ABSTRACT:

My goal with the Exploratory Data Analysis module was to use the MTA Turnstile Data to determine a few interesting patterns in commuter usage. I was hoping to use these patterns to help the MTA to adjust their maintenance schedule to minimize the disruption to the subway system and also to maximize maintenance worker efficiency by allowing them less distractions while they work.

DESIGN:

The design of my project focused on the movement of MTA commuters and usage of MTA turnstiles and stations. Using the data provided by the MTA, my goal was to help the corporation serve their clients the best way they can, as well as try to improve client experience whilst using the subways.

DATA:

My project uses 54 weeks of MTA turnstile data focusing primarily on specific dates and daily users of the subway stations on those specific dates. In those 54 weeks there are 11,319,718 rows of data with many features such as date, time, station, c/a, unit, scp and others. This data was imported into JupyterLab using SQLAlchemy.

ALGORITHMS:

The main algorithm used in this project was a function that turned unclean data from the MTA into usable daily entries that are nice to compare in various graphs.

TOOLS:

Ingested Data to Python using SQLAlchemy
Pandas exploratory data analysis
Matplotlib visualizations

COMMUNICATION:

This project was communicated through a series of slides that display my findings regarding the MTA data.