*Determine where and when to distribute masks in New York subway stations to reduce Covid spread.*

By: Archana Turumella

**ABSTRACT**

The objective of this analysis is to determine top MTA stations to maximize distribution of masks so that there is a high reach, and a greater number of people have access to them. Ultimately, when more people wear masks, the Covid spread in MTA trains and stations decreases. This analysis will help MTA prepare for next wave, if necessary. I worked with two data sets- MTA data and Covid Cases data from NYC Open Data source. I cleaned and transformed the data to derive insights such as what are the top 15 busiest stations to place masks. Since the ridership is dropped during a spike in covid cases, the best possible period to distribute masks is before there is a Covid wave predicted.

**DESIGN**

Studying the ridership of MTA trains before, during and after a Covid wave can give us insights of riders patterns. Analyzing the Covid cases data along with riders patterns would enable MTA to effectively determine the prime stations and best possible time period for mask distribution.

**DATA**

I collected data for the past three months (November 2021, December 2021, and January 2022) from *MTA data* and *Covid-19 daily case counts* data from “NYC Open Data”. MTA data has categorical features which can be transformed to arrive at station level data and daily riders count. Covid-19 data has daily cases and probable cases, of which probable cases is used as it is more relevant.

**ALGORITHMS**

Three datasets were mainly used:

* MTA top busiest stations.
* Daily count of riders in all the stations combined.
* Daily Covid cases count.

Visually plotting the above datasets, gave a clear picture of the top busiest stations and also showed there was a decrease in ridership as the Covid cases spiked. With this insight, we can propose that the masks can be reached to greater number of riders either before the spike or after the Covid cases drop but not during a wave.

Further I calculated the average number of riders before, during and after Covid wave. This is helpful for allocation of number of masks in Stations. As there was a 27% drop in the ridership at the beginning of a wave, the number of masks allocated to stations during this period can also be reduced and probably distributed in other places like hospitals and testing centers.

**TOOLS**

* NumPy and Pandas for data manipulation.
* Matplotlib and Seaborn for data visualization.
* SQLite and SQLAlchemy for data storage and querying.