MTA Traffic Analysis for Coffee Corp.

By Afnan Alsirhani

Introduction :

Data Science Solutions (DSS) is a company specialized in Data Analytics and Data Science. We offer breakthrough and effective solutions for clients to help them achieve their own goals through the utilization of data to draw better decision making and solutions.

Problem Statement:

- Coffee Corp. is a small business that aims to expand their coffee shop branches in the train stations at New York City. in this project, DSS will analyze the turnstiles data for each station to find the busiest stations. Hence, recommending the best location for the new branches so that Coffee Corp. can achieve their goals to expand their coffee business.

The analysis is intended to answer two questions:

1- What are the top 5 busiest stations in NYC?

2- What are the busiest days in NYC stations?

TOOLS:

GIT-GitHub

SQLite- SQLalchmey

Python-Jupiter

Libraries:

Pandas
Matplotlib
Seaborn
plotly.express
plotly.graph_objects

Tableau

Dataset's description:

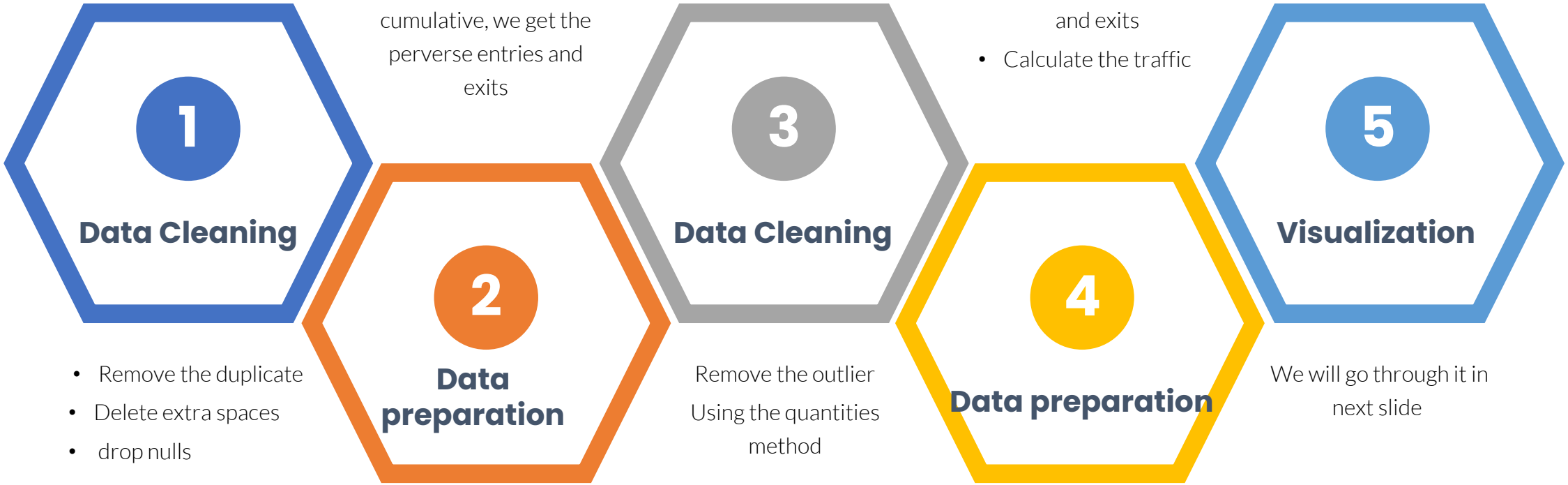
- Metropolitan Transportation Authority (MTA) Turnstile dataset exported from http://web.mta.info/developers/turnstile.html. The NYC MTA Turnstile data has a cumulative number of entries and exits for each station in different dates and time.
- Population of NYC (Source) : https://www.census.gov/quickfacts/newyorkcitynewyork,
- Stations lookup (connecting each station with locations in map), used in Tableau.

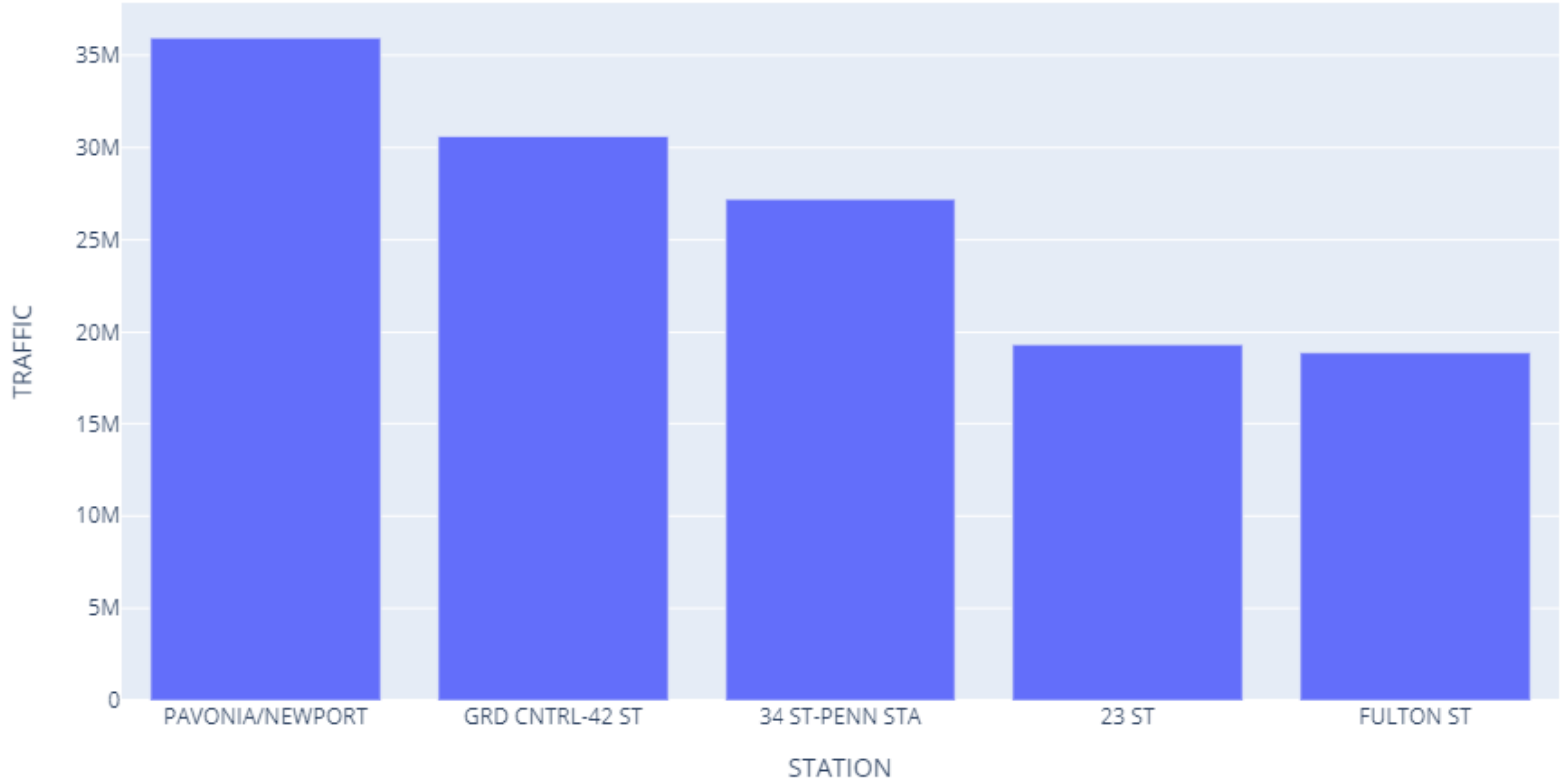
Methodology:

In this EDA, started to get the data from 1/8/2019 to 30/11/2019.

```
(3708610 row, 11 columns)
```

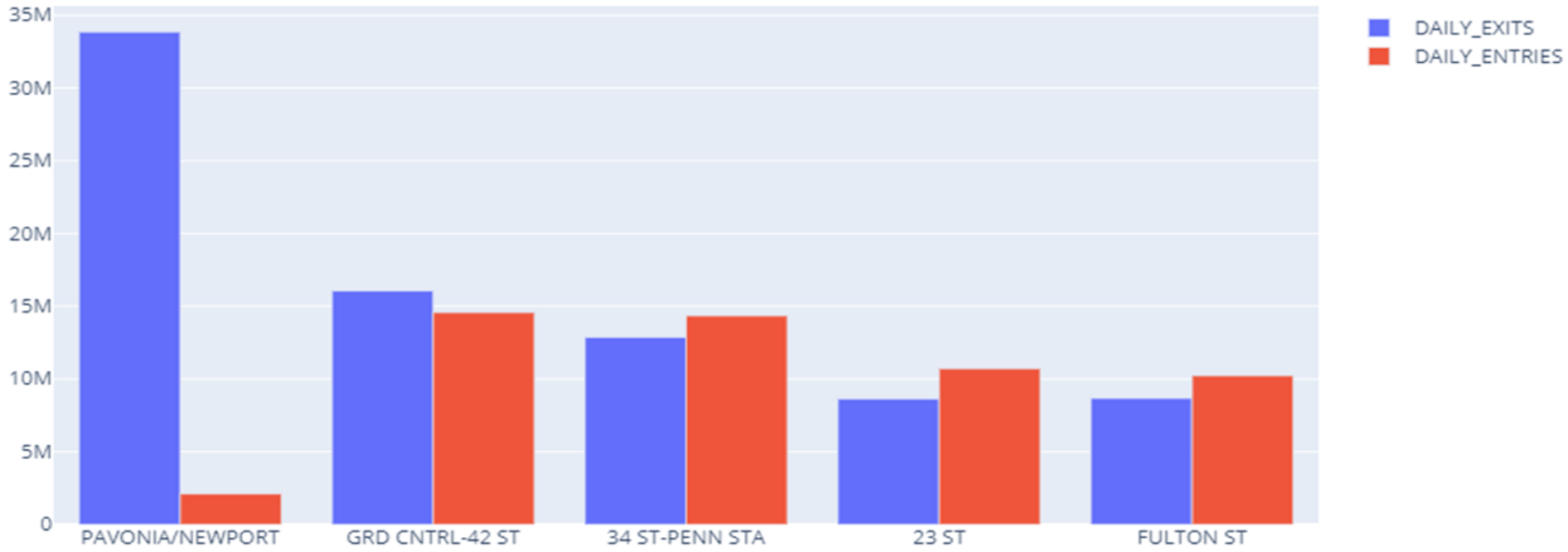
PROJECT WORKFLOW





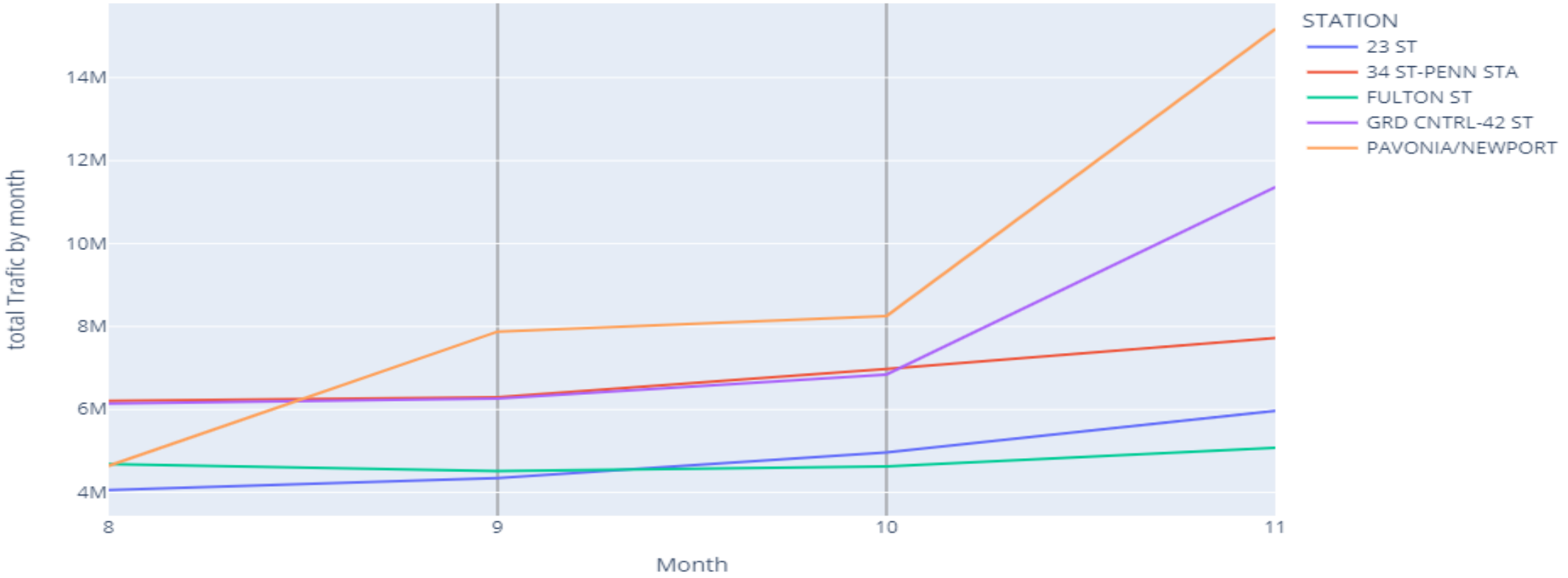
The graph represents the top 5 stations per Traffic,(the Traffic is the sum of the entries and exits for each station)

So, we needed to investigate more the top 5 stations

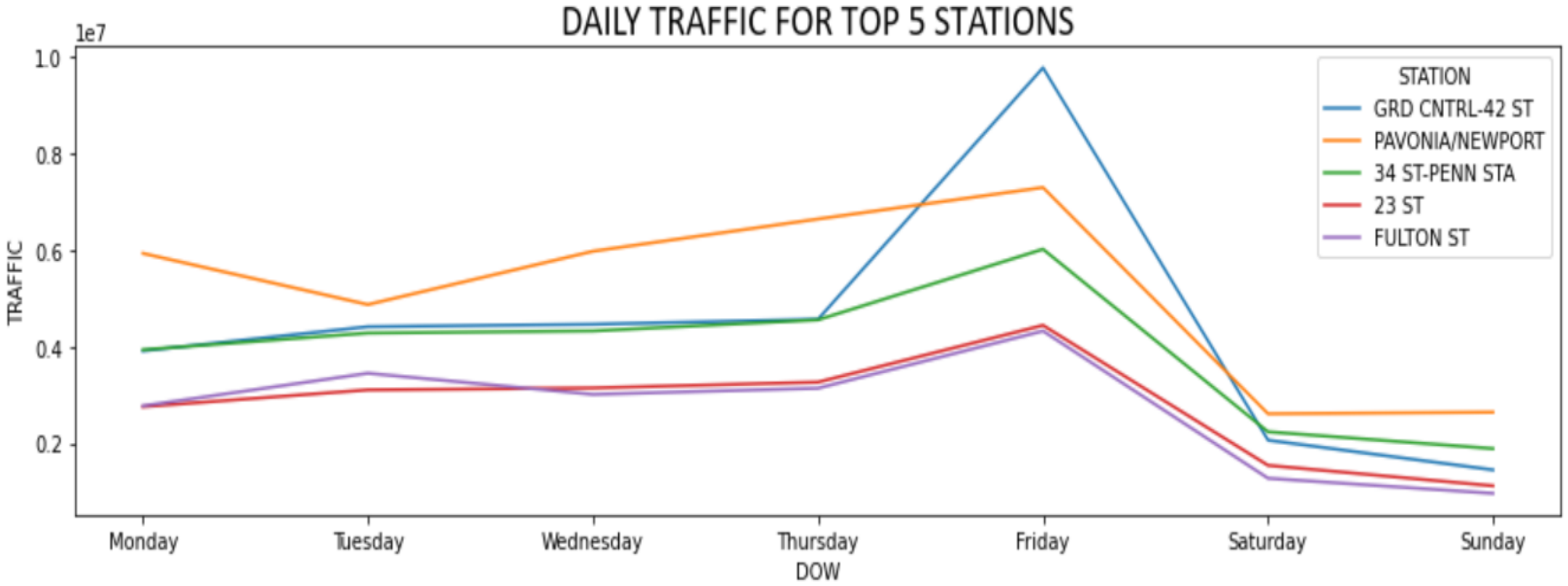


we can see in graph above the PAVONIA/NEWPORT station had the highest exits followed by GRD CNTRL-42 ST

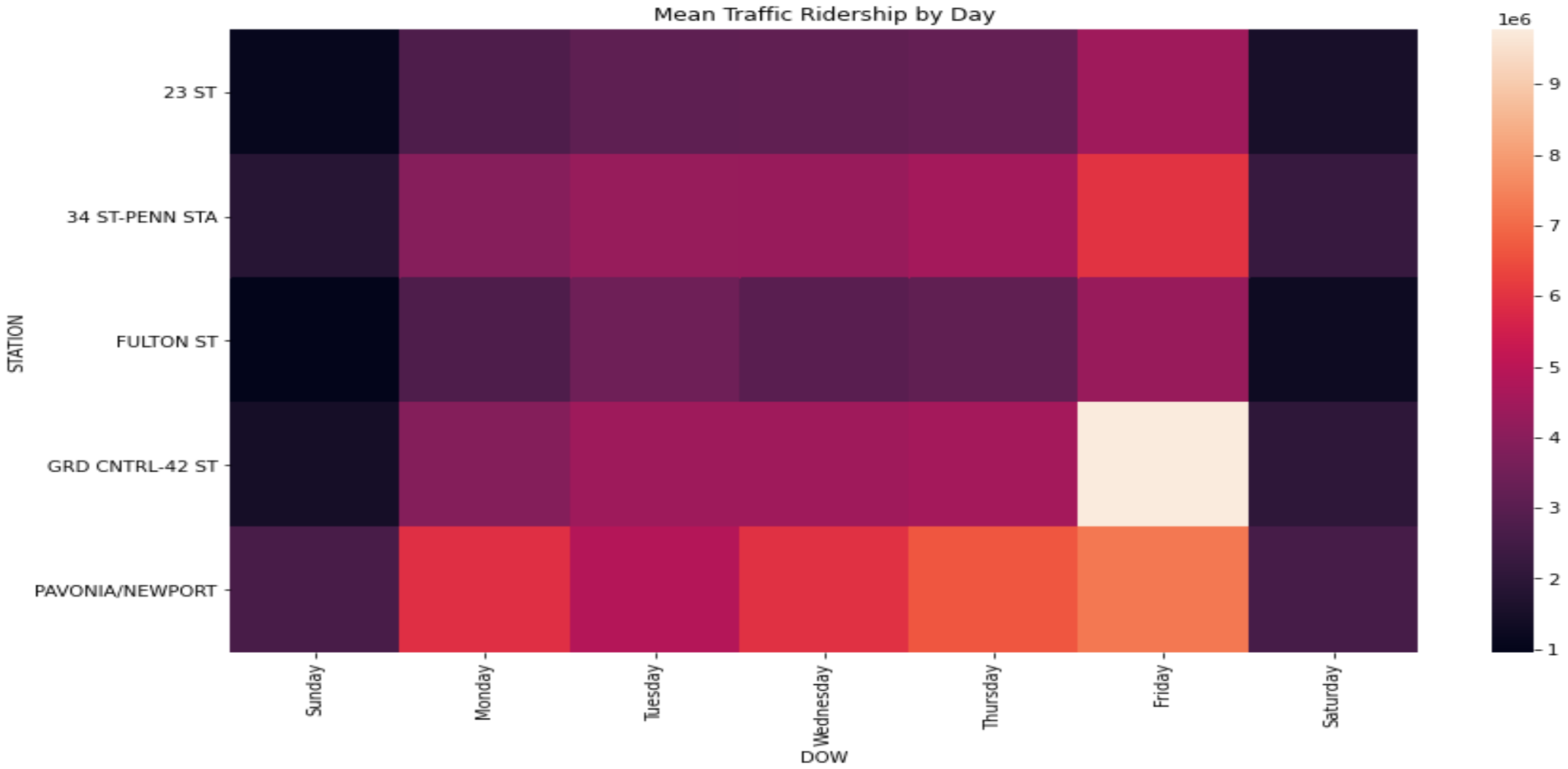
so will go deeper to understand the busiest day of the for these stations



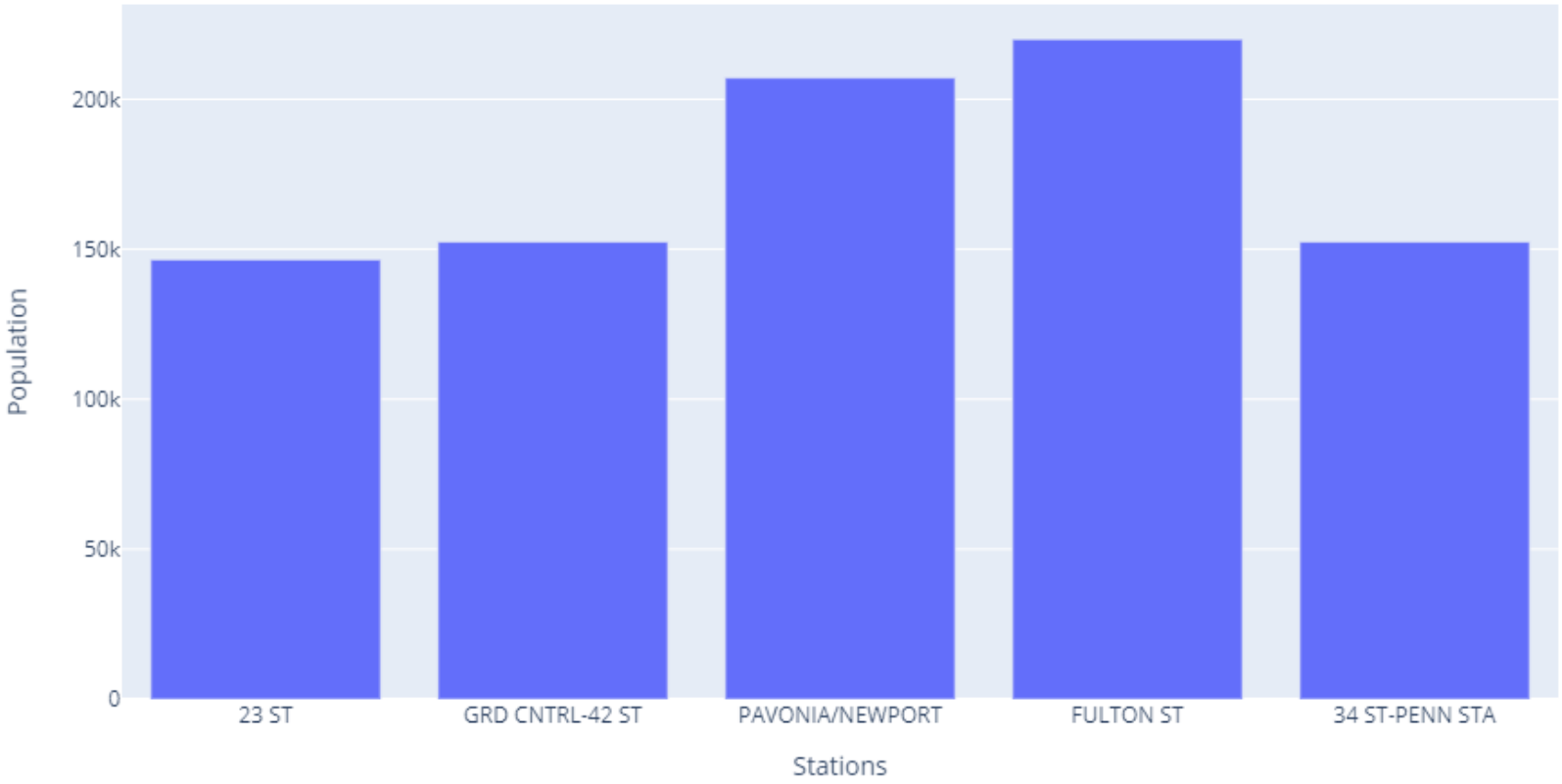
The line chart represent the sum of traffic for each station by months, notice the increases in November due to the fact that is a holiday month.



we can see in the line chart above the traffic for each of the top 5 stations, GRD CNTRL-42 ST station had the highest traffic in Friday.

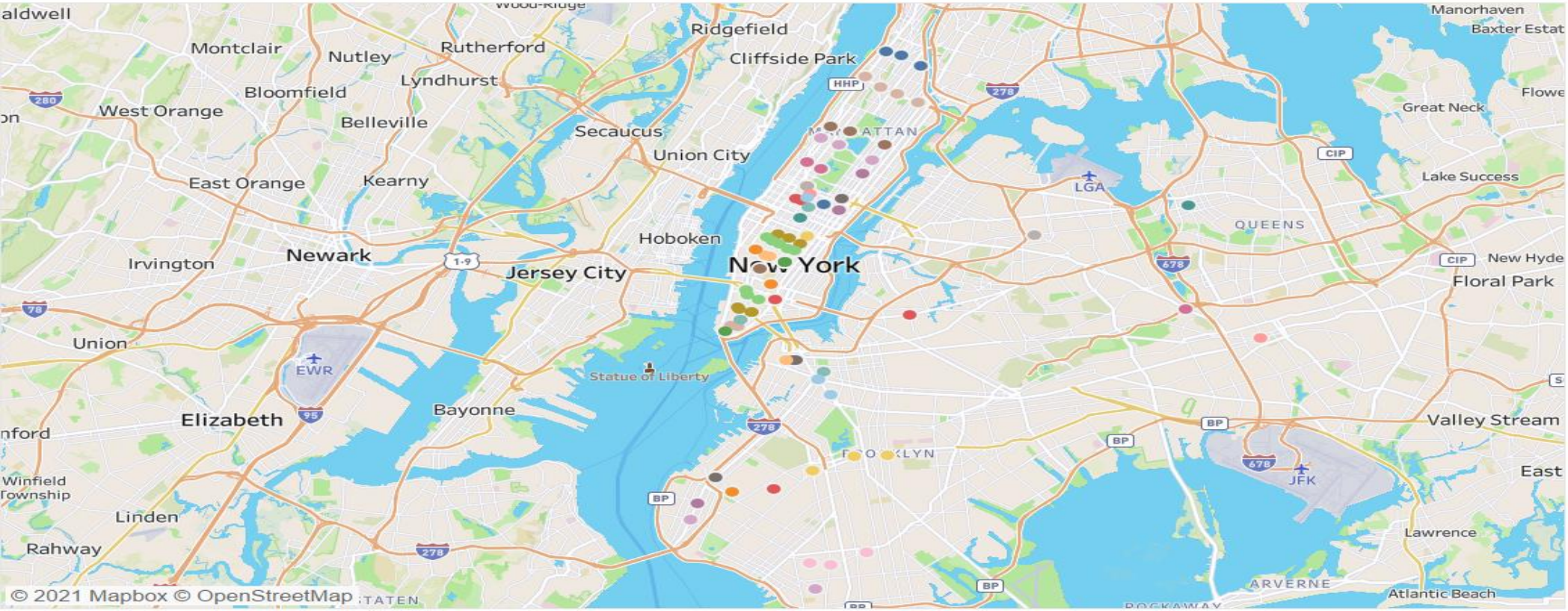


Heatmap represents the mean of the traffic distribution through the top 5 stations per the day of the week.



The bar chart above represent the population over the top 5 stations, PAVONIA/NEWPORT and FULTON ST has the highest population.

NYC Station

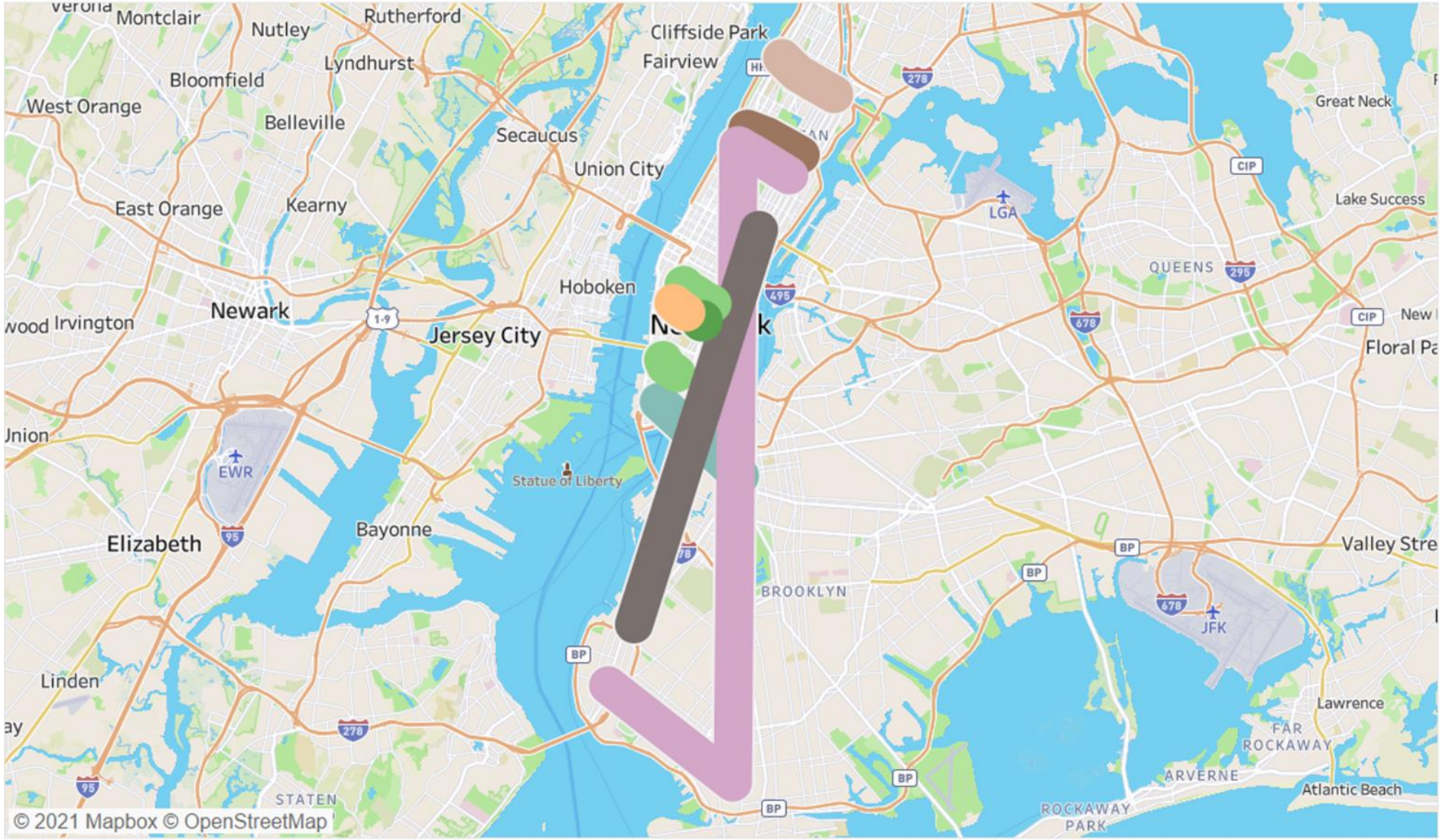


Map based on Longitude and Latitude. Colour shows details about Station. Details are shown for Line Name. The view is filtered on Longitude, which keeps non-Null values only.

Station

- 5 AV/53 ST
- 7 AV
- 8 AV
- 14 ST
- 14 ST-UNION SQ
- 23 ST
- 28 ST
- 33 ST
- 42 ST-BRYANT PK
- 47-50 STS ROCK
- 50 ST
- 57 ST-7 AV
- 59 ST
- 59 ST COLUMBUS
- 72 ST
- 77 ST
- 86 ST
- 96 ST
- 125 ST
- 145 ST
- ATL AV-BARCLAY
- B'WAY-LAFAYETTE
- BOROUGH HALL
- BOWLING GREEN
- CANAL ST
- CHAMBERS ST
- CHURCH AV
- FLUSHING-MAIN
- FULTON ST
- GRAND ST
- JAMAICA CENTER
- JAY ST-METROTEC
- JKSN HT-ROOSVLT
- KEW GARDENS
- KINGS HWY
- LEXINGTON AV/53
- W 4 ST-WASH SQ
- WALL ST

Top 5 stations by location



STATION (Weekend 10p-3a data.csv)
- 14 ST
- 14 ST-UNION SQ
- 23 ST
- 59 ST
- 86 ST
- 96 ST
- 125 ST
- CANAL ST
- FULTON ST

Map based on Longitude and Latitude. Colour shows details about STATION (Weekend 10p-3a data.csv). The view is filtered on STATION (Weekend 10p-3a data.csv) and Latitude. The STATION (Weekend 10p-3a data.csv) filter keeps 10 of 41 members. The Latitude filter keeps non-Null values only.

Recommendations:

- Highly recommends coffee corp. to open new branches next to the top 5 stations (23 ST,ST-PENN STA,GRD CNTRL-42 ST ,PAVONIA/NEWPORT, FULTON)
- open coffee shop branch next to exit of pavonia/newport station.
- recommends that you prepare to receive more customers in Friday .

Future works:

- Use geocoding to locate each of the top 5 stations with coffee shops next to it .
- Adding another data set for populations for each state .

Thank you ..please raise your questions