MTA Turnstile Data

Project Proposal

Alanoud Almutairi and Rahaf Alyousef

## Abstract:

The Metropolitan Transportation Authority is the largest transportation network in North America. The MTA network consists of the nation's largest bus fleet and more subway and commuter rail vehicles than all other United States. Nevertheless, transport companies were affected by the Corona epidemic economically. Therefore, officials and taxi owners will be provided with data on stations, crowded locations, and peak times so that they can increase the number of cars for the convenience of the passengers of each train and provide them with crowded stations or estimate fares in those areas and increase their income.

## Question/need:

With the Corona pandemic crisis, transport companies have been affected a lot, specifically the economic aspect. Therefore, officials and taxi owners will be provided with data on stations, busy locations, and peak times so that they can increase the number of cars in crowded stations or estimate prices in those areas, and increase their income. Taxi owners will benefit from this system by estimating delivery rates according to peak times or increasing the number of cars in crowded places.

## Data Description:

The New York MTA publishes weekly turnstile data on its developer page. data is a series of data files containing a cumulative number of entries and exits by station, turnstile, date, and time. Data files are produced weekly, data records are collected typically every 4 hours with some exceptions.

The data set consists of 11 columns, but 6 of them will be used are:

C/A = Control Area.

Station = Represents the station name the device is located at.

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

Time = Represents the time (hh:mm:ss) for a scheduled audit event.

entries = The cumulative entry register value for a device.

exits = The cumulative exit register value for a device.

## Target:

* After providing the taxi owners with the data, it is possible to increase the number of cars in those specified places and times.
* Raising the income of those responsible for taxis when raising the prices of delivery to those crowded stations.
* increase the number of cars to make it easier to ride each train and provide more business for taxis in the mall and restaurant around the station.
* We expect after the implementation of this system to increase the income of taxi drivers.

## Tools:

Project tools requirements will be presented using SQL and python, jupyter, and using the Python libraries: pandas, numpy, matplotlib.pyplot and seaborn.