Hayat Aldhahri

Data Scientist

Project MVP

EDA on MTA Turnstile Data

I was assigned by the New York Orphanage Organization to assess them in their upcoming fundraising campaign in November 2021 by analyzing the data available from Metropolitan Transportation Authority. An Exploratory Data Analysis using the freely available data from the MTA website. Our goal is to identify the location or areas with the most traffic to place the fundraisers at those locations. Data for three months were analyzed and explored to identify top 5 busiest stations from the data set in terms of total number of travelers through the stations (exits and entries).

Chart, bar chart

Description automatically generatedChart, bar chart, funnel chart

Description automatically generated

Figure 1: Total Traffic for (Sep-2019)

Figure 2: Total Traffic for (Oct-2019)

Chart, bar chart

Description automatically generated

Figure 3: Total Traffic (Nov 2019)

The above graphs indicate the total traffic for the top 5 stations from the MTA data Set for the period (September to November 2019). Those are:

* 34 St-PENN station
* GRD CNTRL-42 Station
* 34 ST-HERALD Square
* 23 St.
* 14 St-Union Square

The ranking of the stations did not change except for “23 St” Station which was ranked 5th in September. In total, there are about 85 million visitors for the three months where majority are in October and November.

A further analysis has been done by looking at the traffic on each day of the week. As expected, the traffic during the weekday (Mon-Fri) is a lot larger than the weekend and traffic would slow during the weekend. If the fundraising is scheduled for 5-day, it is recommended to target the weekday rather than the weekend.

Chart, line chart

Description automatically generated

Figure 4: Traffic by Weekday