Project -1 **Traffic Flow Prediction and optimization**

Name : ***kratika soni*** *Batch –* ***DST20823***

Batch Date: 01/09/2023

Submitted to : Kevin shah

Problem Statement :

Traffic congestion is rising in the cities around the world. Contributing factors include expanding urban populations, aging infrastructure, inefficient and uncoordinated traffic signal timing and a lack of real-time data.

The impacts are significant. Traffic data and analytics company INRIX estimates that traffic congestion cost U.S. commuters $305 billion in 2017 due to wasted fuel, lost time and the increased cost of transporting goods through crowded areas.

Given the physical and financial limitations around building additional roads, cities must use new strategies and technologies to improve traffic conditions.

Content:

This dataset contains 48.1k observations of the number of vehicles each hour in four different junctions:

1) DateTime

2) Juction

3) Vehicles

4) ID

About the data:

The sensors on each of these junctions were collecting data at different times, hence you will see traffic data from different time periods.

Some of the junctions have provided limited or sparse data requiring thoughtfulness when creating future projections.

Number of Instances: 48120 Number of Attributes: 4

Source: A comma separated value file containing the dataset in tabular format – traffic.csv.

Solution:

a) Detailed analysis of data is performed using sql, plotting figures and feature engineering

b) I trained a GRU, CNN, MLP, LSTM and my custom Model to predicted the traffic on four junctions

Technology Stack:

1) Anaconda – Jupyter notebook
2) Excel

Libraries: numpy, pandas, matplotlib, seaborn, tensorflow,keras etc.

Knowledge: Python and their Libraries, Machine Learning concepts and Gated Recurrrent unit, Long short term memory which is designed to model or more.

Visualization:

Feature Engginearing

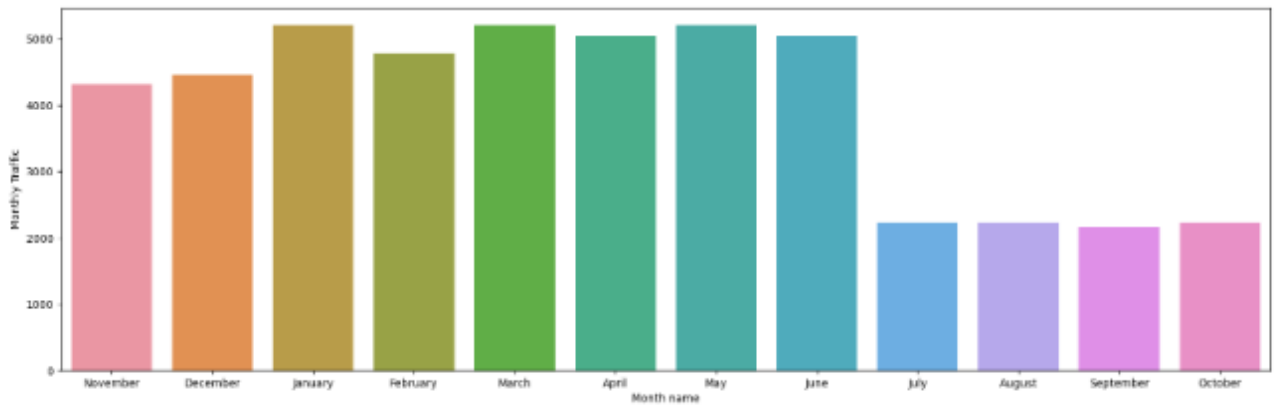
Extracting Year, Month, Month Number, Day, Date Number, Hour From Datetime.

Extracting dawn, mid-noon, dusk, night time from datetime.

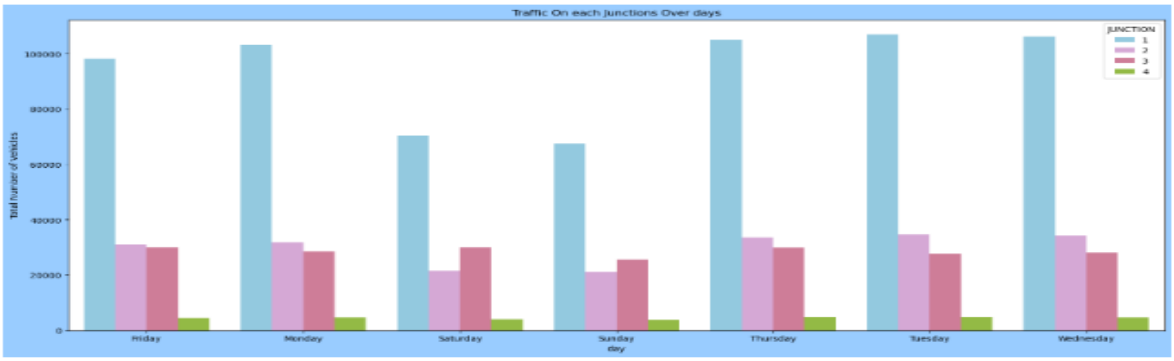
1)There are 4 Junctions reffered as 1 to 4.

2)The maximum vechiles passed is 180 in a particular datetime.

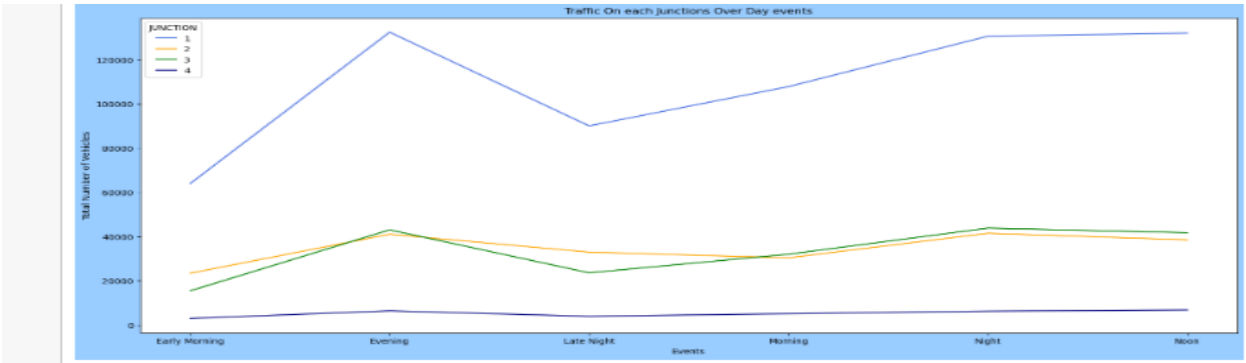
3)The data is from November 2015 to june 2017.



1)January, March, May has most number of travelling history.



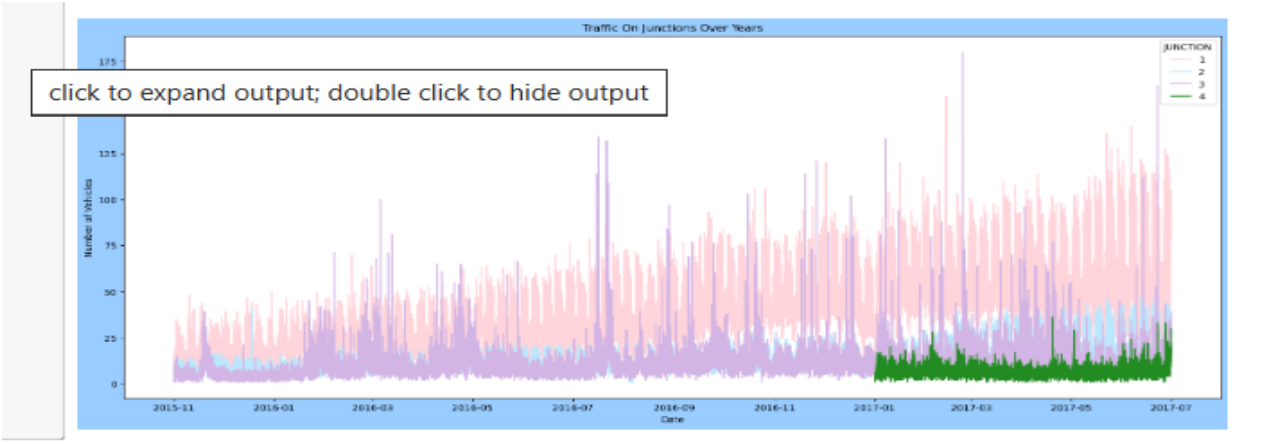
1)Tuesday, thursday, wednesday, monday these days people travel more may be working-days.



1)From junction 1 most people travel and maximum in noon to night andless in late night .

2)From junction 2,3 people travel maximum in noon to night.

3)From junction 4 less people travel and maximum in evening.



It can be seen here that the first junction is visibly having an upward trend.

The data for the fourth junction is starting only after 2017

In the notebook file there is detailed analysis and charts for data visualization and various model used for predicting the traffic congestion on four junctions.

Results:

In this project, I trained a GRU, CNN, MLP, LSTM and my custom Model to predicted the traffic on four junctions. I used a normalisation and differencing transform to achieve a stationary timeseries.

As the Junctions varry in trends and seasonality, I took diffrent approach for each junction to make it stationary. I applyied the root mean squred error as the evaluation metric for the model. In addition to that I plotted the Predictions alongside the original test values. Take aways from the data analysis:

The Number of vehicles in Junction one is rising more rapidly compaired to junction two and three. The sparsity of data in juction four bars me from making any conclusion on the same.

The Junction one's traffic has a stronger weekly seasonality as well as hourly or mid day to night seasonality. Where as other junctions are significantly linear.