Abstract:

This analysis will provide information which stations are most by entries and exits to open vaccine center in traffic station. Then, we will give recommendation based on this analysis on where to start implementation of this project. The MTA turnstiles data was used to in this project. The process to reaching results consisted of cleaning the data, aggregating the data and visualizing the data using different techniques.

Design:

This project if focused around identifying traffic stations found in the MTA turnstiles data by aggregating the entries and exists. The goal of the analysis is to find the busiest station to open vaccine center.

Data:

Will use available dataset about turnstile for 3 months worth of data from April to February, 2021. The features that I expect to work with is the number entries for each station, number of exists for each station to know busiest station.

Algorithms:

Exploratory data analysis and visualization techniques were used to clean, aggregate, and visualize the data. Cleaning data consisted missing values were checked using pandas and duplicate rows were deleted. Unnecessary columns were removed from the table. Aggregation of daily entries and daily exits were made using pandas, then the sum of ridership for each unique station was calculated using the sum of entries and exits. For visualization, matplotlib and plotly were used to create bar graphs.

Tools:

Technologies: Python language, Jupyter Notebook and SQLAlchemy

Libraries: Pandas, Matplotlib and Numpy.

Communication:

PowerPoint slides were created to describe the need and visuals of findings were included in these slides.