Assignment - 1

Problem Statement - The main motive of this assignment is to predict the taxi trip duration with the help of features in the dataset “Taxi Trajectory Data” from Kaggle.

The feature “POLYLINE” is a string, with a list of GPS coordinates. The trip duration is calculated by using a formula - (len(POLYLINE)-1)*15), this gives the trip duration a particular record in a dataset in seconds.

- The Dataset used for this assignment is Taxi Trajectory Data from Kaggle (Dataset).
- Data cleaning and preprocessing, building a model done in goggle collab.
- Data is loaded from drive to collab.
- Dataset downloaded from Kaggle is around 2GB.
- Dataset has a total of 9 features.
- Columns in the dataset:
    - TRIP_ID is a unique identifier for each trip
    - CALL_TYPE is the way used to demand the taxi
        - ‘A’ id it is from central, ‘B’ is from Taxi stand and ‘C’ otherwise
    - ORIGINCALL is a unique identifier for the customer phone from which the trip was requested
    - ORIGINSTAND is a integer number given based on the column ‘CALL_TYPE’
    - TAXI_ID is a unique identifier for taxi driver
    - TIMESTAMP is the trip start time
    - DAYTYPE is the day when the trip started
    - MISSINGDATA is true if there is any missing data in a particular row
    - POLYLINE is the list of GPS coordinates.
- TRIP DURATION in seconds is calculated by (len(POLYLINE)-1)*15 sec.