Impact of COVID – 19 on Metropolitan Transportation Authority: Analysis of Pre-covid and Post-covid data

Study conducted by: Himani Kaushik

Backstory

New York City MTA (The Metropolitan Transportation Authority) needs to determine the impact of COVID-19 and the resultant remote work on the daily commute in the city in the last two years. This will enable MTA to plan and optimize the available resources and be prepared for the change in demand, assuming there will be no new variant and the current situation is the new normal.

- **Question/need:**
  - What is the framing question of your analysis, or the purpose of the model/system you plan to build?
    - The scope of the project is to ascertain the changes in the travel pattern in last two years. Hence, we aim to analyze data from two different timeframes - *Pre-COVID:* November 2019, December 2019, January 2020 and *Post-COVID*: November 2021, December 2021, January 2022. MTA will use this analysis for effective resource management and forecasting future travel patterns, if possible.
  - Who benefits from exploring this question or building this model/system?
    - The Metropolitan Transportation Authority, New York City.
- **Data Description:**
  - What dataset(s) do you plan to use, and how will you obtain the data?
    - MTA turnstile data - http://web.mta.info/developers/turnstile.html
    - I will load the data from the site into SQL database. I may also explore using APIs or web scraping.
  - What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?
    - The individual sample includes all the entrances and exits for an entire station over a 4-hour period.
    - I expect to work on C/A, UNIT, STATION, DATE, TIME, ENTRIES and EXITS columns from the data.
  - If modeling, what will you predict as your target?
    - I will try to predict the future traffic patterns for MTA, if applicable.
- **Tools:**
  - How do you intend to meet the tools requirement of the project?
    - SQLite and DB Browser: Creating database and tables
    - Sqlalchemy: Accessing SQL database in Python
    - Pandas and Numpy: Manipulating data
    - Matplotlib and Seaborn: Visualizing and plotting data
  - Are you planning in advance to need or use additional tools beyond those required?
    - Yes, I may use APIs/web scrapping to obtain the data and other Python packages to visualize data.