**Name:** Ruba Alnashwan

**Email:** Ruubba72@gmail.com

**Github :** RubaAlnashwan

**Background:**

AAT 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.

**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 82.1 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 three months, depending on some events like the month of January, it was chosen because it was the beginning of the year, and curfew was opened, and the month of March because life in it is very normal and August because of the summer vacation and a large number of tourist trips.

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 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 number of row in January is 210434 , March is 207055 and August is 209416

**Tools:**

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

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