MTA EXPLORATORY DATA ANALYSIS PROJECT

RAGHAD ALAWAD

PROBLEM

Have you ever been stuck on a subway ride or late for a meetup due to the traffic you encounter at the stop? 90% of people have faced this problem before so it is struggling in our lifestyle and we have to find a solution to this problem to make our life easier.

GOAL

1/ Find the 5 stations that have most traffic

2/Find the five stations that have least traffic

3/Suggest events and activities at the crowded statings

THE PRE-PROCESSING

If there is spaces in columns names we remove it

```
copy_df.columns = [column.strip() for column in copy_df.columns]
copy_df.columns
```

SHOW THE DUPLICATED ROWS

```
(copy_df
 .groupby(["C/A", "UNIT", "SCP", "STATION", "DATE_TIME"])
 .ENTRIES.count()
 .reset_index()
 .sort_values("ENTRIES", ascending=False)).head(5)
```

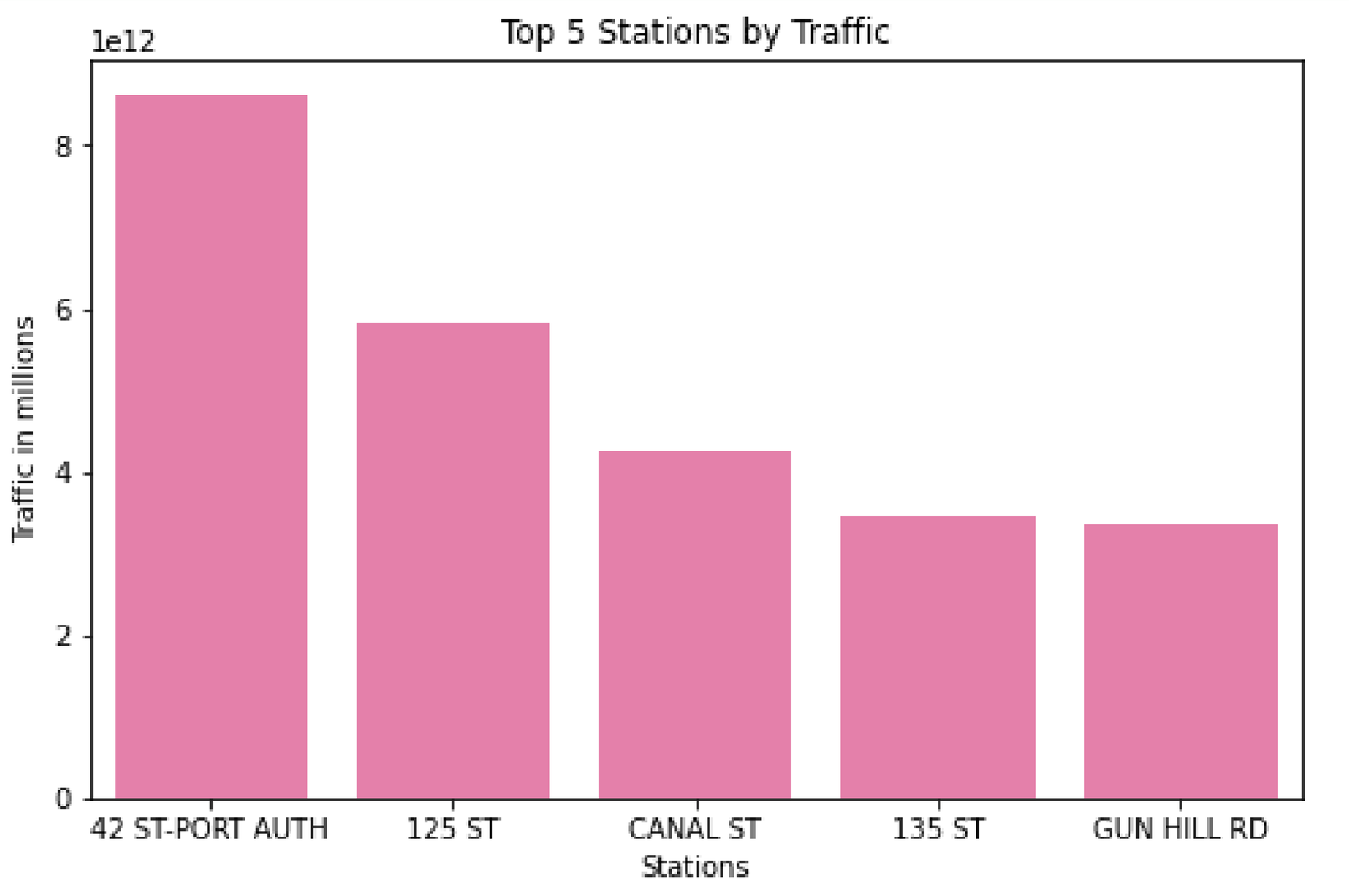
DROP COLUMNS

```
df_station = copy_df.drop(["DESC","C/A","UNIT", "SCP", "LINENAME", "DIVISION"], axis=1, errors="ignore")
```

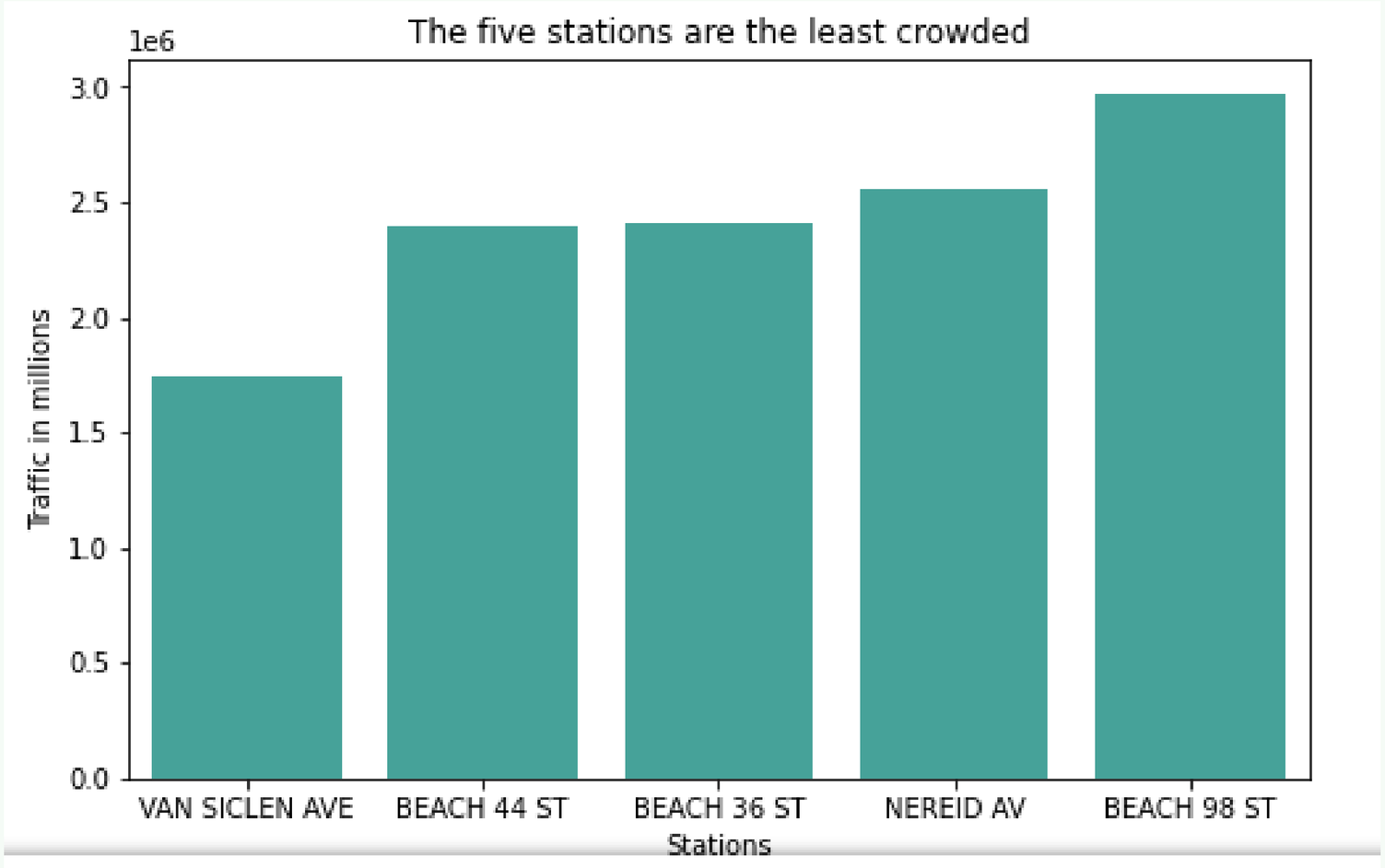
```
df_station.head(5)
```

	DATE_TIME	STATION	DATE	TIME	ENTRIES	EXITS
0	2021-01-30 03:00:00	59 ST	01/30/2021	03:00:00	7524539	2564693
1	2021-01-30 07:00:00	59 ST	01/30/2021	07:00:00	7524543	2564703
2	2021-01-30 11:00:00	59 ST	01/30/2021	11:00:00	7524566	2564755
4	2021-01-30 19:00:00	59 ST	01/30/2021	19:00:00	7524739	2564811
5	2021-01-30 23:00:00	59 ST	01/30/2021	23:00:00	7524821	2564823

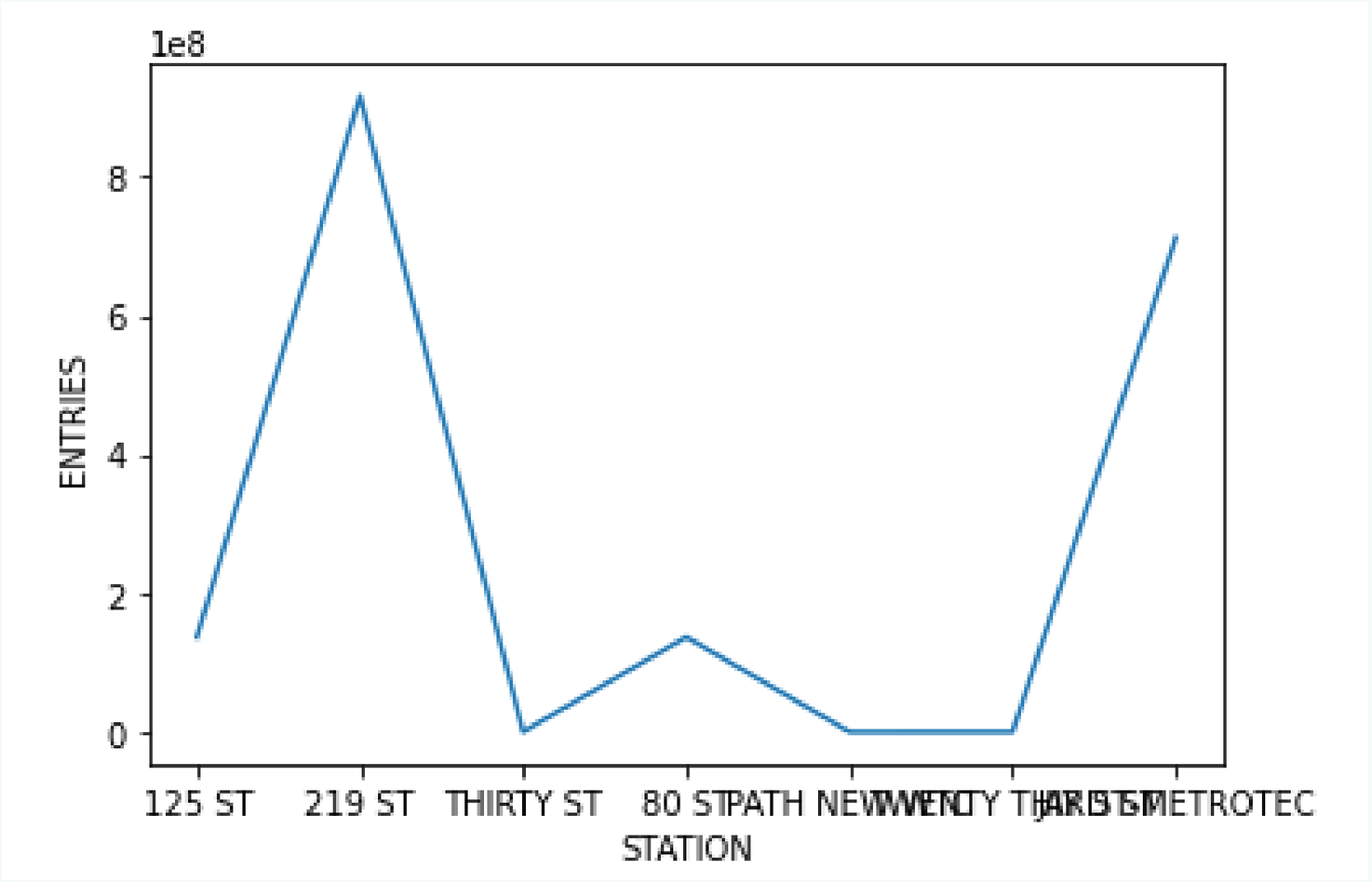
The top 5 stations by traffic and "**42ST-PORT AUTIH**" were the most trafficked stations



The five stations are the least crowded and "**van sicken ave**" is the least crowded



Seven stations were randomly selected and the crowds were known



CONCLUSION

As a result from this analysis we can conclude that 42ST-PORT AUTIH station is the most crowded station in NYC, whereas VAN SICKEN AVE station is the least crowded station.

THANK YOU