis is based mainly on intuition, observing the days of the week of the cycles and the consistency of it, and comparing it to other data sets such as COVID-19 cases and deaths in NYC in early 2020.

Future Work- In order to gain actionable insights related to the movement pattern of tourists at various tourist destinations in NYC, other data sets are necessary. Initially the goal was to discern daily and even hourly trends in the data. After observing the unnatural periodicity in the data, I concluded this was impossible. Also, methods other than remote data analysis might be more effective for this purpose. Such as going physically to these locations to try to figure out the proportion of tourists vs other types of city goers.

The MTA data set was very helpful for learning skills related to data cleaning and plotting, but not for making assertions related to MTA users over hours or days. If I were to do another project with this data set, I would focus on issues that span a much longer time horizon and don’t depend on very accurate data.