Data Capstone project

Conan Sum

Problem definition

- In this project I will be investigating traffic during peak hours in NYC metropolitan area

Target audience

- The main target of the report will be people who have to travel a lot in the metropolitan area
- For example, taxi drivers or post service
- The results of the project will give them information about which route to avoid during peak hours

Dataset

- The data is recorded by the department of transportation and it contains data for every hour of the day
- The dataset contains more than 1000 streets around New York City
- Data is recorded from year 2014 - 2018

Data Cleaning

- The dataset from the website is large and it contains a lot of overlapping information
- In order to clearly present these information, I have grouped the data based on their street name

Analysis

- The dataset contains information for 24 hours each day, which does not reflect peak hours traffic count
- The peak hours I have selected for this dataset is from 7am – 9am and 4pm – 6pm
- I have filtered the data using these hour ranges for a better representation of traffic conditions

Conclusion

- Based on the results, I have concluded that most traffic occurs at expressways, with an average of 60,000 cars per hour
- Therefore if someone has to navigate through the city, they would want to avoid expressways