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

Project Writeup by: Himani Kaushik

Abstract

The project goal is to analyze the traffic data from MTA for pre covid and post covid time frames and understand the shift in the volume due to COVID and resultant changes in lifestyle.

The analysis of the current data will identify the changes in the traffic pattern and hence, enable MTA to make informed decisions and optimize the resources. The project assumes that there will be no new variant and the current situation is the new normal.

Design

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. MTA will use this analysis for effective resource management and forecasting future travel patterns, if possible.

Data

Initially, the time frames for the project were - *Pre-COVID:* November 2019, December 2019, January 2020 and *Post-COVID*: November 2021, December 2021, January 2022. However, to make analysis more relevant and current, timeframes were changed as follows:

- *Pre-COVID:* February and March 2019
- *Post-COVID*: February and March 2022

I obtained the MTA turnstile data from the website for the above-mentioned months. The dataset has 3509986 rows and 11 columns, totaling 38609846 observations. I utilized STATION, DATE, TIME and ENTRIES information from the data to perform EDA and draw conclusions.

Algorithms

I followed these steps to complete the project analysis:

1. Ask the right questions: In order to analyze the traffic patterns, I asked several questions to reach the goal:
    a. Which are the top 20 stations before and post COVID?
    b. Which are the bottom 20 stations before and post COVID?
    c. Are traffic patterns different for weekdays and weekend.
    d. How much decline in the volume for top stations?
    e. Should I focus on one specific station?
2. Collect and clean data: I collected and stored the data in database using SQLite and SQLAlchemy. I cleaned the data using python and pandas framework.
3. EDA: I explored and analyzed data using several pandas methods like groupby, sort_values, to_datetime, merge, and many more.
4. Communicate: I plotted the results using Matplotlib and Seaborn.
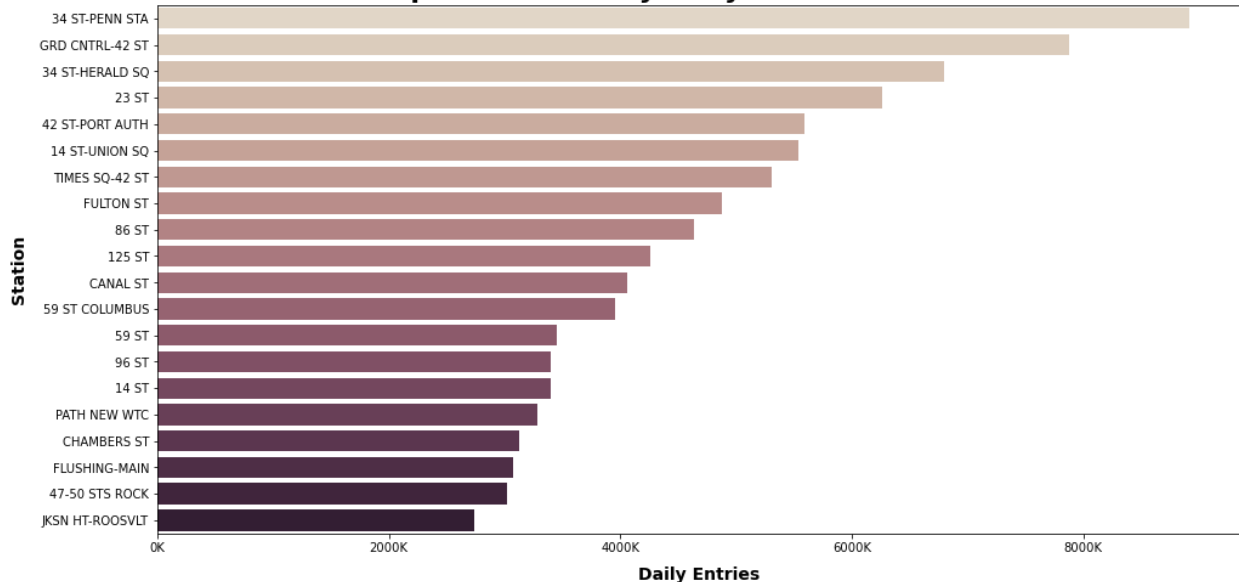
Communication

There is a huge decline in the traffic volume at all the stations. Even the busiest stations are operating at more than 60% decline at the moment. The bar chart displays the names of the station on the horizontal axis and number of entries on the vertical axis. The chart depicts the precovid and postcovid daily entries for stations, which were the top stations in February-March 2019.
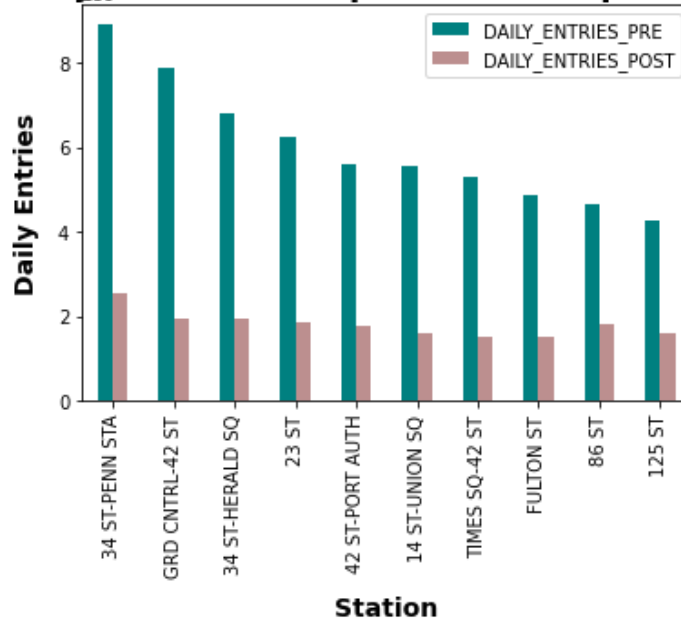
MTA should continue to focus their resources on these stations:

Top 20 Stations by Daily Entries - PostCovid