Project 1 : EDA MTA Dataset

By : Nada Alhamad

Project idea :

- Today we want to do data analysis to help us lead rental space, whether it's advertising screens, food, and drink machines, stores, or pharmacies, etc...
- We want : busiest stations and the days of the crowd.

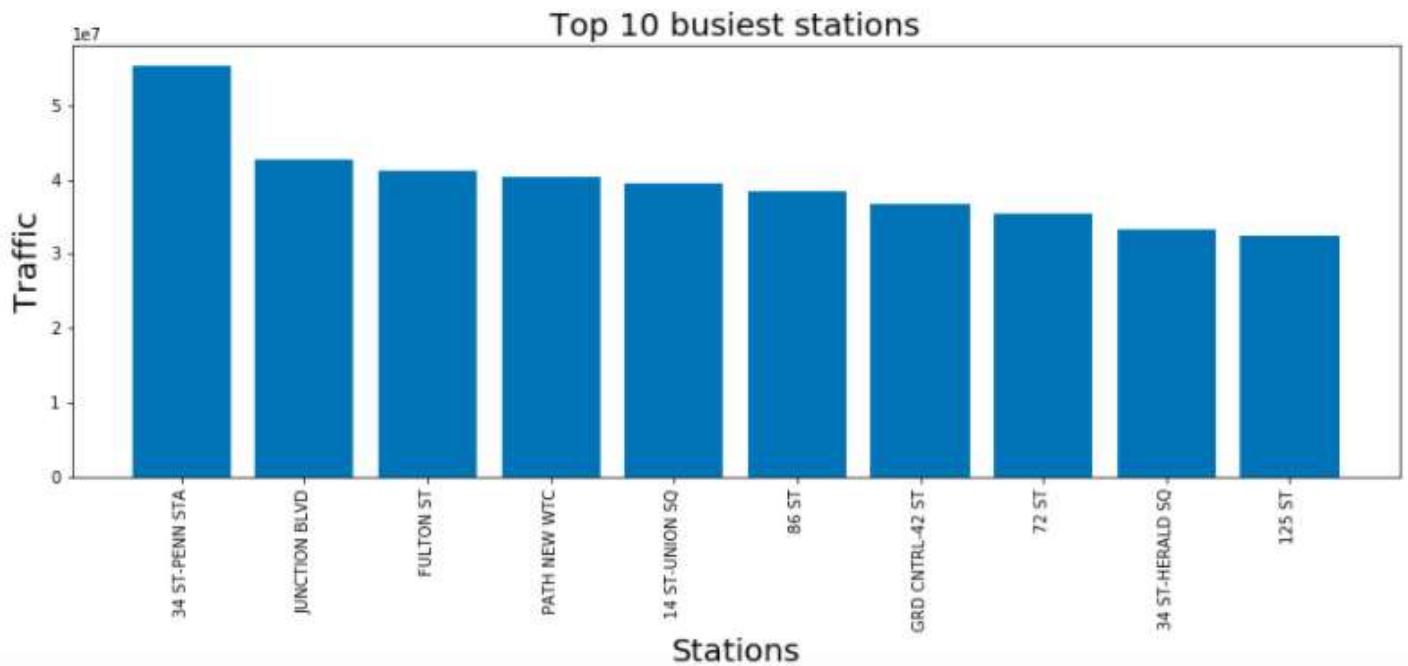
To perform this analysis we applied the following steps:

- load the dataset (data frame & SQLite)
- Convert dates and time into a string
- Clean (missing data, duplicate data, outliers)
- Statistical operations
- Data Visualization

Visualization:

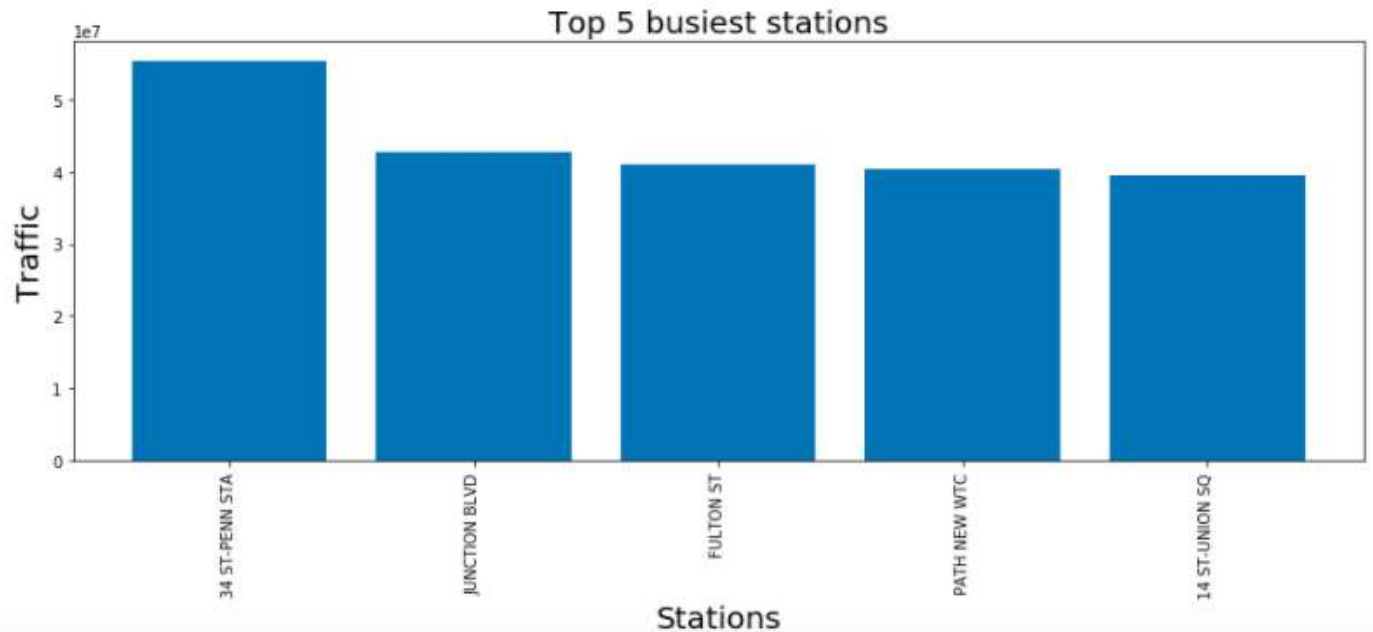
In [54]:
```
#Top 10 stations
plt.figure(figsize=(15,5))
plt.bar(x=station_totals['STATION'][:10], height=station_totals['DAILY_TRAFFIC'][:10])
plt.ylabel("Traffic",size=20)
plt.xlabel("Stations", size=20)
plt.xticks(rotation=90)
plt.title('Top 10 busiest stations',size=20)
```

Out[54]: Text(0.5, 1.0, 'Top 10 busiest stations')



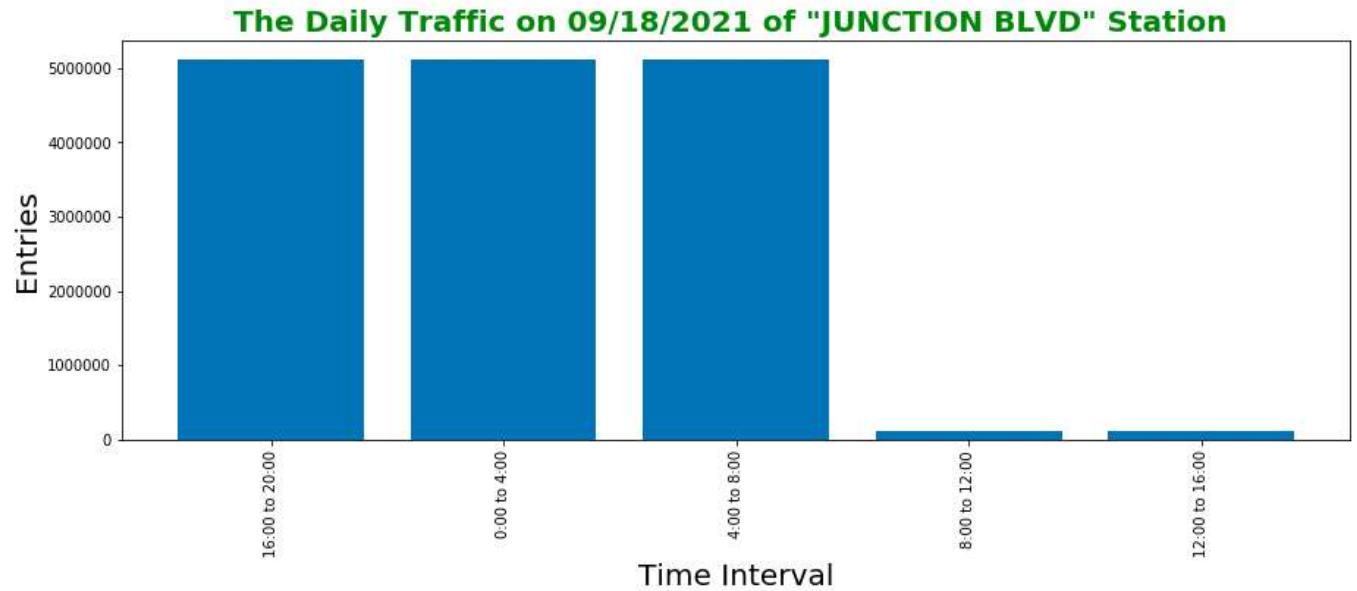
In [62]:
```
#top 5 stations

plt.figure(figsize=(15,5))
plt.bar(x=station_totals['STATION'][:5], height=station_totals['DAILY_TRAFFIC'][:5])
plt.ylabel("Traffic",size=20)
plt.xlabel("Stations", size=20)
plt.xticks(rotation=90)
plt.title('Top 5 busiest stations',size=20)
```

Out[62]:
```
Text(0.5, 1.0, 'Top 5 busiest stations')
```



In [68]:
```
# specefic date for one of the top station :
plt.figure(figsize=(15,5))
plt.bar(x=Days_one['TIME_INTERVAL'],height=Days_one['DAILY_TRAFFIC'])
plt.ylabel("Entries",size=20)
plt.xlabel("Time Interval", size=20)
plt.xticks(rotation=90 )
plt.title('The Daily Traffic on 09/18/2021 of "JUNCTION BLVD" Station ', color='g', size=20, weight="bold")
```

Out[68]: Text(0.5, 1.0, 'The Daily Traffic on 09/18/2021 of "JUNCTION BLVD" Station ')



Thanks for listening