****

**Name:** Ruba Alnashwan

**Email:** Ruubba72@gmail.com

**Github :** RubaAlnashwan

**Background:**

tss is an advertising company responsible for the billboard in train stations. Its main problem is to determine the price of advertisement at each train station to increase the income of the advertising company and the advertiser.

**Abstract:**

The goal of this project was use EDA to discover and analyze New York subway data, I took advantage of the matplotlib library to display some illustrations, such as the bar chart and line chart, which shows the busiest stations and which days of the week are the most traffic.

**Data description:**

The New York subway [MTA](http://web.mta.info/developers/turnstile.html) turnstile data is a series of data files containing cumulative number of entries and exits by station, date and time. Data files are produced weekly, data records are collected typically every 4 hours with some exceptions.

**Scope:**

In this analysis we use data from three months of 2021( January , March and August ). The total size of data for three months is 88.1 MB .

we chose the data for the year 2021 because it is the most recent data and as we know that with the beginning of 2021 the world came out of the covid-19 crisis and life return to normal. in addition to the presence of new areas that are increasing in population density that's why we constantly need new data. we chose four months March, April, June and August.

Each file represents a specific month containing 11 columns. The column names are( C/A Control Area , UNIT Remote Unit for a station, SCP Subunit Channel Position on represents an specific address for a device , STATION Represents the station name the device , LINENAME Represents all train lines that can be boarded at this station ,DIVISION Represents the Line originally the station belonged to BMT, IRT, or IND , DATE ,TIME ,DESC Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours) , ENTRIES The cumulative entry register value for a device , EXITS The cumulative exit register value for a device ). And total number of row is 3347283.

**Feature Engineering:**

Add a column for the month number.

Combine C/A+UNIT +SCP and add it in the TURNSTILE column.

Add columns Name of day and DateTime.

Calculate the DAILY ENTRIES + DAILY\_EXITS and add it in the traffic column

**Tools:**

**Technologies:** SQL, Python, Jupyter notebook.

**Libraries:** matplotlib, numpy, pandas.

## Communication:

## Through this Bar plot, it becomes clear to us that the (Junction Blvd , 103 ST-Corona, 191 ST) stations are the most crowded station during the four months , and the Bar plot shows the traffic according to the days of the week, and it turns out that Saturday is the most traffic day of the week and Monday is the least

## 