

EXPLORATORY DATA ANALYSIS

MTA Turnstile Analysis

Marjan Rezvani





What can we do for citibike?



Target market:

- Commuters using the Metropolitan Transportation Authority(MTA system)
- Highly likely to use citibike between subway stations and the office



Analytical Approach

Define main criteria



HIGH EXIT RATE IN THE
MORNING



HIGH ENTRY RATE IN THE
AFTERNOON OR EVENING

MTA Data:

- There are 3362235 rows and 11 columns including:

C/A, UNIT, SCP, STATION, LINENAME, DIVISION, DATE, TIME, DESC, ENTRIES, EXITS

Citibike Data:

- There are 3072478 rows and 13 columns including:

ride_id, rideable_type, started_at , ended_at, start_station_name,
end_station_name, start_lat, start_lng ,end_lat, end_lng, member_casual

The Methods and Tools:

Data ingestion and storage:

- Pandas
- SQLite

Data cleaning and manipulation:

- Python Libraries: Pandas, NumPy

Presentation tools:

- Matplotlib
- Seaborn

Data Analysis:

- Analyzed citibike data for September 2021 and some other random months
- Analyzed MTA turnstile data from September 2021 to end of December 2021
- Total amount of commuter traffic by station
- Daily amount of commuter traffic by station
- Morning hours traffic rate of exits from 7 to 10
- Afternoon hours traffic rate of entries from 4 to 7
- Visualize Data Analysis

Outcome:

- Post advertisements in high traffic MTA stations
- Encourage purchases through push notifications with discounts for citibike stations at peak hours

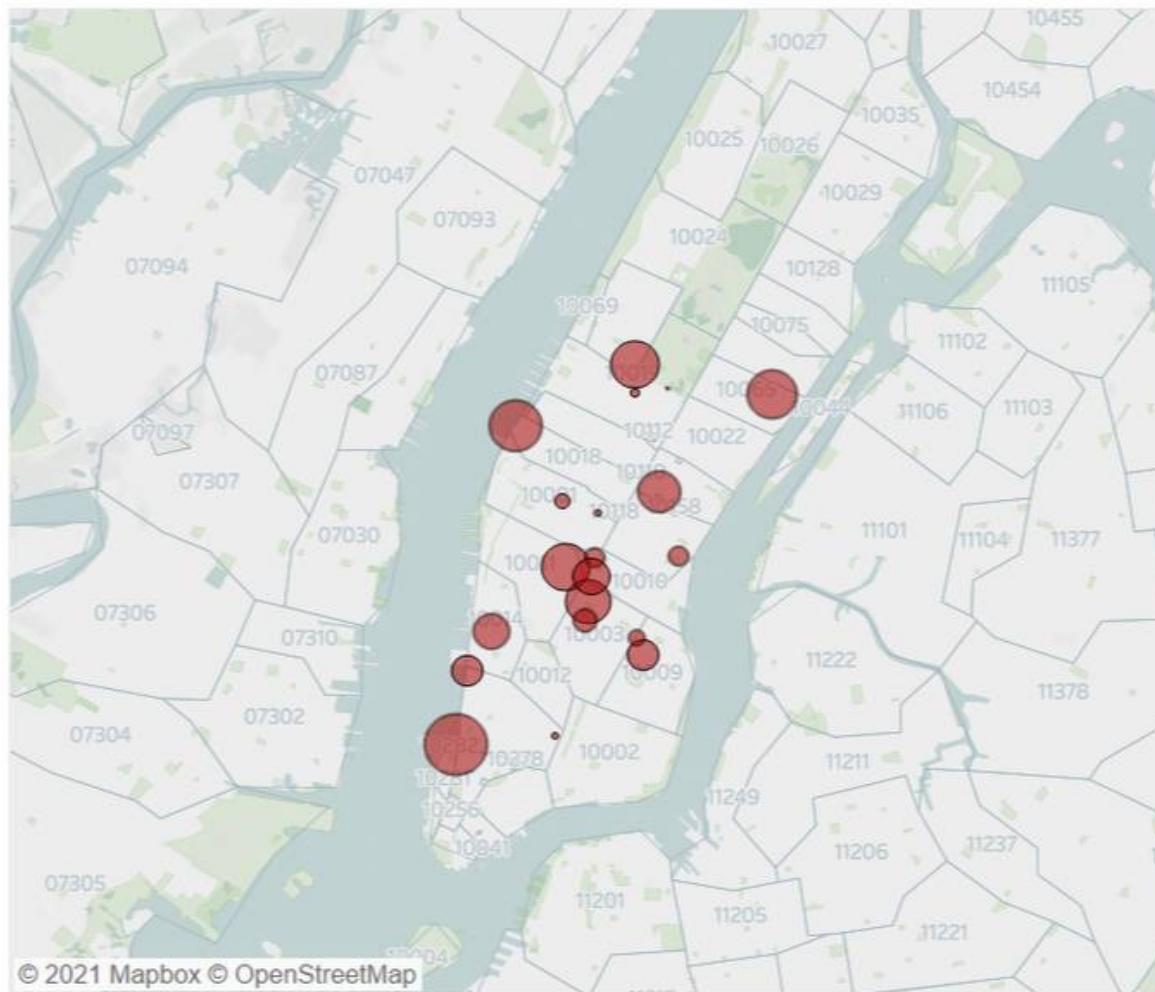
Problem Statement:

- What are the busiest stations?
- What are the best time to prompt app users to get a citibike membership?

Results and Analyses

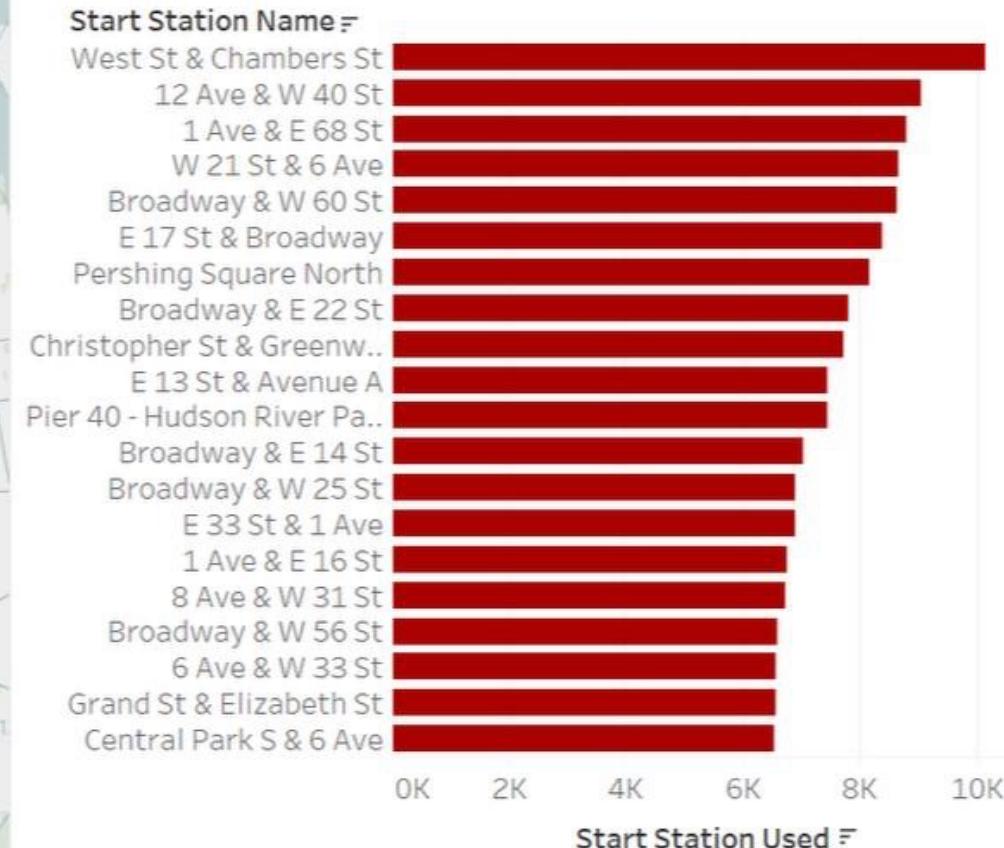
20 most active citibike start stations in September 2021

20 Most Active Start Stations (Map)

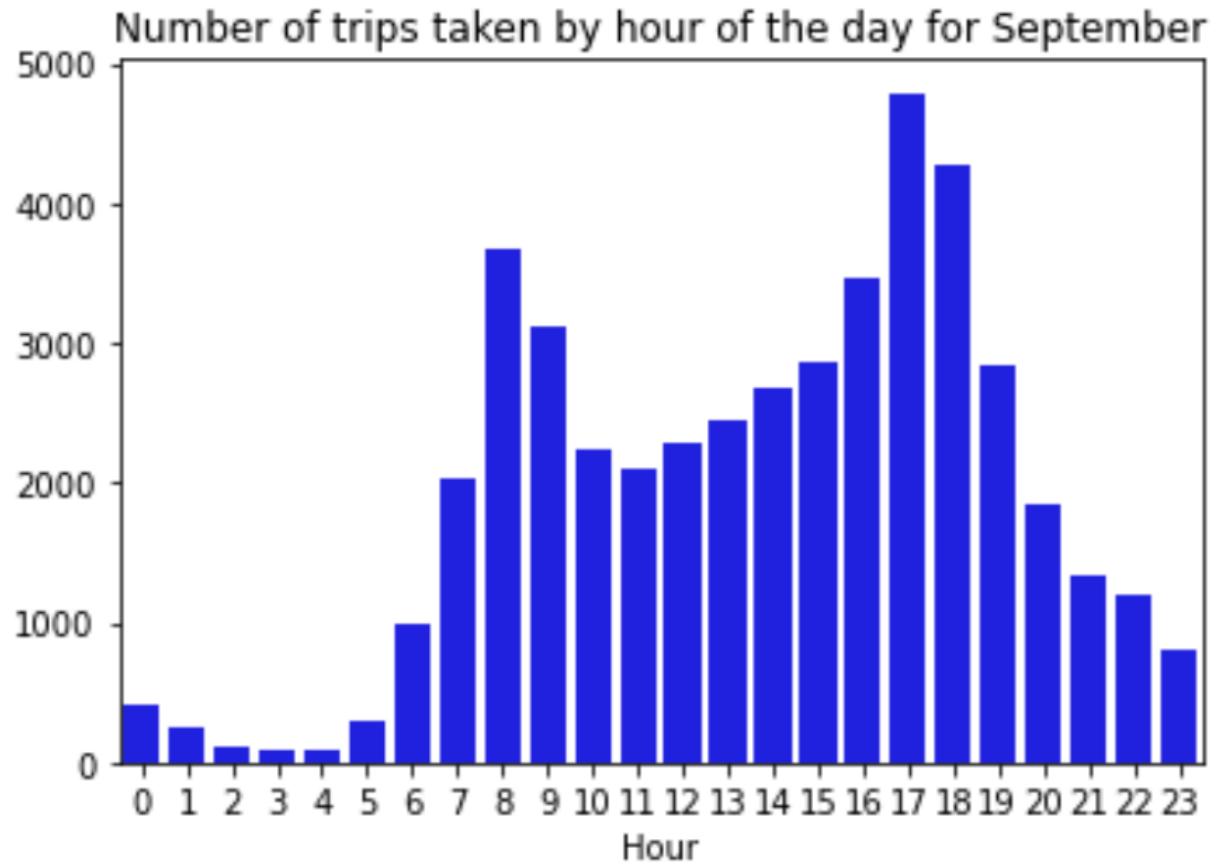


Day of Starttime
Sep 1, 2021 D Oct 1, 2021

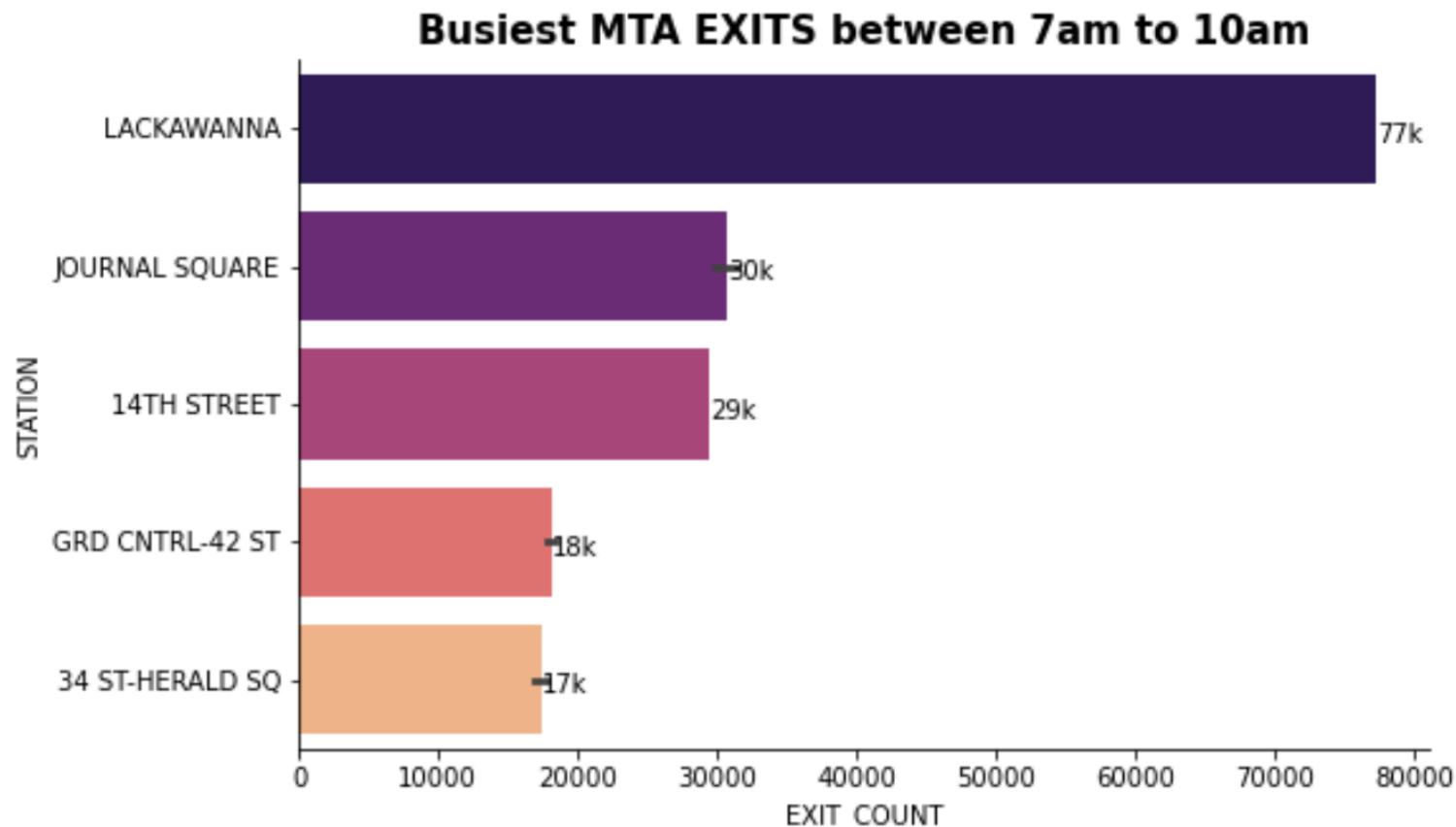
20 Most Active Start Stations (Bar)



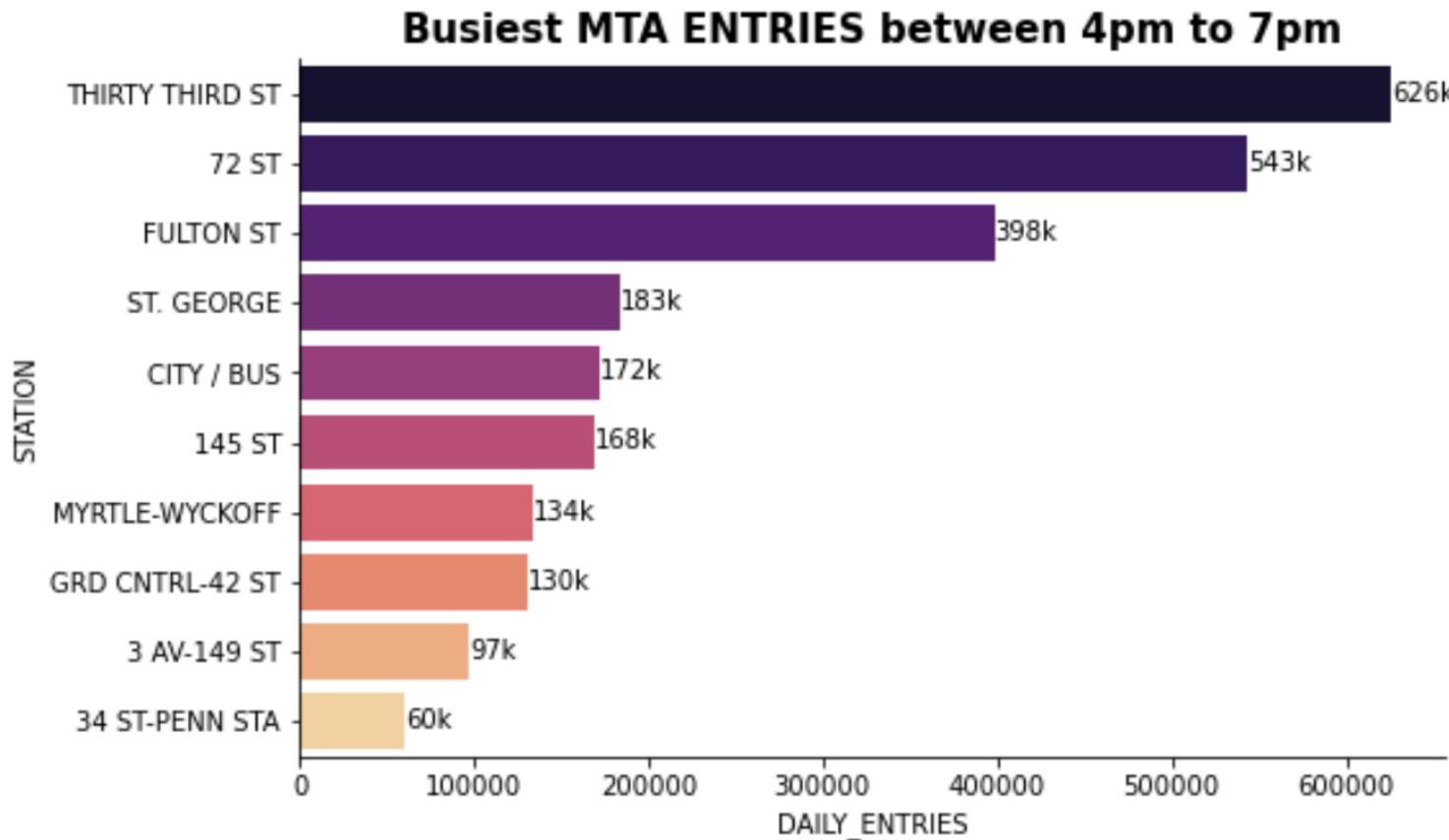
The number of trips taken by hour of the day in 'citibike.csv'



Busiest Exit MTA Commuters In The Morning



Busiest Entries of MTA Stations In The Afternoon



Conclusion:

1

Target the commuters of the busiest ten stations with posted advertisements encouraging citibike app downloads.

2

Push notifications with citibike discounts should be sent out between 7 to 9 in the morning and 4 to 6 in the evening.

3

Develop partnerships with citibike stations in areas of high commuter density and low citibike coverage.

Future Works:

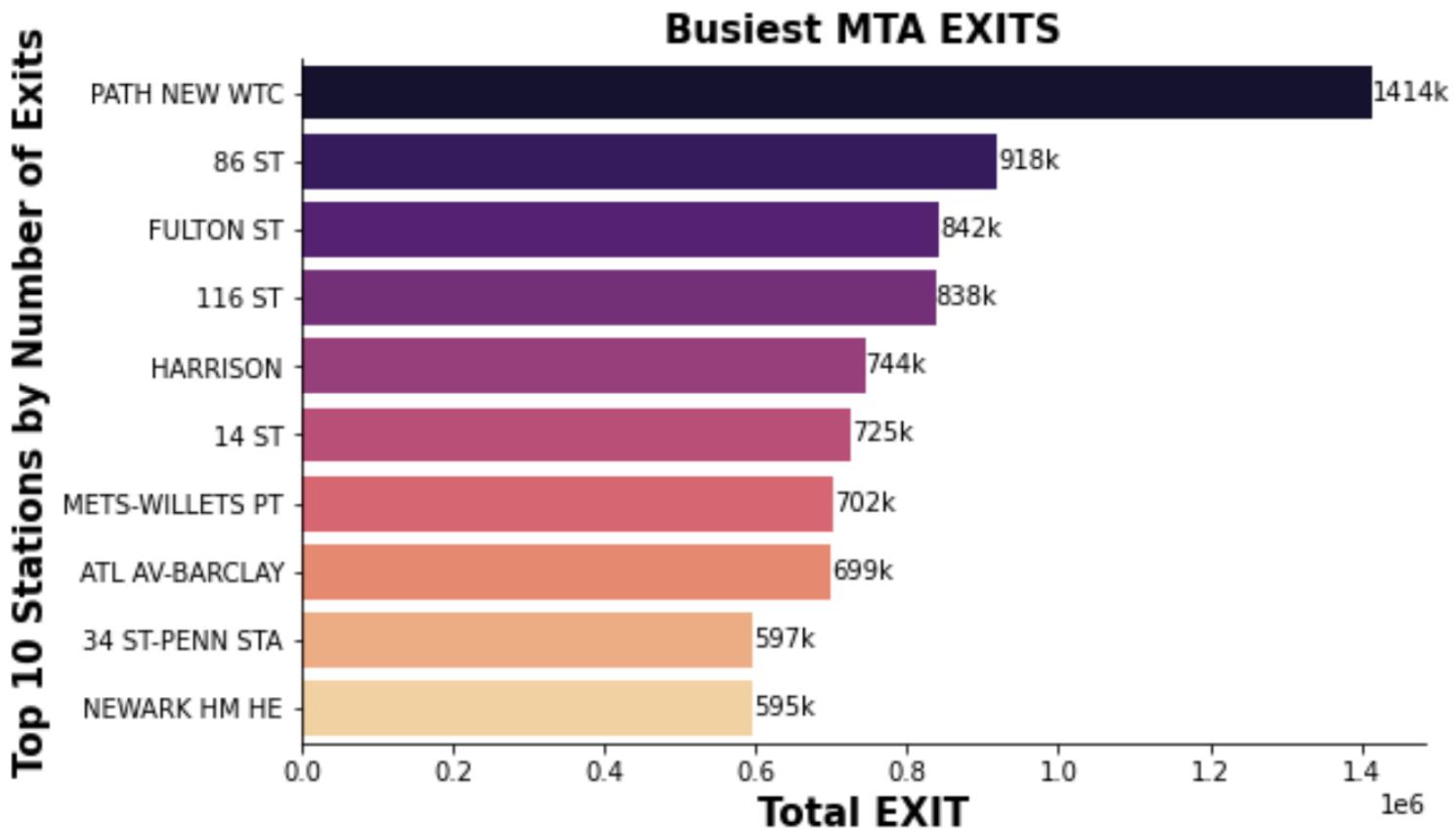


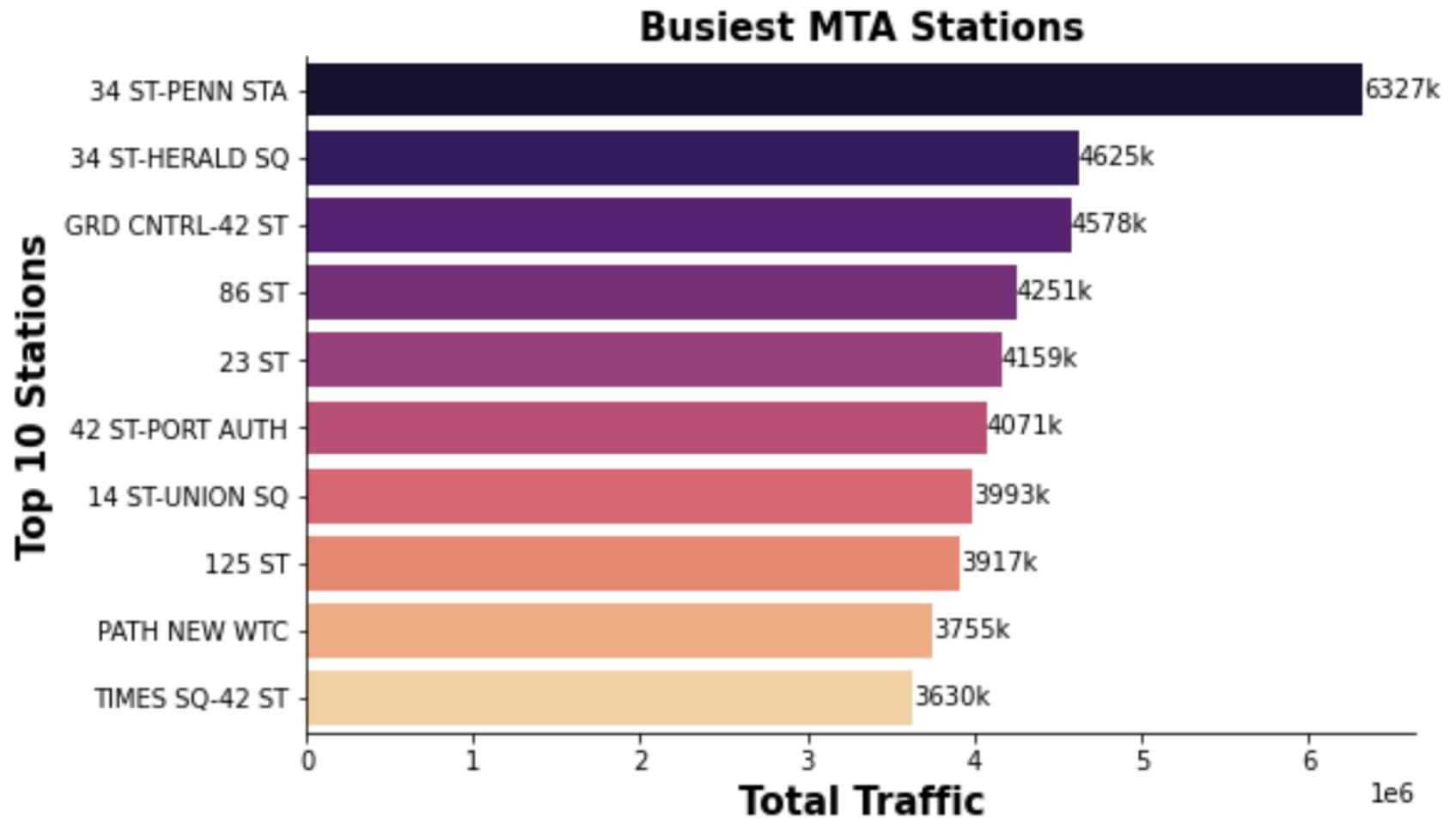
FINDING THE NEAREST CITIBIKE STATION
AND SEE IF IT'S FREQUENTLY LOW ON
BIKES OR ALL THE BIKES ARE GONE DURING
THAT TIME OF DAY.

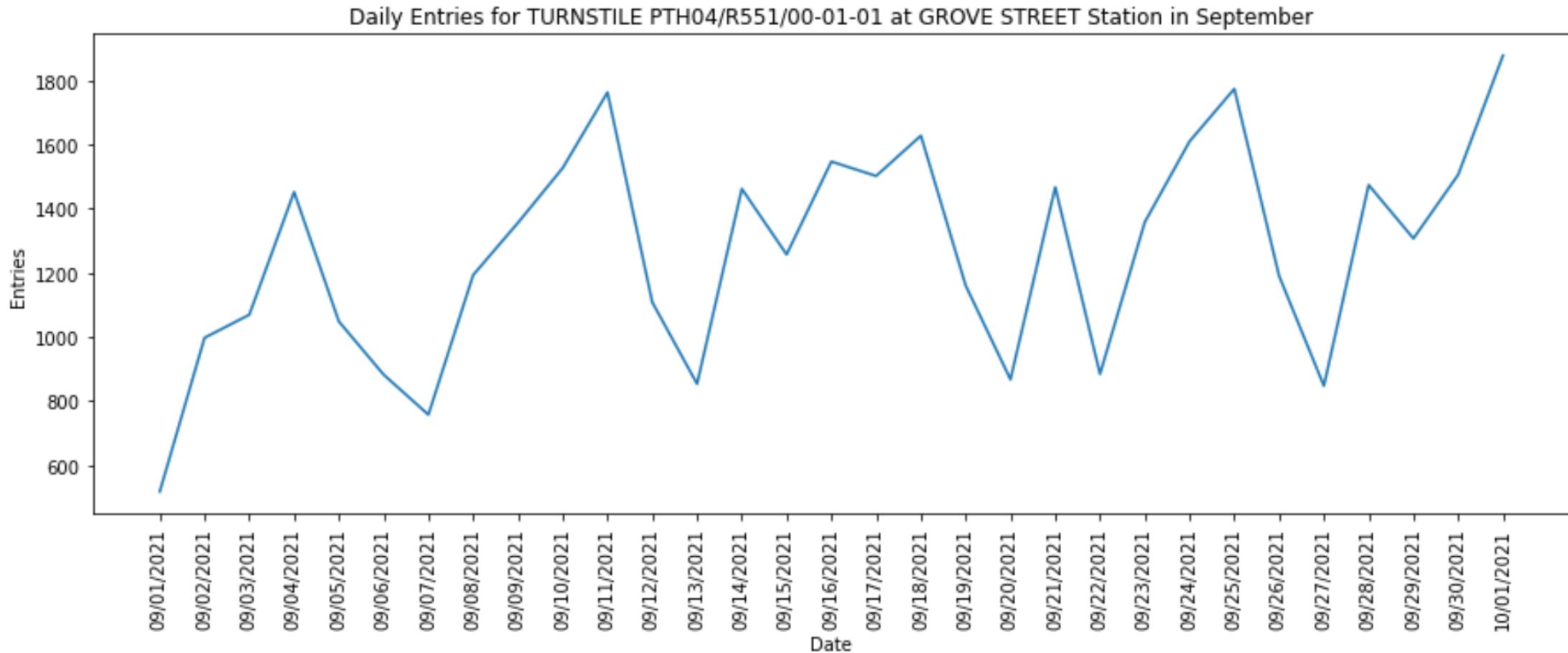


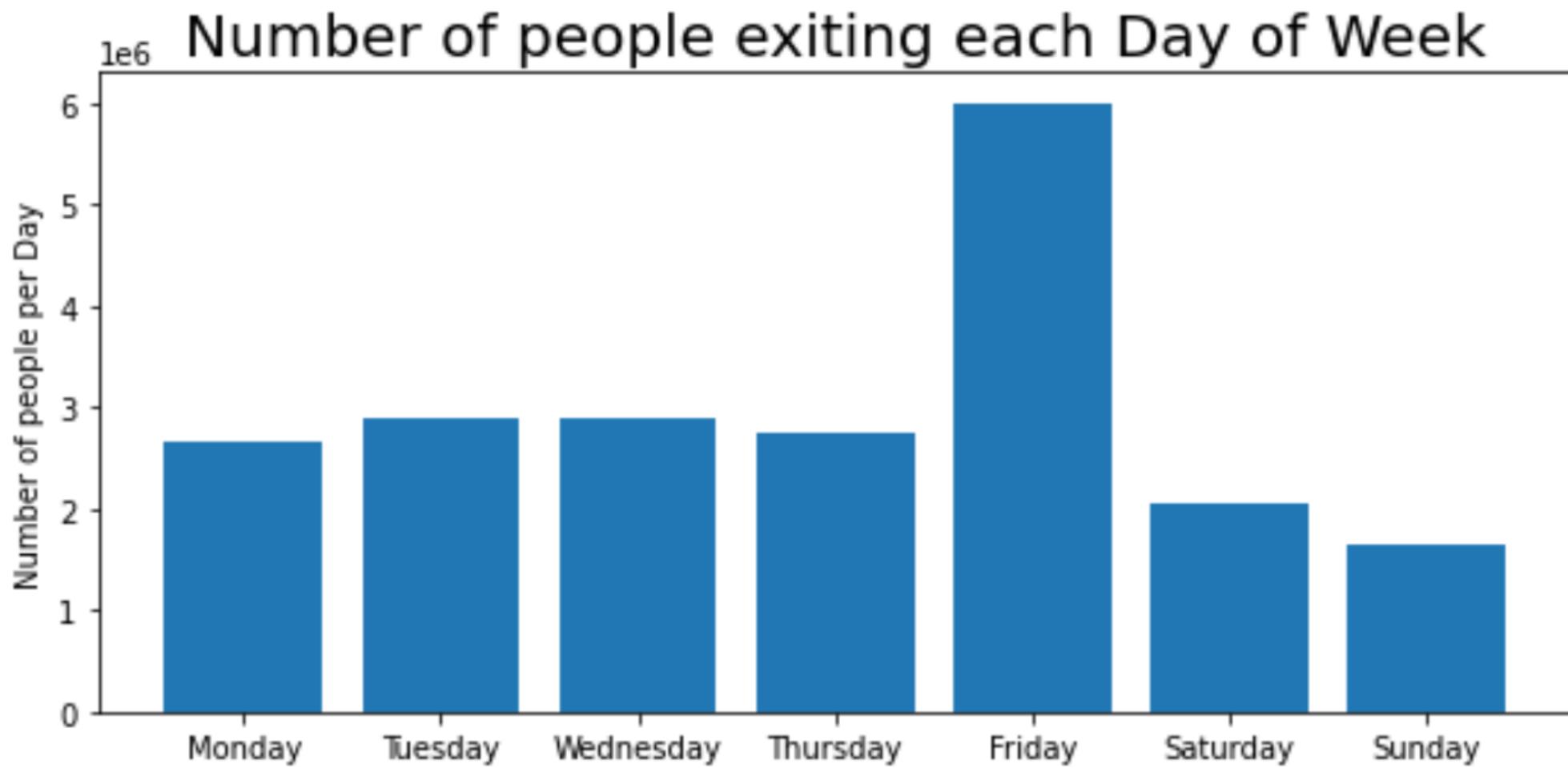
INCORPORATE DEMOGRAPHIC AND
GEOGRAPHICAL DATA TO DETERMINE THE
MAKEUP OF THE COMMUTERS, ANY
GEOGRAPHICAL CORRELATIONS ETC.

Appendix:









Questions