

Problem definition:

In this project, I will use a traffic dataset to build a map showing the traffic condition in New York City at different hours during the day. The main purpose of the project is to give users an insight as to which hours during the day are peak hours, and which area has the most traffic during these peak hours. This can help users navigate around the city avoiding high traffic areas during peak hours. The goal is to help reduce the traffic in a specific area and help reduce the travel time during peak hours.

Target Audience:

The target audience of this project will be people who travel a lot in the metropolitan area, or businesses that depends on traffic conditions such as taxis or post service

How this project helps the target audience:

With a map that shows the traffic conditions for each hour during the day, it can provide information to the user on how to avoid locations with the highest amount of traffic.

Dataset description:

This dataset is from the department of transportation, it contains data such as location, traffic count at each hour, and the date where the data was recorded. This project will utilize the FourSquare location data to create a map showing the traffic counts in a specific area during each hour of the day. Since the devices used to measure the traffic count of the streets will only cover a small area, the map will only show the traffic count where the device is placed.

FourSquare dataset:

The FourSquare dataset will be used to give an approximation to where the traffic count is recorded. This data will be in the format of geojson and it will be used in the python script along side the traffic dataset.

Data Cleaning:

The dataset obtained online contains 31 columns of information, 24 of which are the hours throughout the day. The most important information would be the time and traffic count.

Therefore only these features are preserved in the dataset.

| Features used | Features dropped |
|--|---|
| <ul style="list-style-type: none">• Roadway names• Time (hours) | <ul style="list-style-type: none">• ID• Direction• From/to• Date |

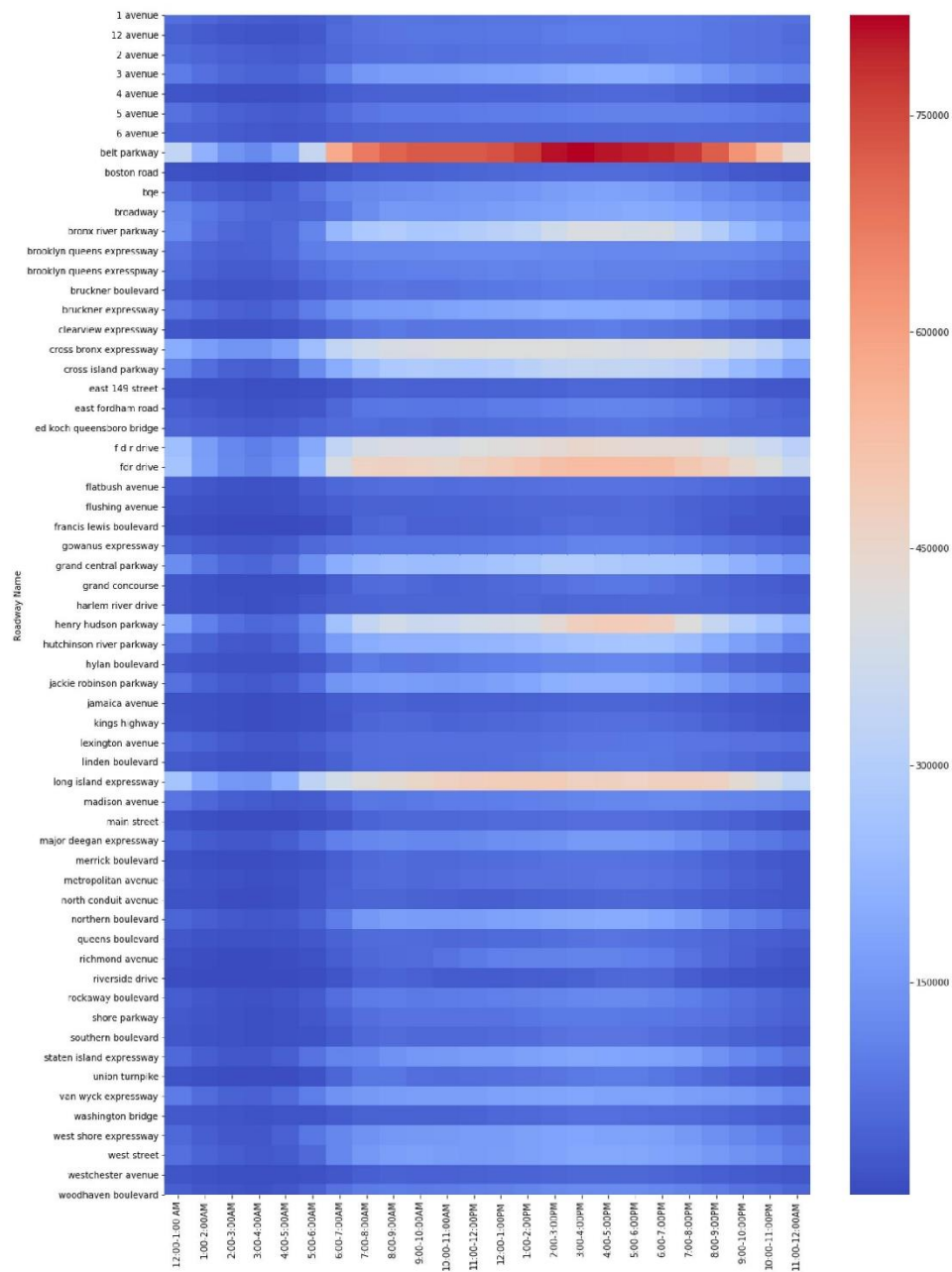
Data analysis:

Since the dataset contains traffic counts for 24 hours every day, the data does not reflect clearly on the traffic conditions during peak hours. To further understand the dataset, I have created a filter that sorts out roads that have less than 40,000 cars per hour during peak hours, such as 7am – 9am and 4pm – 6pm. This way the results will only show roads with heavy traffic and people should avoid using it during these hours.

Data Visualization:

In order to understand more about the dataset, I chose to use seaborn's heatmap to do so, since the data set contains more than 1000 streets, heatmaps are the most effective way to visualize the dataset.

Results:



Based on the results, we can conclude that most of the traffic occurs at expressways during peak hours, with an average of 60,000 cars per hour. The maximum traffic count in around 80,000 cars per hour at belt parkway.