أكاديمية سدايا
SDAIA Academy

Data Analysis of MTA Turnstile Data

Elyas Maghrabi

Analysis Goal

Would like to verify which stations get visited the most and whether or not certain factors(time of day, day of week, etc.) affect the number of riders taking the subway. By understanding what factors affect the number of MTA riders, also can figure out whether or not certain loction and underserved.

Data

Data from MTA involving the number of entries and exits at various turnstiles around each subway station in NYC

For geometric coordinate Open Data NYC, NYC Capital Planning Platform

Also, I join two data set , the MTA and the other dataset that have geometric coordinate.

Tools

I will use python libraries (pandas, Matplotlib, NumPy ,etc.), also for exploratory analysis and data cleaning used SQLite for local database and PostgreSQL for global database and server is google cloud.

Tableau.

Main Attribute Used

- Station: name of subway in which turnstile is located
- Date: month,day and year that turnstile data was recorded
- Time: hour that turnstile data was recorded

Derived Attributes Used

- The attribute Entries given by the MTA turnstile data was a cumlative number of entries at that specific turnstile. By cleaning the data and calculating a new attribute Entries Diff, now we have an incremental value to repersent the number of entries per station.

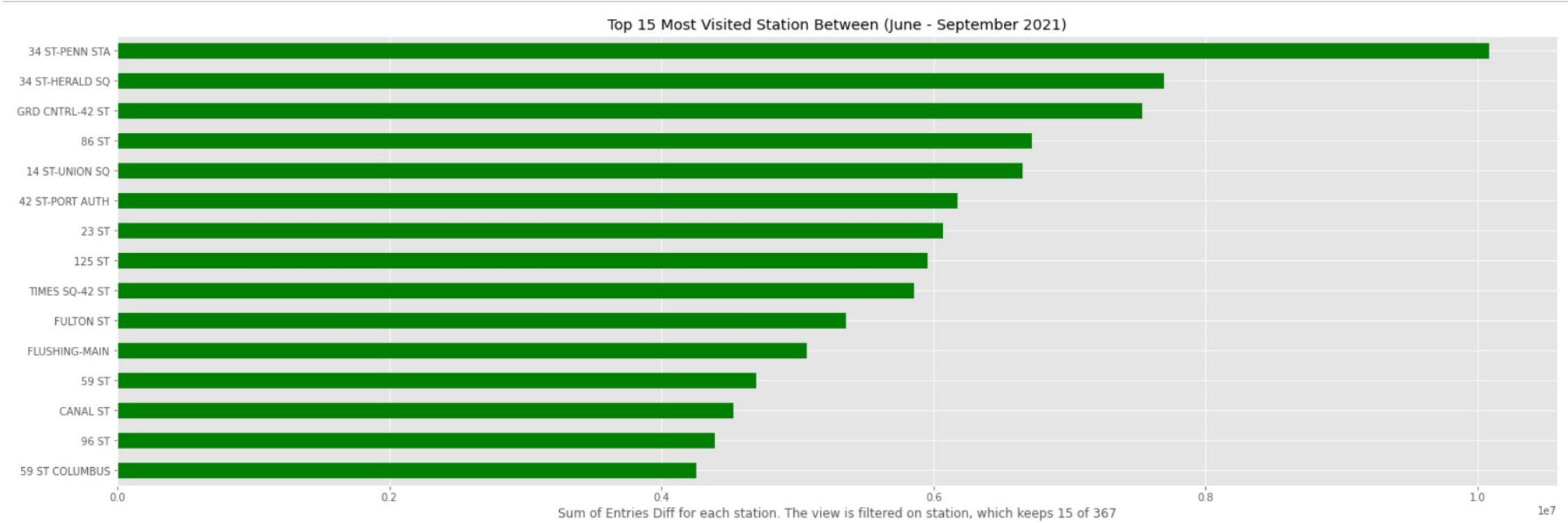
Questions

- Do certain subway station get visited more often than others? Where are these stations located?
- Are there more MTA riders during certain days of the week? (Workdays vs. weekends?)
- Are there more MTA riders during certain hours of the day? (Rush hour vs. late night?)

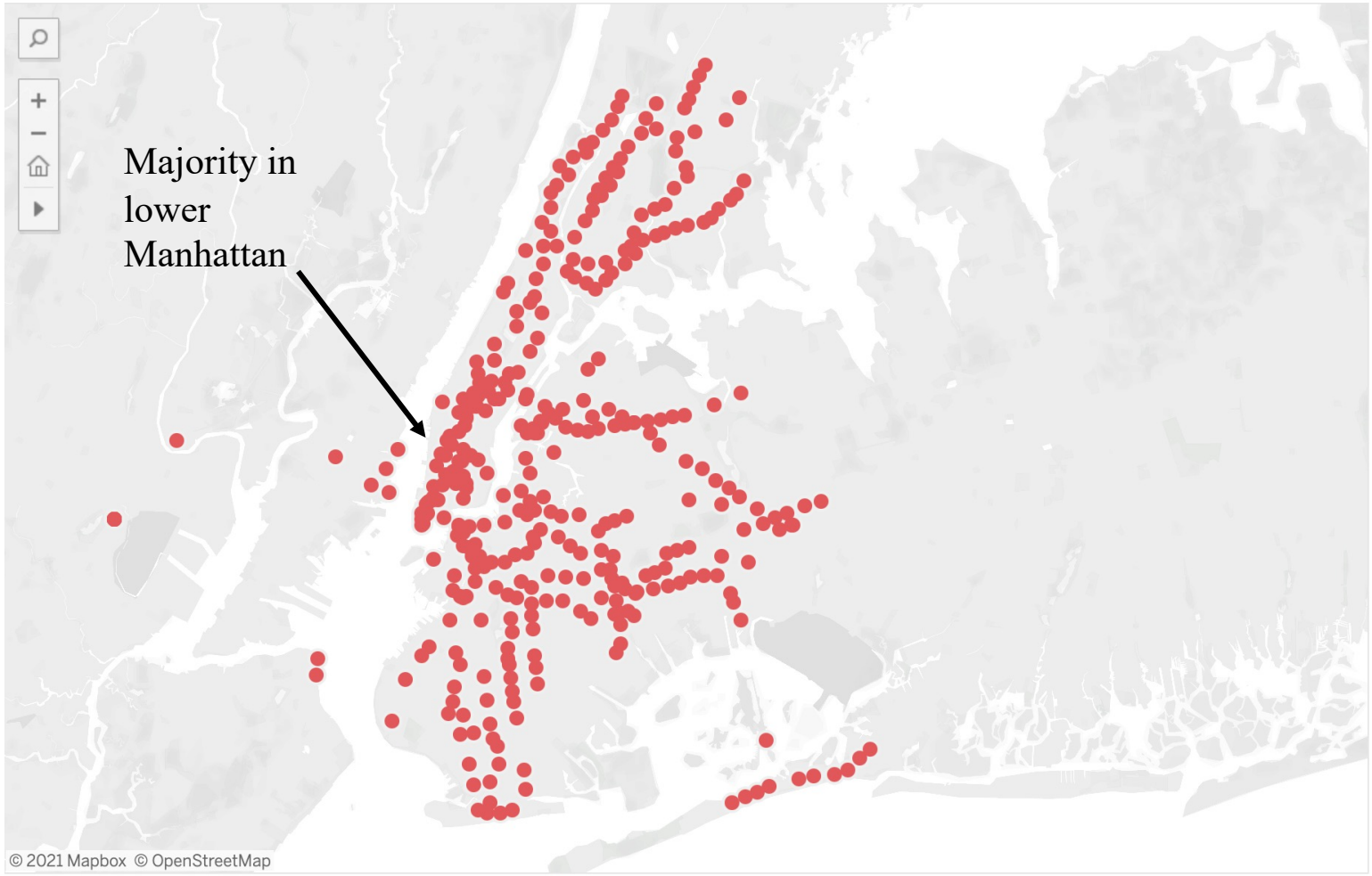
Attributes for Qusetion #1

- Do certain subway station get visited more often than others? Where are these stations located?
- Station: categorical
- Entries: ordinal(cumlative turnstile– dose not get rest dailly)
- Entries Dff: ordinal(calculated from Entries – incremental)
- Latitude: geographical
- Longitude: geographical

What certain station get visited more often than others?



Where are these stations located?



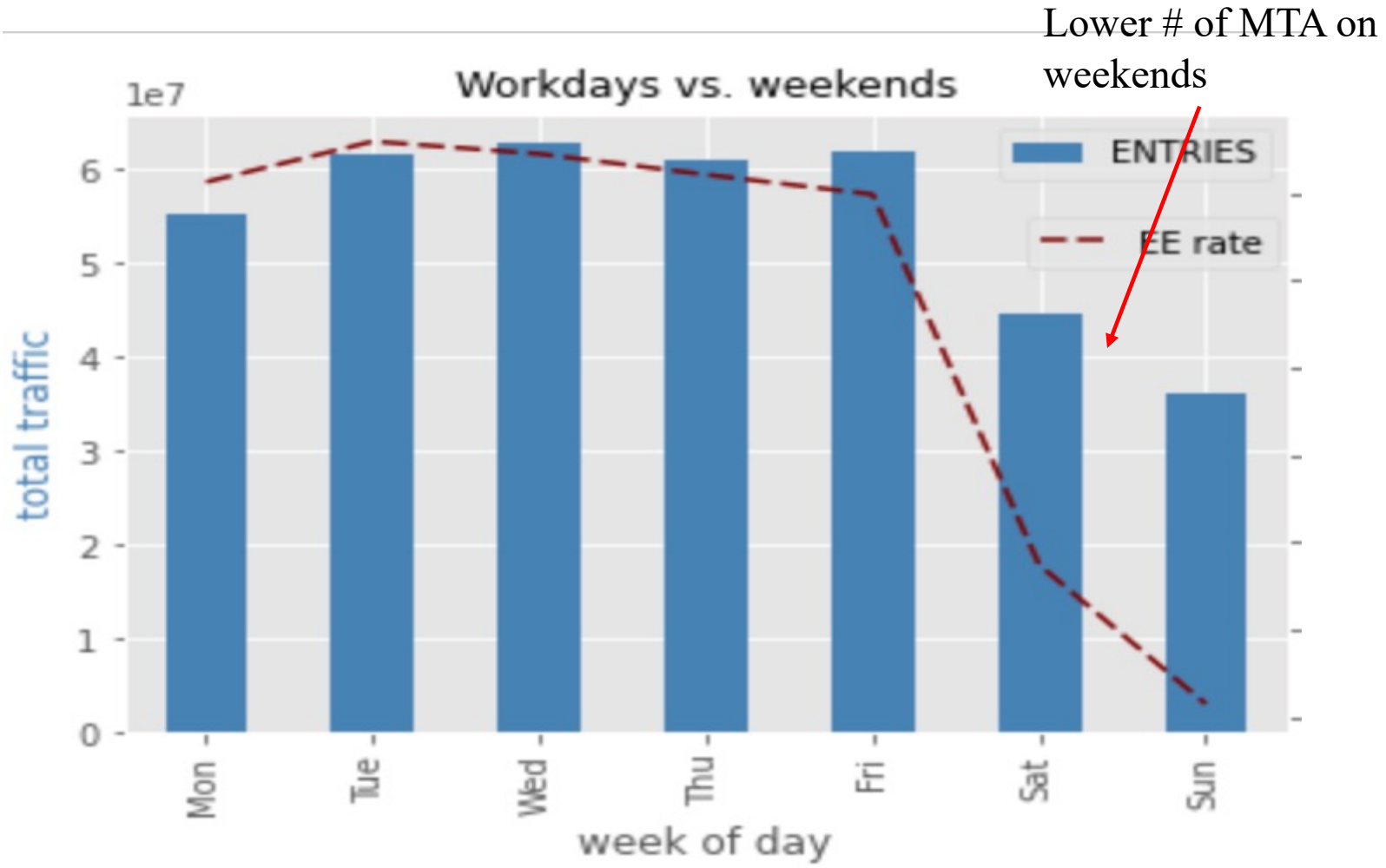
- As seen by the graph on left, the majority of MTA riders enter at subwway station located in lower Manhattan.the second greatest mumber of MTA riders appear to enter at subway stations Brooklyn

Attributes for Qusetion #2

- Are there more MTA riders during certain days of the week? (Workdays vs. weekends?)
- Date: ordinal
- Entries: ordinal
- Entries Dff: ordinal(calculated from Entries)

Are there more MTA riders during certain days of the week? (Workdays vs. weekends?)

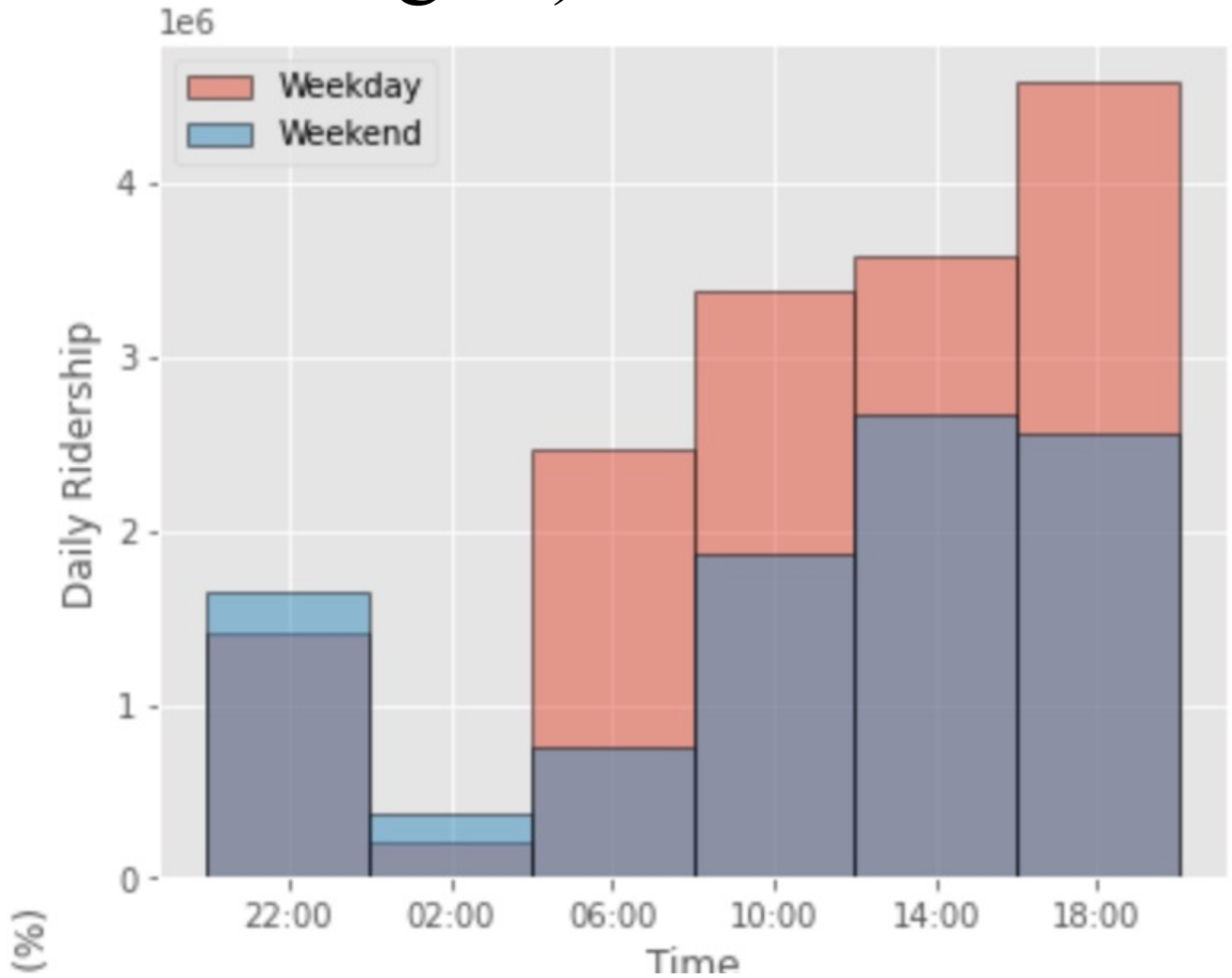
According to the graph to left. The number of MTA riders is greater during the weekdays then during the weekend



Attributes for Qusetion #3

- Are there more MTA riders during certain hours of the day? (Rush hour vs. late night?)
- Entries : ordinal
- Time: ordinal
- Entries Dff: ordinal(calculated from Entries)

Are there more MTA riders during certain hours of the day? (Rush hour vs. late night?)



Conclusion

- We can notice the subway stations that are visited the most are subway stations located in lower Manhattan. This might be because Manhattan is where many people go to work, to shop and to sight see.
- It was also verified by the data that the number of MTA riders are affected by the day of the week and the time of the day.
- During weekdays when people either have to go to work or to school, there are more people taking the subway.

Recommendation

Taxi company should focus on the hours of workdays. Start (9am/10am) on stations with high concentrations of busy places in Manhattan — 34 ST-PENN STA, 34 ST-HERALD SQ and GRD CNTRL-42 ST .in teams at street subway entrances will take advantage of tourism pedestrian traffic that may not be related to the subway. During lulls in the workdays or after hours, they can focus on stations with high ridership during those time blocks to make use of all daytime hours.

Thank You