The purpose of the project was to provide an analysis of commuter behavior between pre-covid data and post-covid patterns.

The client requested this data to understand what the impact of Covid-19 and the current remote work culture has had on public transport use. This will provide insight into possible development opportunities in new neighborhoods that are seeing sustained demand for public transport through the pandemic.

The model is based on six months of MTA turnstile data between March – September 2019 and 2021. After cleaning the data, it was modeled out to show the aggregate daily entries during the nine-month timeframe and concluded to the top 25 most active subway stations in terms of passenger entries via turnstiles across NYC before the pandemic and after.

Design

- The client backstory was fleshed out and an overview was provided during the presentation.

Data

- The data set includes 12 months of MTA turnstile data in total. This includes six months between March – September 2019 and between March – September 2021.

Algorithms

- The analysis was done in Jupyter Notebook and data was cleaned through Python code and filtered.

Tools

- SQL was effectively used to manage the initial raw data file which was then exported into Python via SQLAlchemy.
- Python packages including matplotlib and pandas were used to clean, analyze, and visualize the data.

Communication

- A presentation was provided to the client that provided major findings, key insights, and conclusions to their questions.