

EDA Project Proposal Template

Munwon Jung

Question/need:

With the breakout of Covid-19, there have been some changes to the MTA service. MTA has been installing hand sanitizers at various stations and has been requiring individuals to wear a mask. In case individuals do not have a mask, it is stated that they can ask for a free mask at the subway station booth.

With the MTA turnstile dataset, this project aims to examine and investigate the turnstile data to draw insights regarding the most populated station to effectively allocate the amount of hand sanitizer + free masks needed in each station. Furthermore, ridership will also be examined for each station over the course of a few months to help MTA collect useful data for the amount of masks/hand sanitizer they would need to replenish depending on the ridership through various stations.

If there is more time to add on to the project, Covid-19 data provided by the NYU Health Department will be added to compare whether there was a noticeable change in the number of patients after having the hand sanitizer installed or giving out free masks by MTA throughout the pandemic period.

Data Description:

The data will be mainly focused on the Turnstile Data(Dec. 2020 - Mar. 2021) provided by MTA: (<http://web.mta.info/developers/turnstile.html>). There will be a main focus on the Entry(cumulative entry register value), Stations, Lines in each station in NewYork to ultimately find which station is most populated through the day.

Tools:

Data will be downloaded as .txt files then be converted to a big one chunk of .csv file. After setting up an engine for the MTA database, the database will be connected through Python and further data cleaning will be done through SQLAlchemy.

- Data Clean through Python Pandas / Sqlite
- Deal with irregularity in the data(account for number of Subunit Channel Position located in station), consider how to deal with inaccurate data recorded...etc

MVP Goal:

The minimum viable product will be a graph of the most entered station compared to other vaious stations on a daily or monthly basis.