**Abstract**

The goal of this project is to use The New York subway MTA turnstile to select the most crowded stations and fill them with vending machines for games nearby schools, books machines nearby business areas, snacks, and drinks in the most station has maximum number of entries

**Business need:**

In this project we used the MTA turnstile dataset that was provided by NYC subway to collect and analysis data and use them to help machines owners to see which station has the highest traffic at peak times and also select the top five stations ordered by highest entries, to allow the owners to use that data to provide a fun and useful machines for games ,books, newspaper and snacks that people can use it while waiting for their trains in these stations.

**Approach and methodology:**

We start by cleaning the data, exploring, and removing duplicated data and white space,

and then we select subset data based on the analysis that we need which is (Entries, Exits, Date, Time, Station Code, and Station name).

after that we use the sub data to find The Peak Times by selecting which station has the greatest number of entries in that time. The Peak times that we choose is from 8 AM to 12 PM and from 16 PM to 20 PM.

Also, we use Pandas for selecting top five station by sum the number of entries within three months.

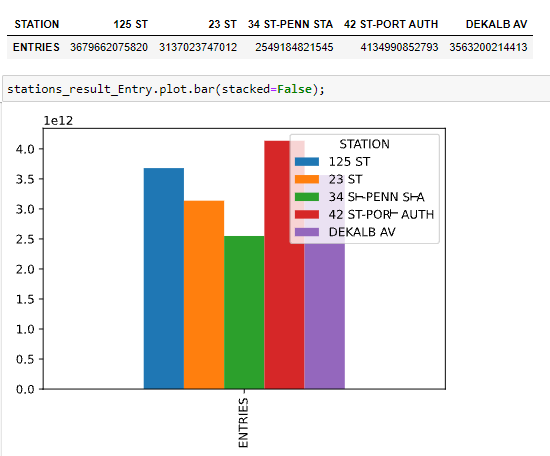
Finally, we Are visualizing peak time data using (Seaborn), and total entries using (Matplotlib)

**Analysis:**

The first step of our project is using pandas to clean data and removing duplicated columns and null values. secondly, splitting a data to sub data for the needed data which are entries, exits, time, date, C/A, and stations. After that, we looked for the most crowded stations, and we plot that data by using seaborn. The last step is seeking for select the top five stations by the highest number of entries and exits and we used matplotlib to plot that data to use it as a reference to the economic companies

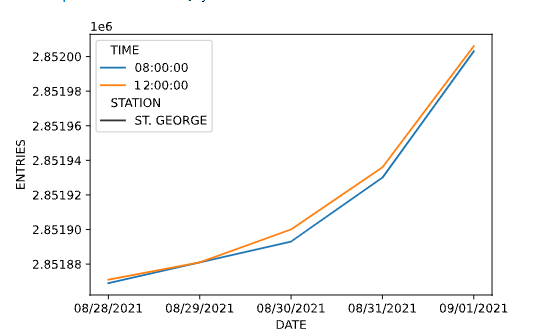
**Result:**

At the end of our research we found the result for top five station by entries number presented by Matplot graph And we did an extra search to see where the station is located if it under Business and tourism areas or commercial area so the top station (in red) is 42 Station with + 4 b entries per 3month, and is located under Times Square ,The second station (in blue) is 125 Station with + 3.6 entries per 3 months, and it is Near to The Empire State Building ,The third station (in purple) DeKalb Avenue Brooklyn station with 3.5 entries per 3 month and it is near to mall and parks ,The fourth station (in orange)23 Station with 3.1 entries per 3month, and is located on Lower Manhattan aka “financial capital” ,The last station is (in green) 34 ST-PENN STA with 2.5 entries per 3month.

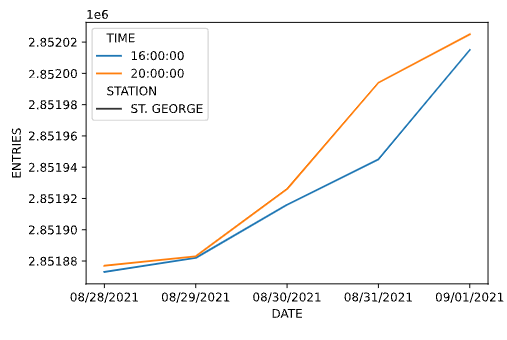


Furthermore, our result shows the result of most crowded station on the peak time between 8AM to 12PM , and 16PM to 20 PM is S101A, was station (S101A) which located on Staten Island surrounded by many of colleges and schools.

Morning peak time:

****

**Evening peak time:**

****

**Recommendations:**

our recommendation is to provide a customized vending machine based on where the station is located so, in first 3 station which is under Business and tourism areas

we suggest providing a vending machine containing business magazines, books, and Newspapers.

and the other 2 station Commercial areas:

we suggest providing a vending machine containing Ice cream, Snacks, Coffee, Games

Also, in most crowded station in peak time and because it Surrounded by many of colleges and schools so that it needs to provide many numbers of vending machines such as foods, drinks, ice creams and books