**Abstract**

The goal of this project is to find use the data of NYC MTA Turnstile data to find the largest station among the 375 stations which has the largest number of turnstiles in New York to schedule maintenance in September 2021. Therefore, choosing the correct data is an important step to make use of it. The maintenance should be done when the station has the lowest number of people, to avoid any inconvenient for the people who use the station. Therefore, doing such analysis is very important for the maintenance company and MTA as will.

**Data**

The data that was used in this project is from MTA Turnstile Data, Since the maintenance will be in September 2022, the idea was to collect the data of the same month from the previous year. Therefore, the used data was for September 2021.

The columns are:

C/A = Control Area.

UNIT = Remote Unit for a station.

SCP = Subunit Channel Position represents an specific address for a device (turnstile).

STATION = Represent the name of the station.

LINENAME = Represent the line of the train trip.

DIVISION = Represent the operating company (IRT , IND or BMT)

DATE = Represents the date (MM-DD-YY).

TIME = Represents the time (hh:mm:ss).

DESC = Represent the "REGULAR" scheduled audit event (occurs every 4 hours) and it could be “RECOVERED”.

ENTRIES = The cumulative entry register value for a turnstile.

EXIST = The cumulative exit register value for a turnstile.

**Tools:**

1. Numpy
2. Pandas
3. Matplotlib
4. Seaborn
5. Jupyter
6. glob

**Communication**

The [slides](https://github.com/AliMufeed/project1/blob/aea1d0bc4ea8e0ab923a611ba79c2e2f926e2a03/Maintenance%20company%20for%20turnstiles%20-%20presentation.pdf) of the presentation as well as the [python code](https://github.com/AliMufeed/project1/blob/aea1d0bc4ea8e0ab923a611ba79c2e2f926e2a03/MTA_Turnstile_Data_for_turnsile_Maintenance.ipynb) are explaining the work done in this project in full details.