BERNARD OPOKU

INVESTIGATING MTA RIDERSHIP FOR A NEW TUNNEL CONSTRUCTION

ABSTRACT
The goal of this project is to see the ridership patterns with the MTA data for a new underground tunnel construction. New York City is a city well known for congestion and higher growth rate. With this I mind, I want to use to use this data to suggest to constructors on how big stations should be, the staffing and security prioritization

DATA
As stated, the data is from the mta developer website which is available for the public.
-I used 5 sets of the data
- 3 months or 13 weeks each for 5 consecutive months
2015- 10/17/2015 – 1/9/2016
2016- 10/15/2016 – 1/7/2017
2017- 10/14/2017 – 1/6/2018
2018- 10/13/2018 – 1/11/2019
2019- 10/12/2015 – 1/4/2020

ALGORITHM &DESIGN
I created an empty database in sqlite, I then used python to get the txt files from the mta website, I then combined all the text files in one csv file for each data before I exported them into their respective sql tables in a single database. I lastly exported the database to pandas dataframe using sqlalchemy to prepare for the pandas, python part of the eda.
I worked on one data, then rapped the process into a function to repeat the cleaning processes with the remaining 4 datasets.

TOOLS
Python,
SQL,
Pandas,
Numpy,
seaborn,
matplotlib,
pandas bokeh,