Traffic Route Prediction

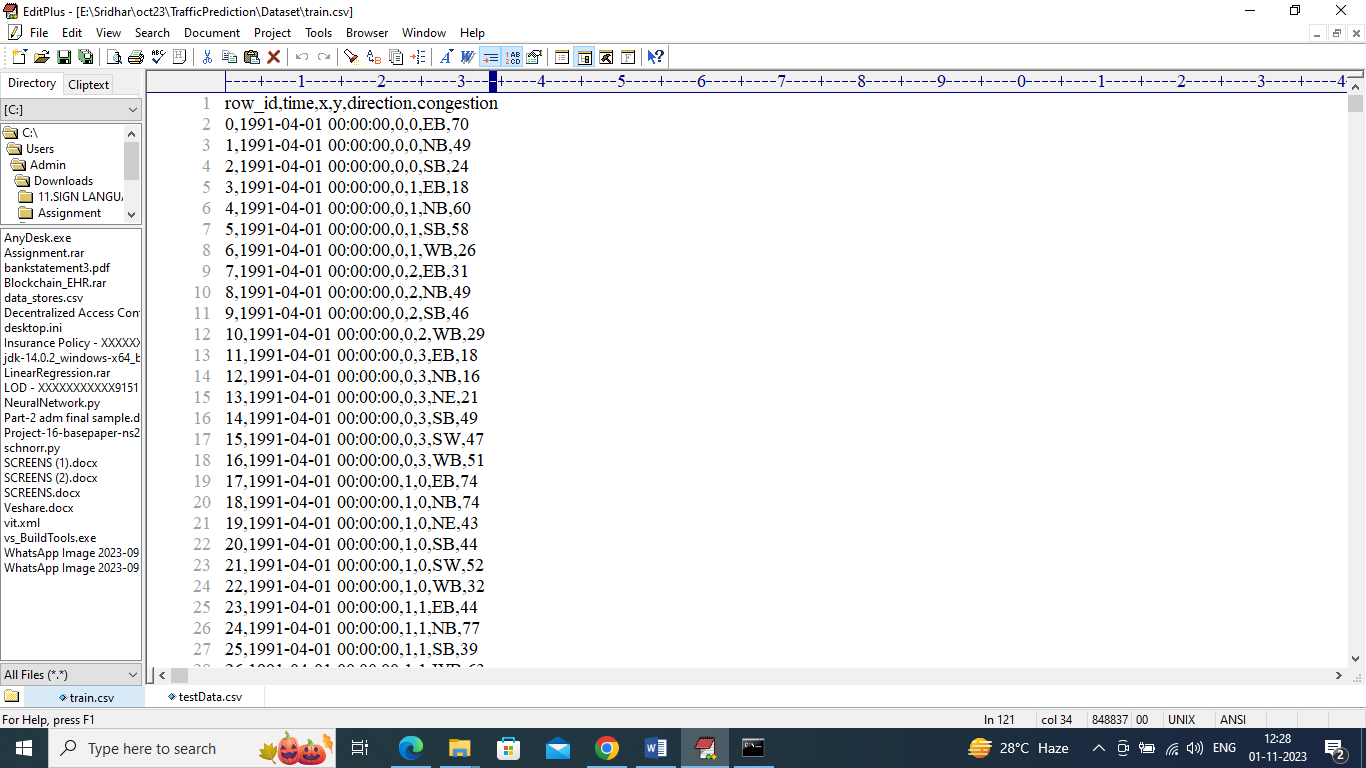
In this project based on traffic congestion we are predicting alternate route for the passengers, to predict route we are training various machine learning algorithms such as SVM, Decision Tree and Random Forest and each algorithm performance is evaluated in terms of accuracy, precision, recall, FSCORE and confusion matrix graph.

Before applying ML algorithms we have performed data analysis via graph visualization to understand traffic flow and congestion in different routes and after analysis we have employed ML algorithms for route prediction.

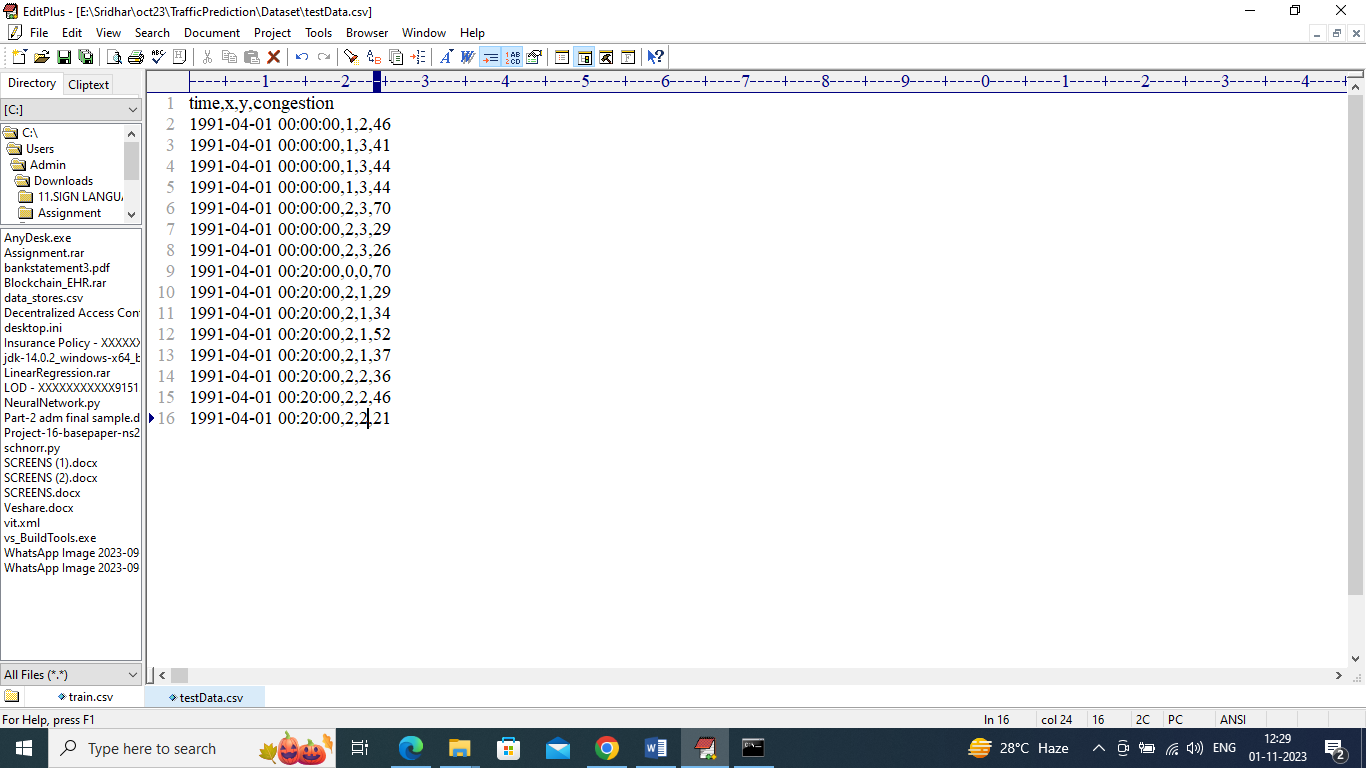
Among all algorithms SVM performing worst and Decision Tree perform well. To train all algorithms we have utilized traffic congestion dataset which can be downloaded from below KAGGLE URL

<https://www.kaggle.com/competitions/tabular-playground-series-mar-2022/data?select=train.csv>

In below screen we are displaying dataset details

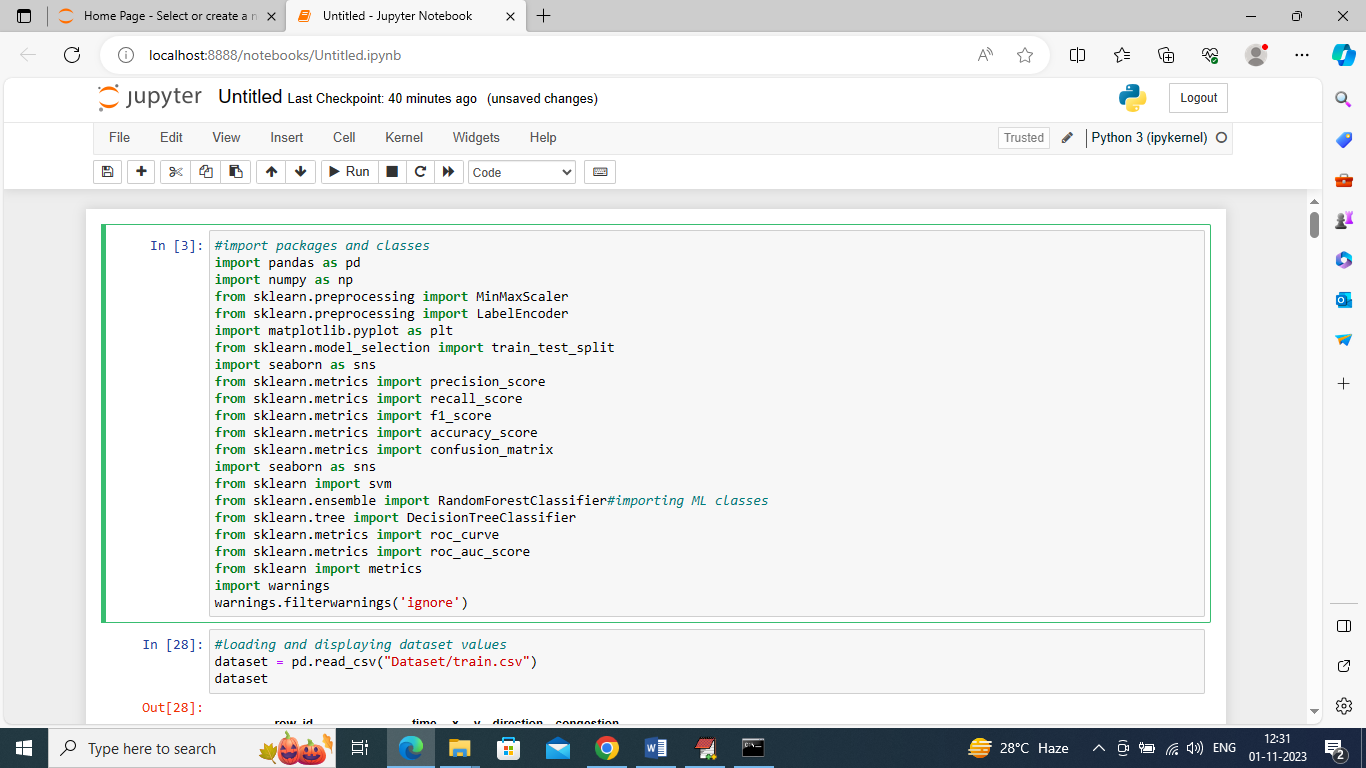


In above dataset we have Date, X and Y as latitude and longitude and then direction and then last column contains traffic congestion and by using this module we will train all ML algorithms. After training we will employ test data to predict alternate direction and below is the TEST data.

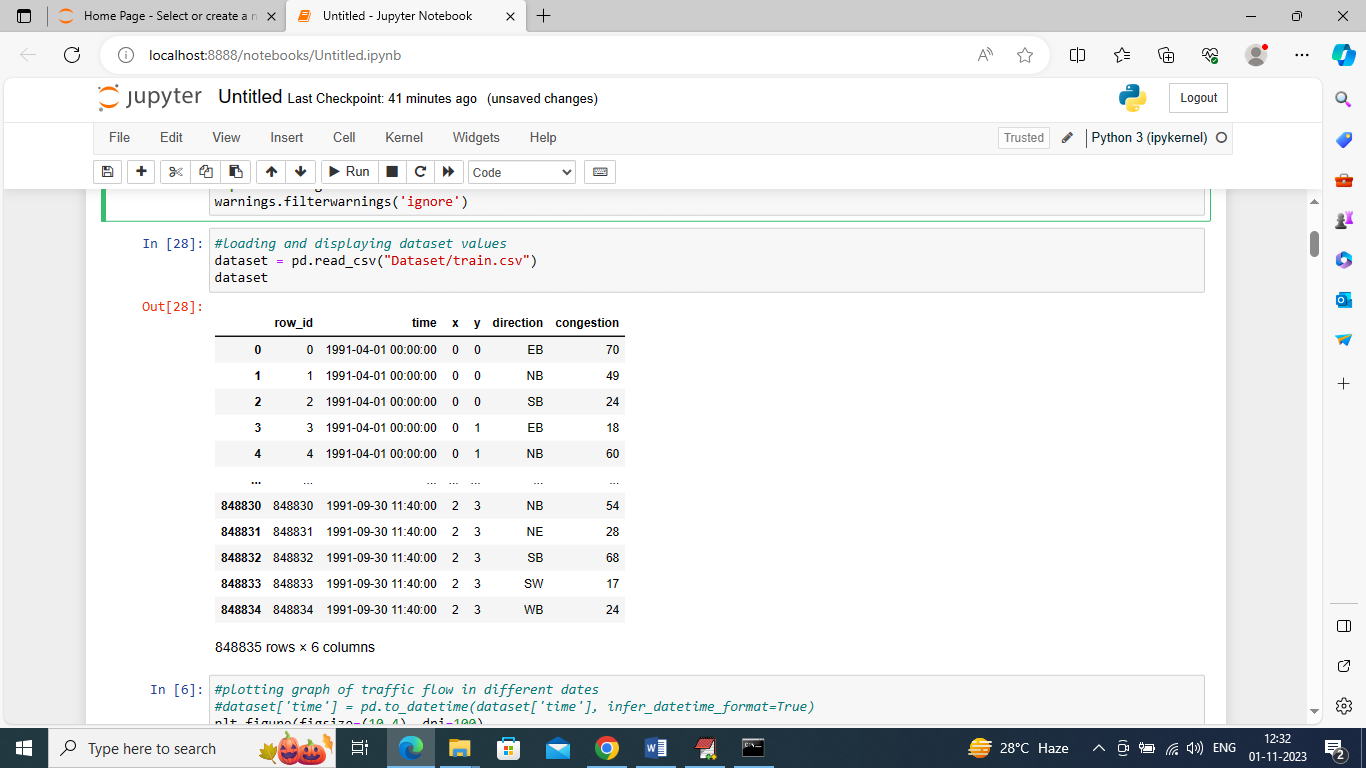


In above test data we have Date, X and Y location and traffic congestion but direction or route column is not available and this direction will be predicted by ML algorithm

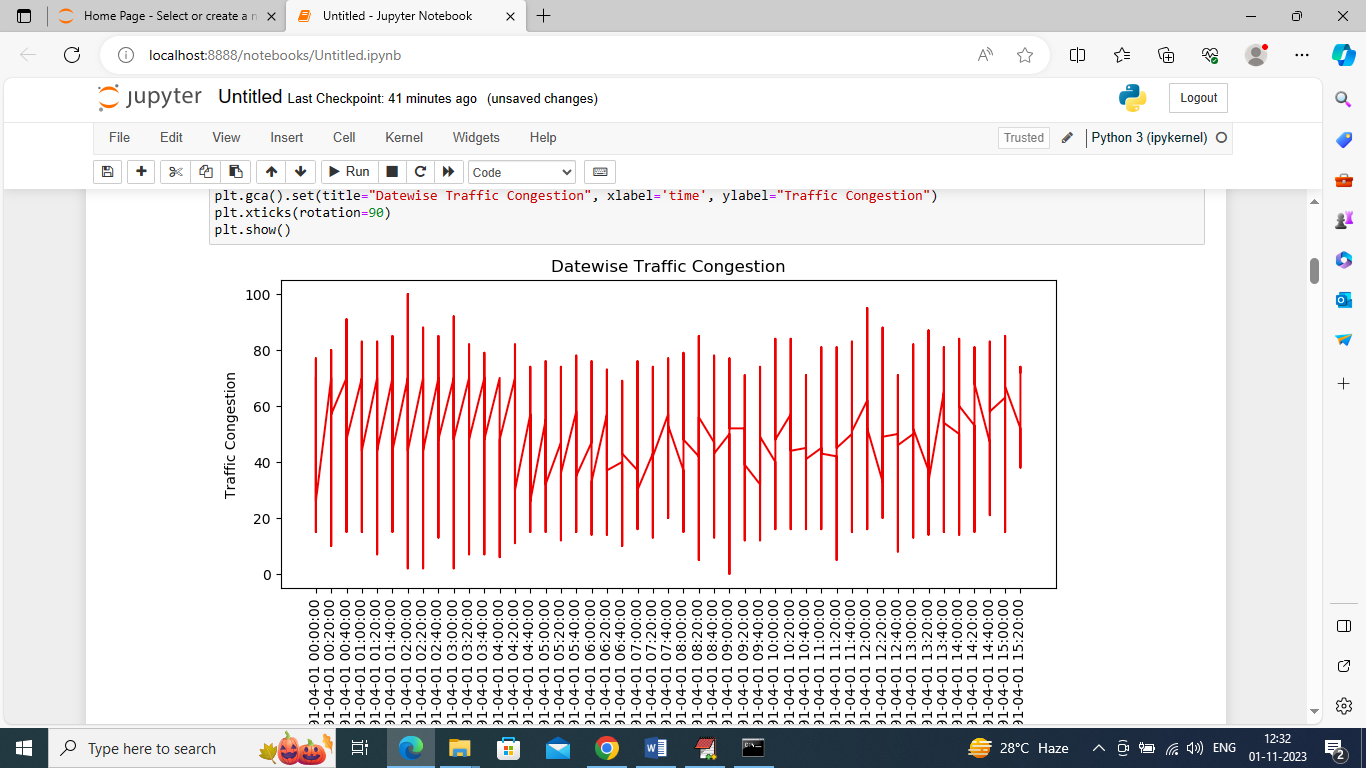
We have coded this project using JUPYTER notebook and below are the code and output screens with blue color comments



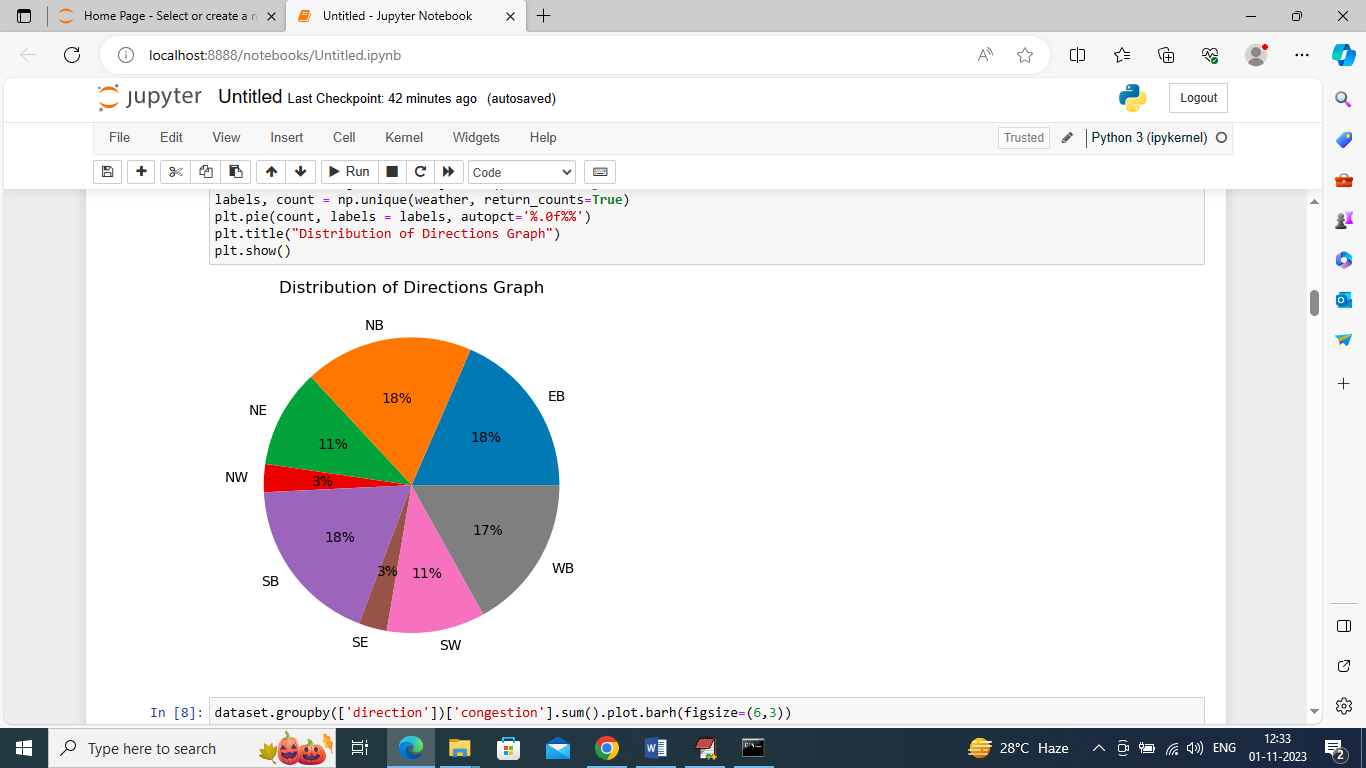
In above screen we are importing require python classes and packages



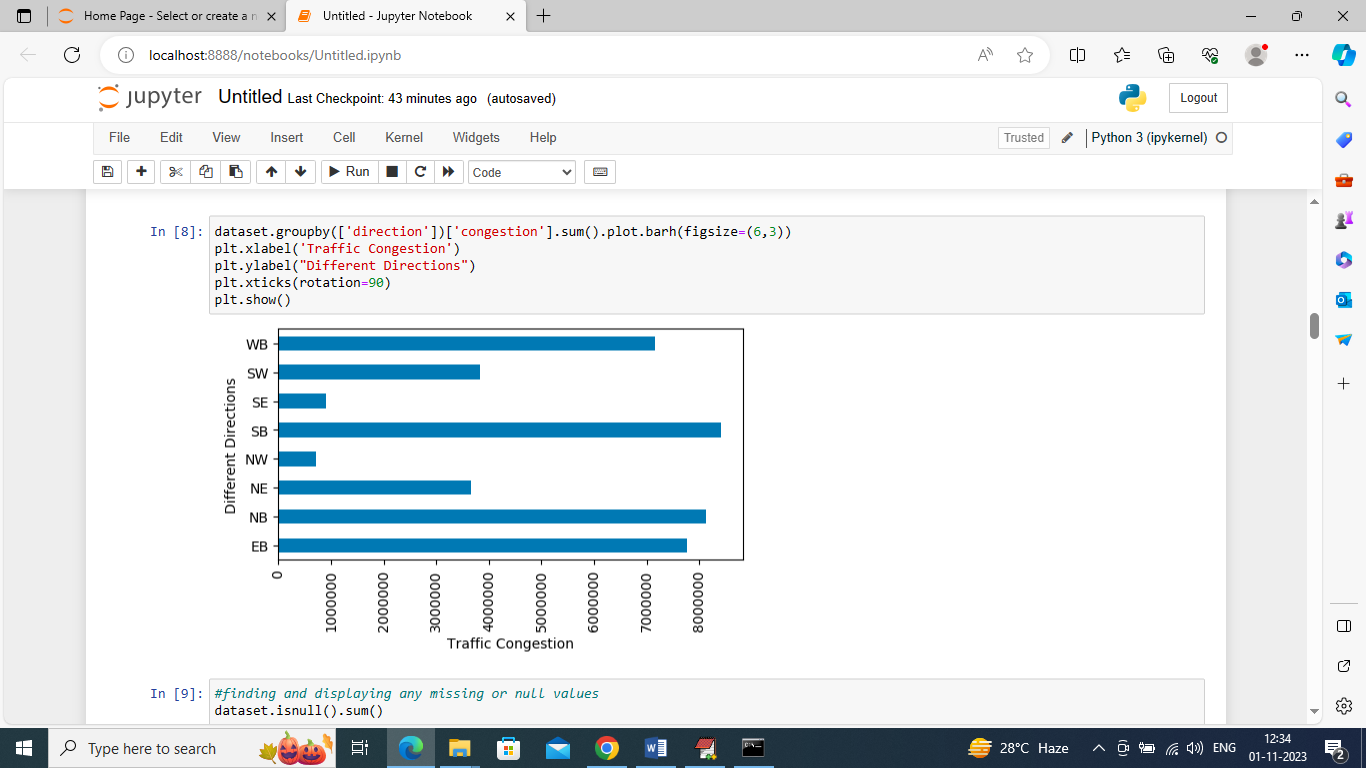
In above screen loading and displaying dataset values



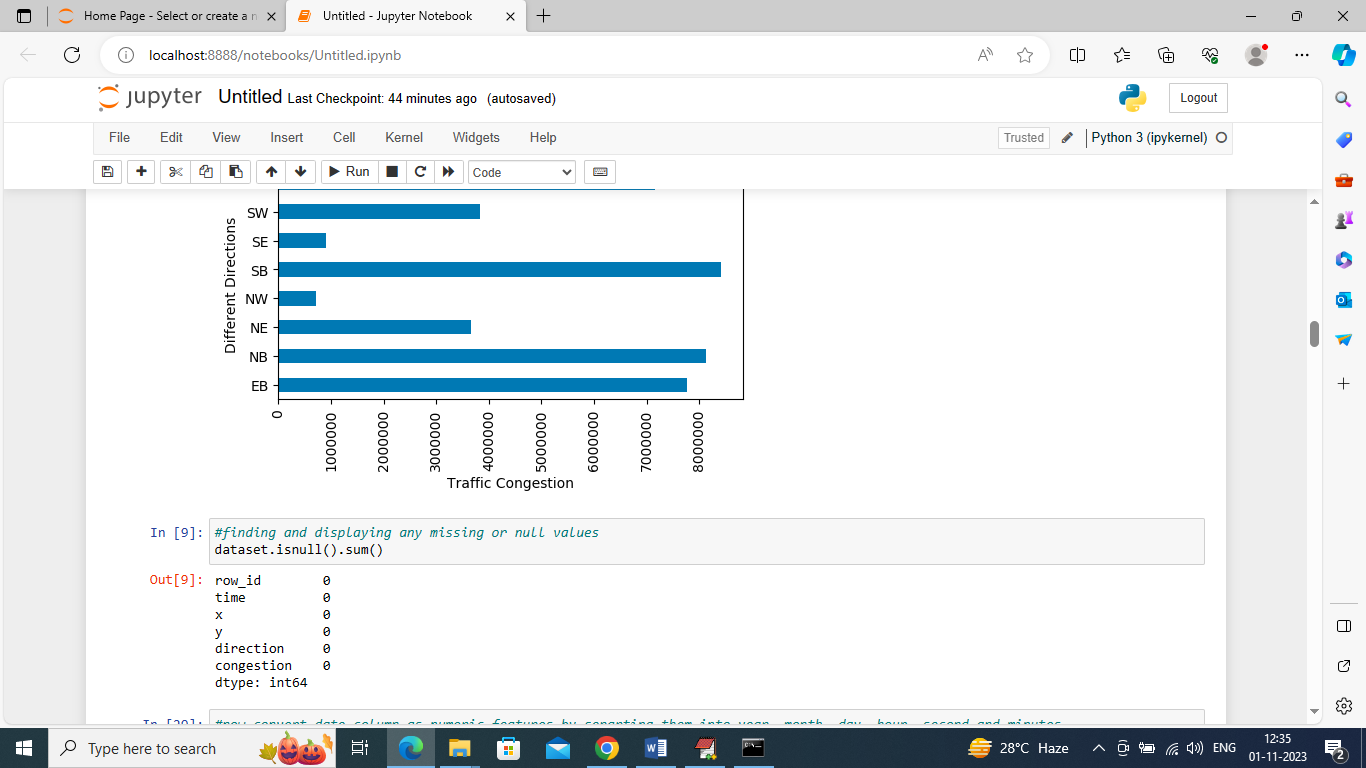
In above graph displaying traffic congestion on different dates where x-axis represents Date and y-axis represents traffic congestion



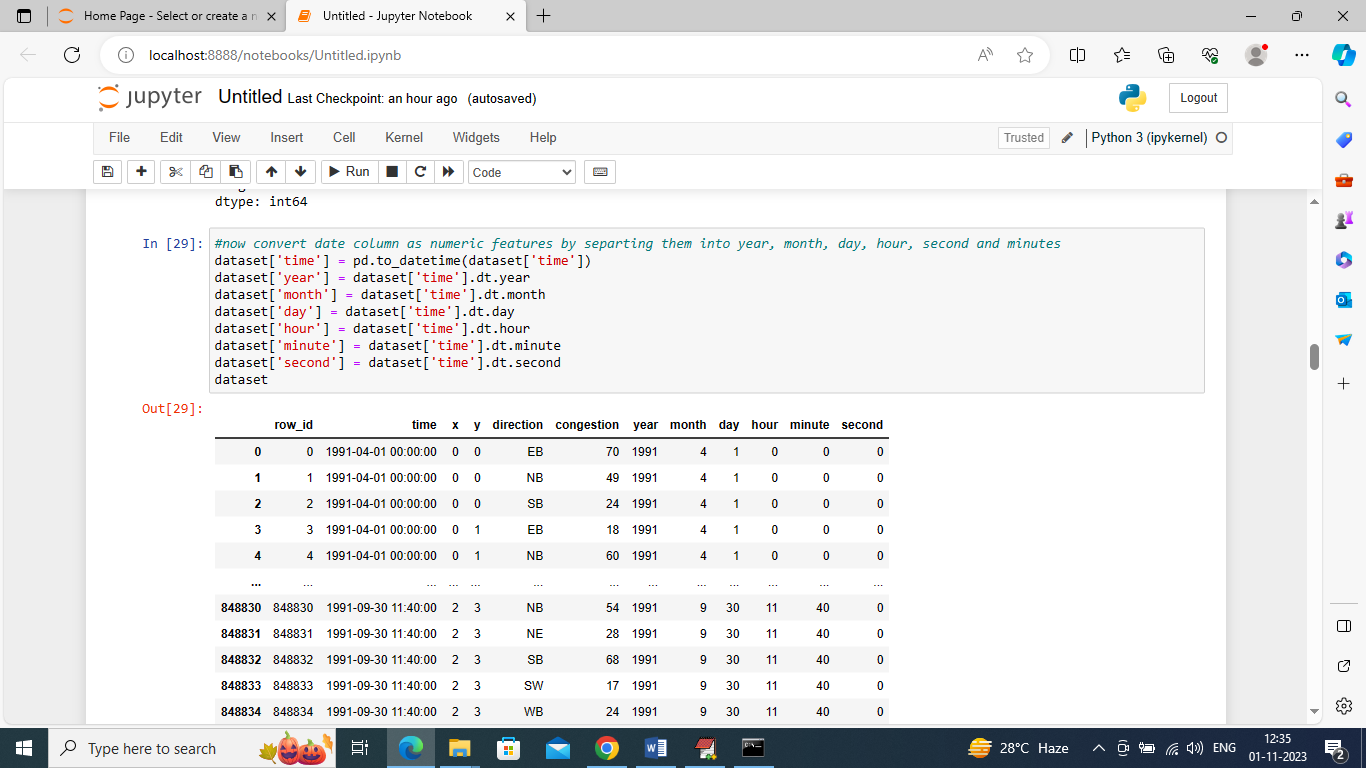
In above graph we are finding percentage of different directions or route exists in the dataset



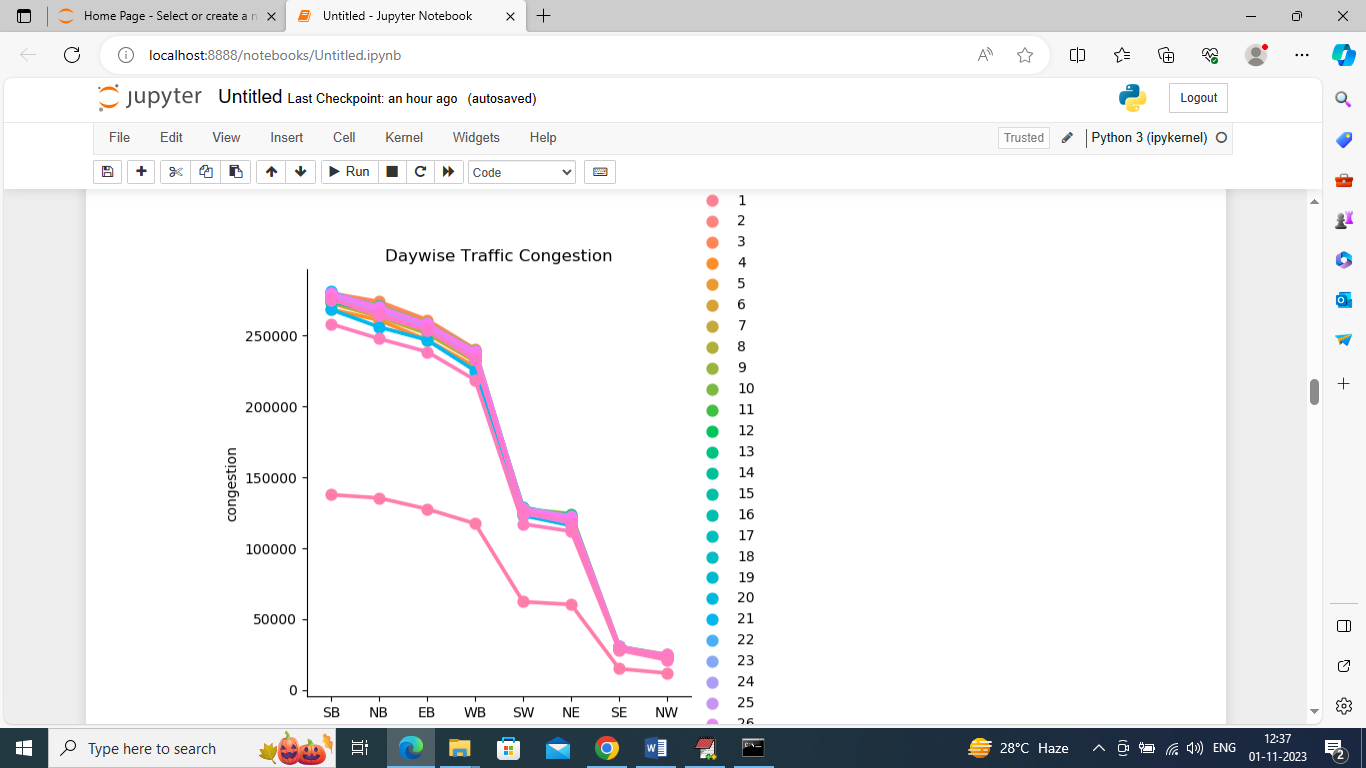
In above graph we are finding sum of traffic exists in each direction where x-axis represents traffic count and y-axis represents direction



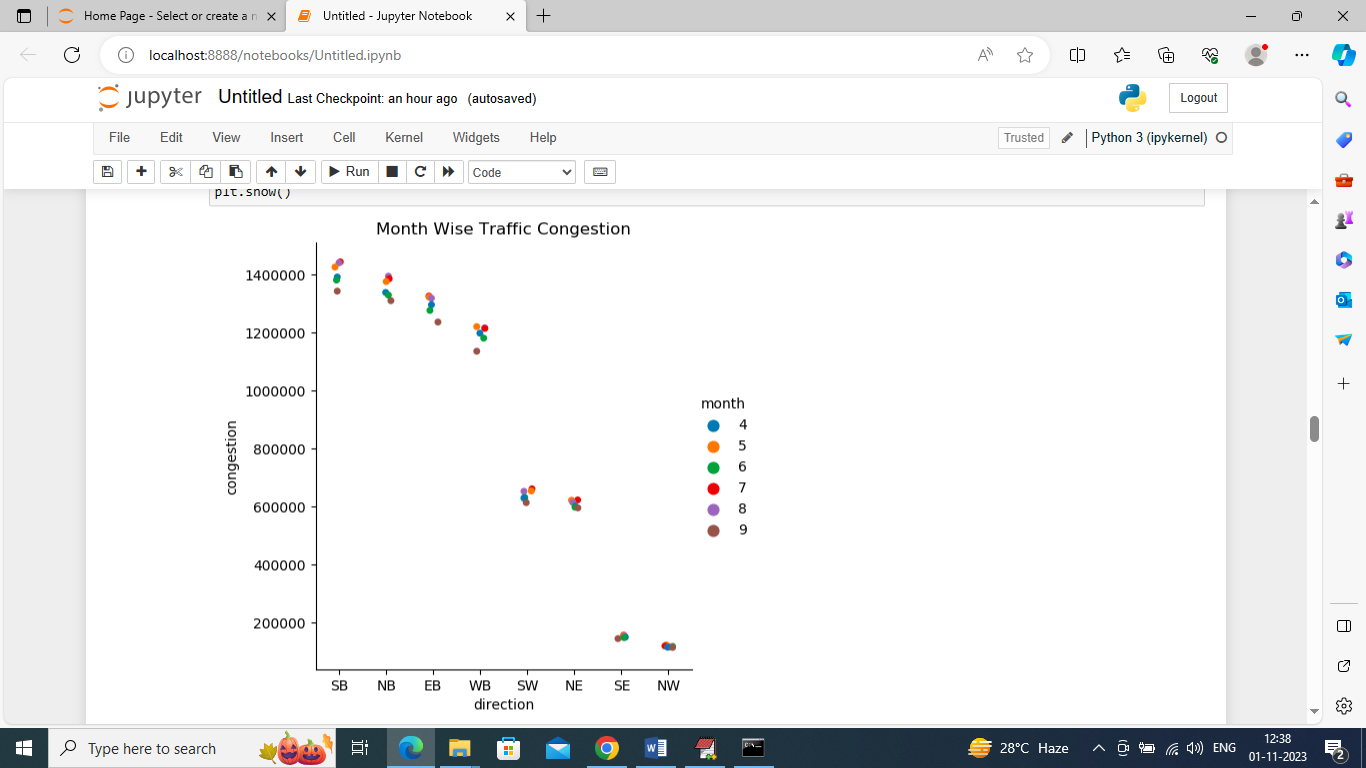
In above screen we are finding weather dataset contains any missing or null values but this dataset has no missing values



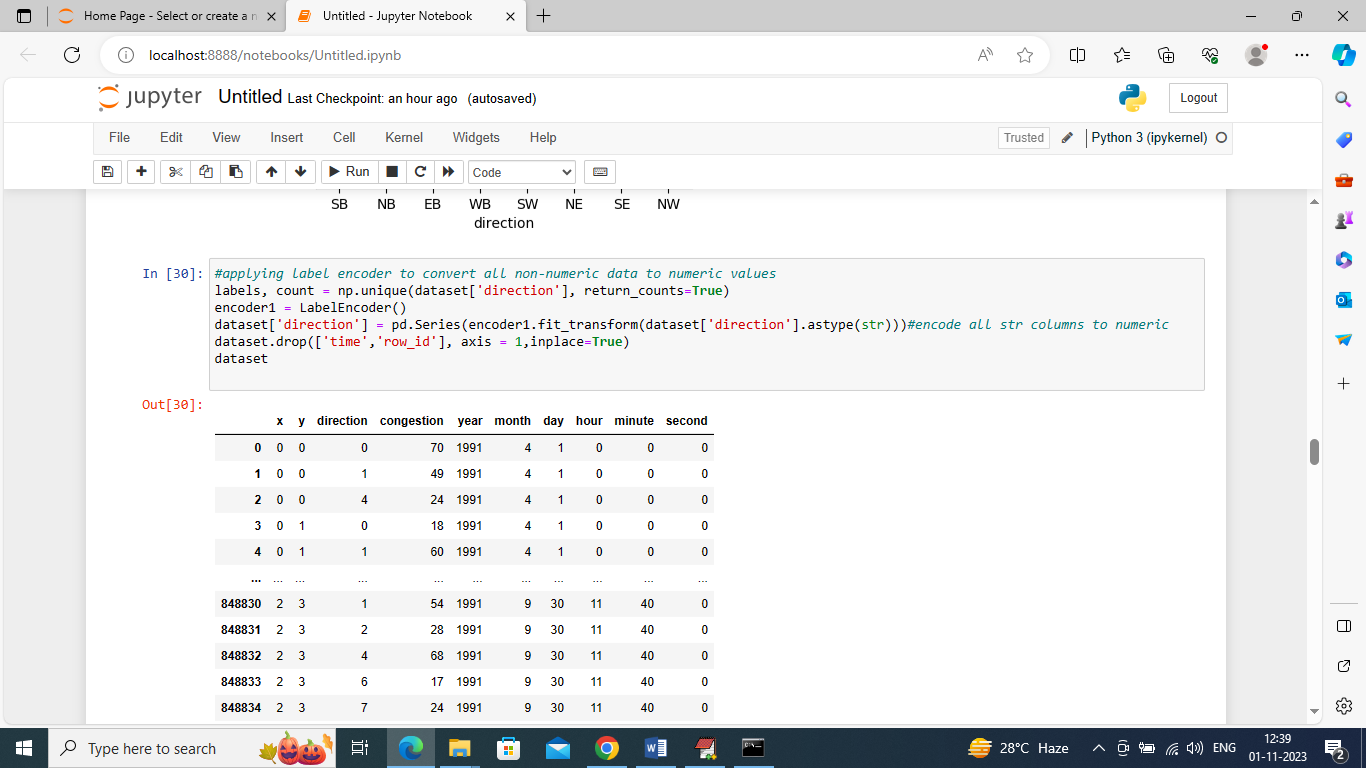
In above screen we are processing dataset to convert date into Day, Month and Year format so we can analyze traffic day or month wise and in above output we can see now dataset has day, year and month columns



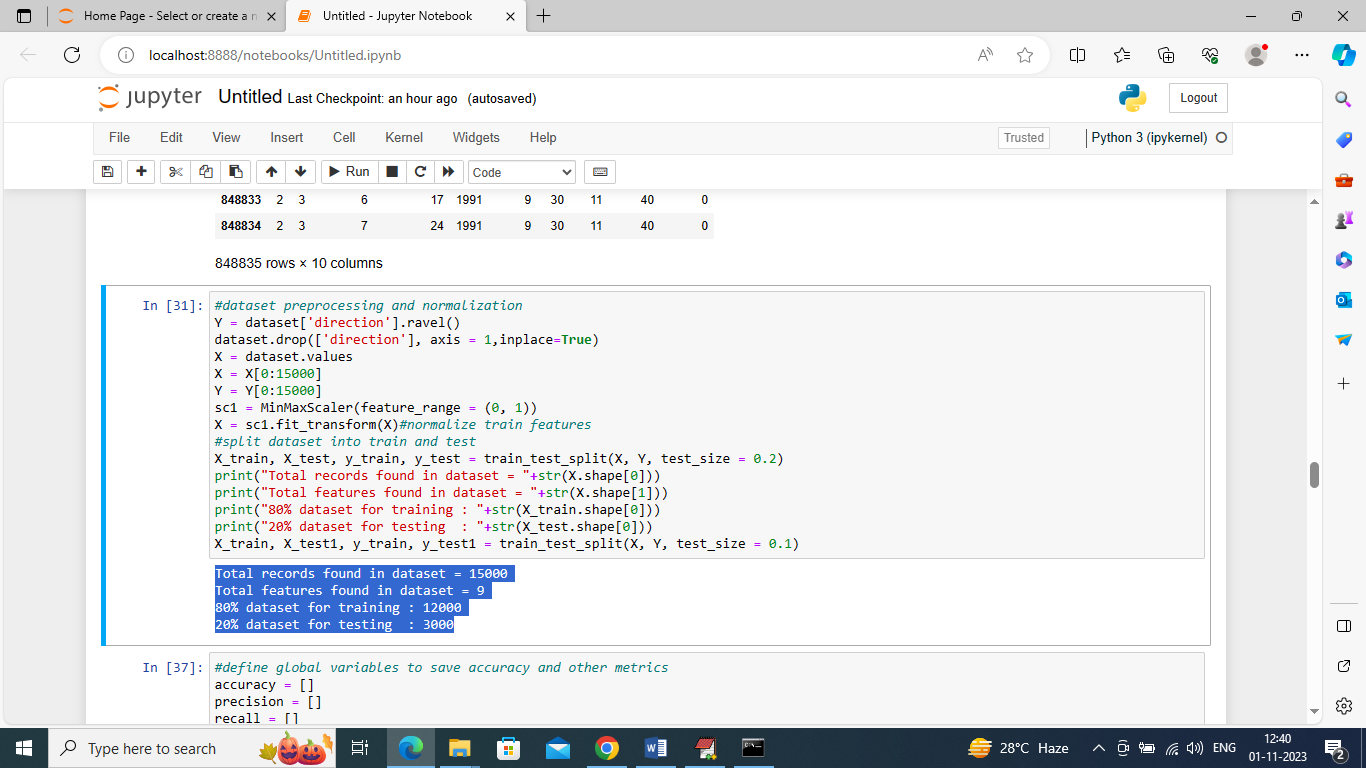
In above graph we are finding traffic day wise and each different color line represents different days traffic where x-axis is the direction and y-axis is the traffic congestion count



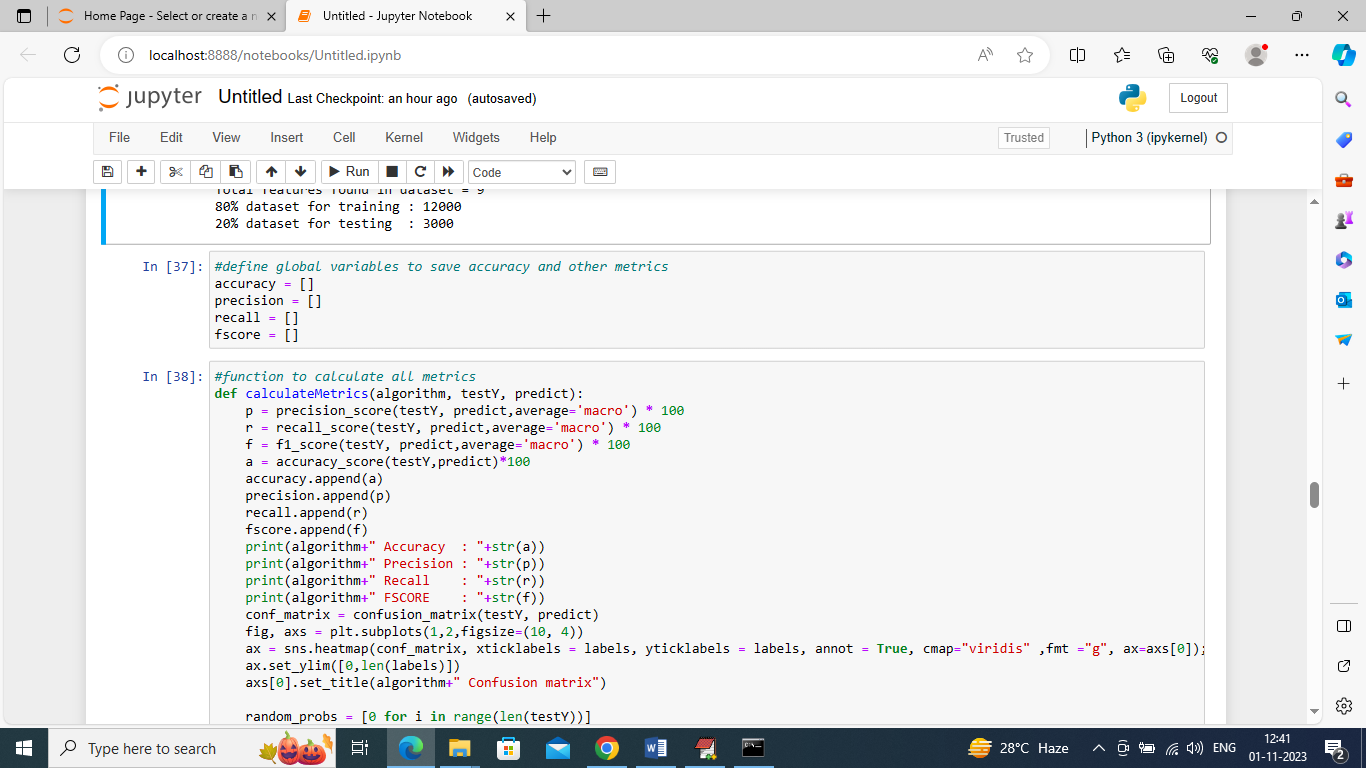
In above graph we are finding month wise traffic in different directions where x-axis represents direction and y-axis represents traffic congestion and different color dots represents months.



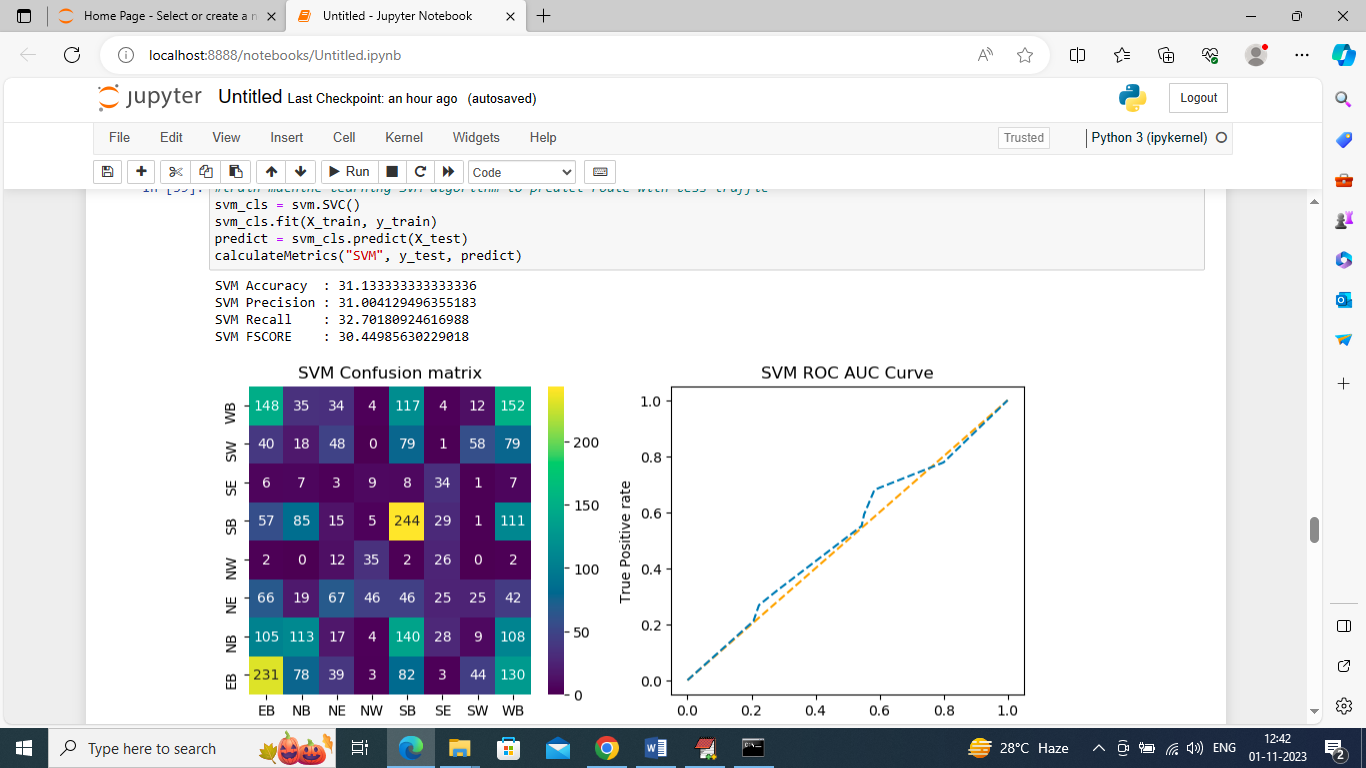
In above screen different directions are in string format so we have converted them into numeric format as ML will take all labels are numeric format so we have converted them into numeric labels



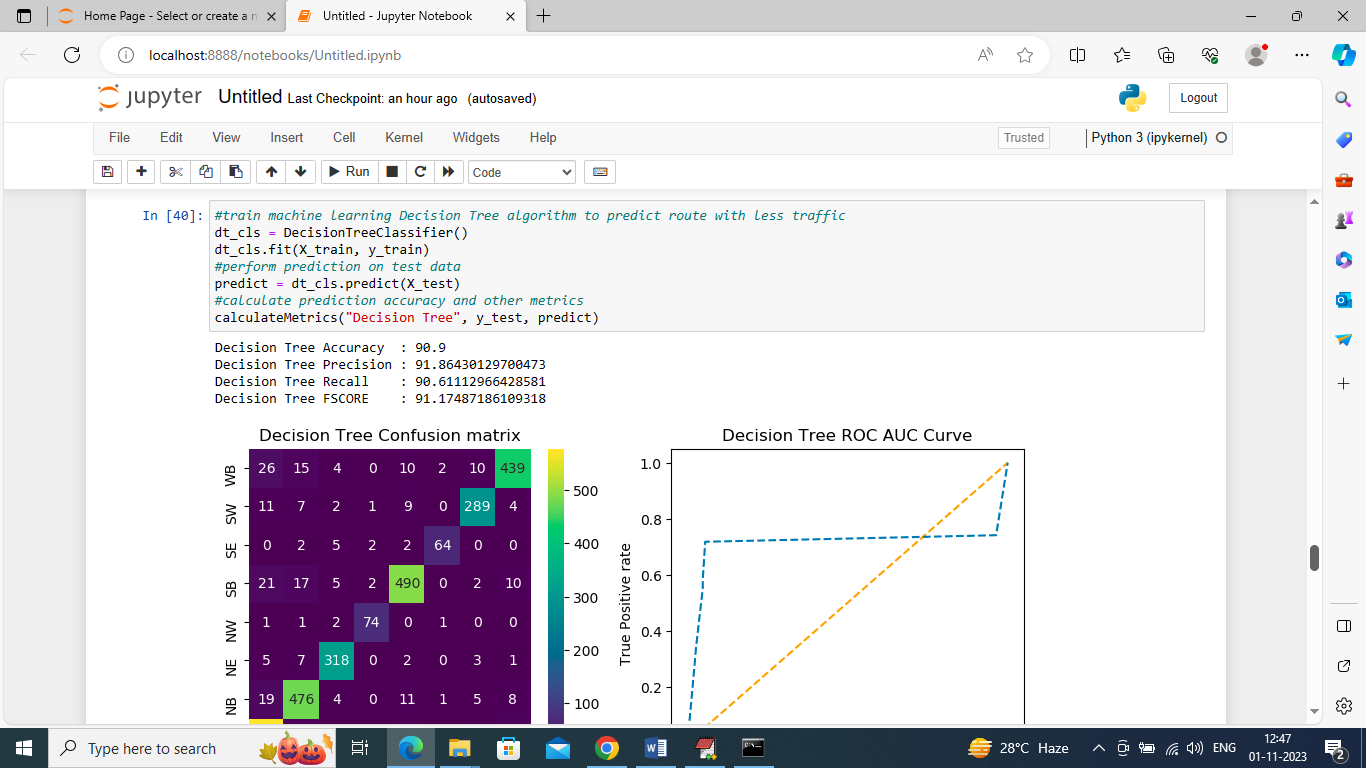
In above screen we are processing dataset such as normalization and then splitting into train and test where application using 80% dataset for training and 20% for testing and in blue color we can see train and test size



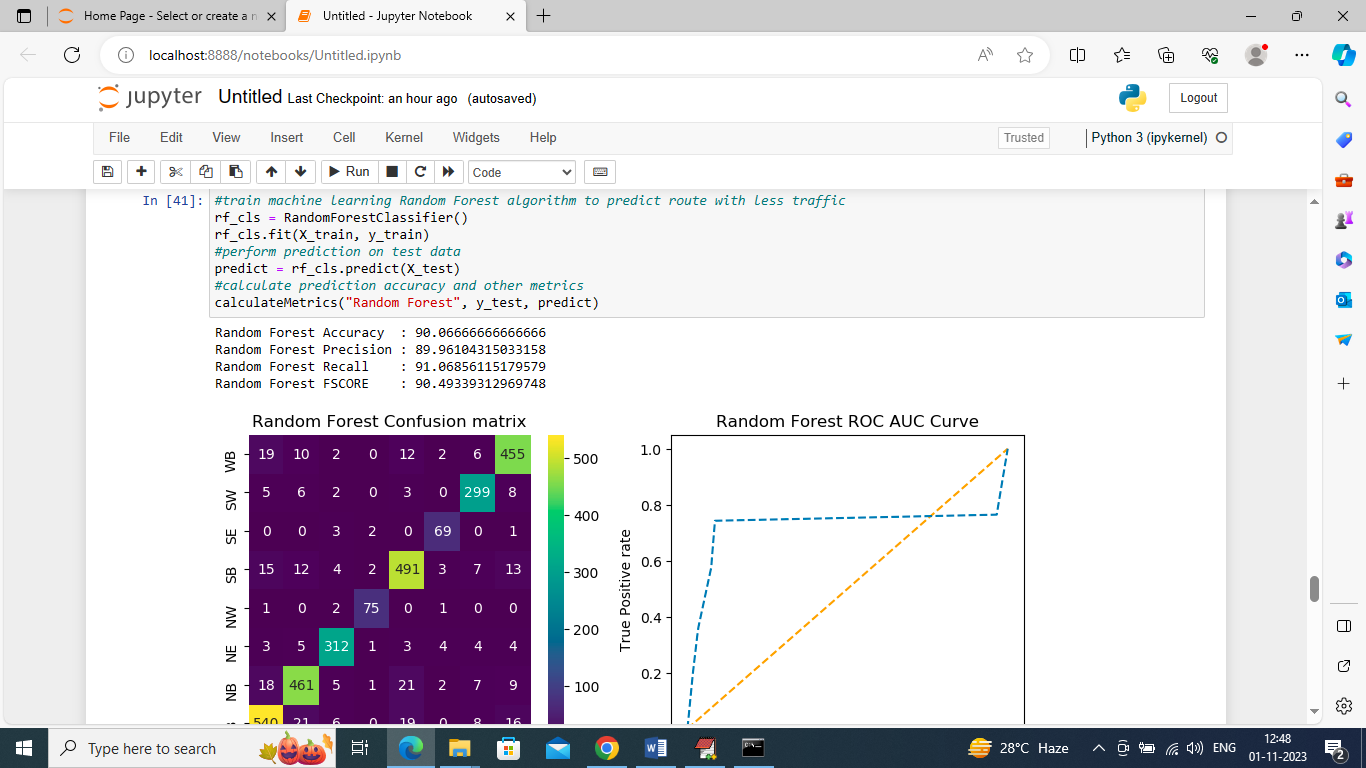
In above screen defining function to calculate accuracy and other metrics



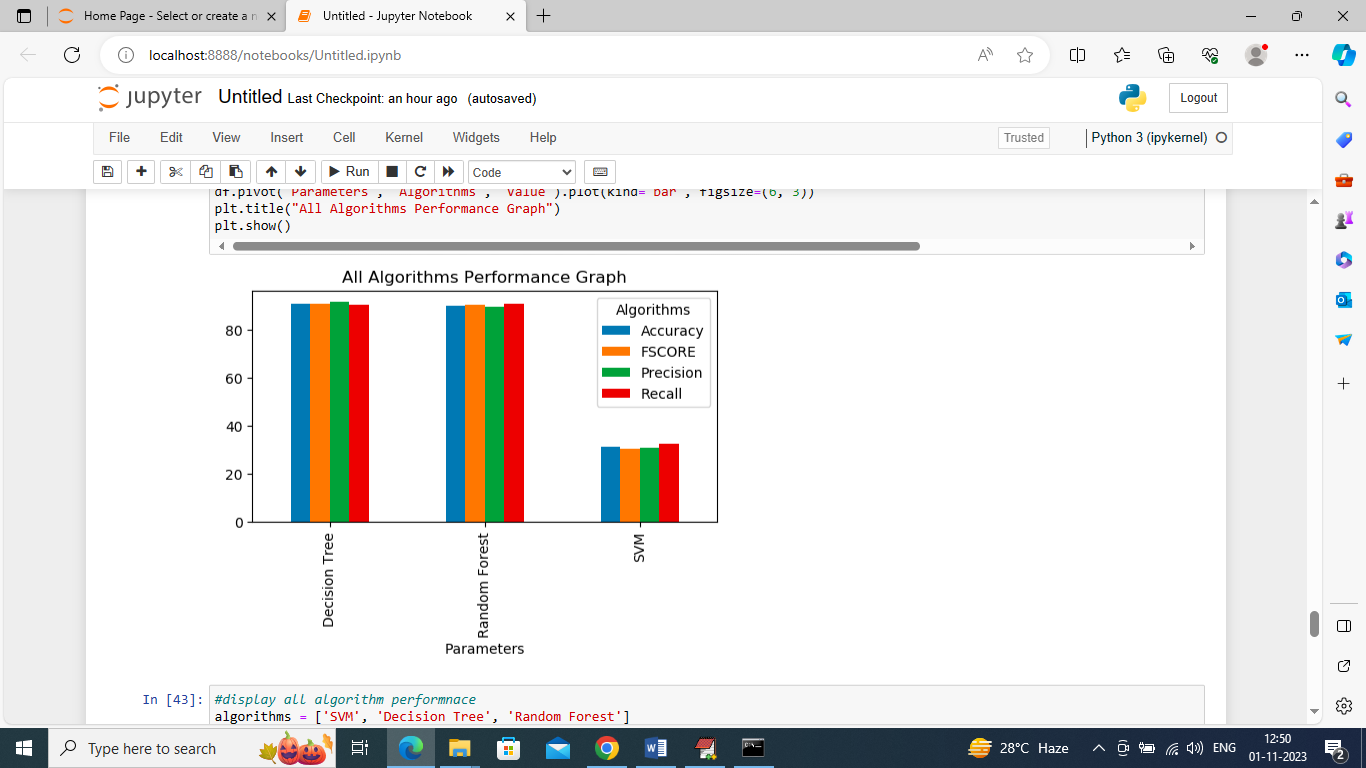
In above screen training SVM algorithm and after training SVM got 30% accuracy and can see other metrics also. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and all boxes in diagnol represents correct prediction count and remaining boxes represents incorrect prediction counts and from above graph we can notice SVM predicted many records incorrectly. In ROC curve graph x-axis represents False Prediction and y-axis represents True Predictions and if blue line comes on top of orange line then predictions are correct and if goes below orange line then predictions are incorrect and in above graph we can see only few predictions are correct



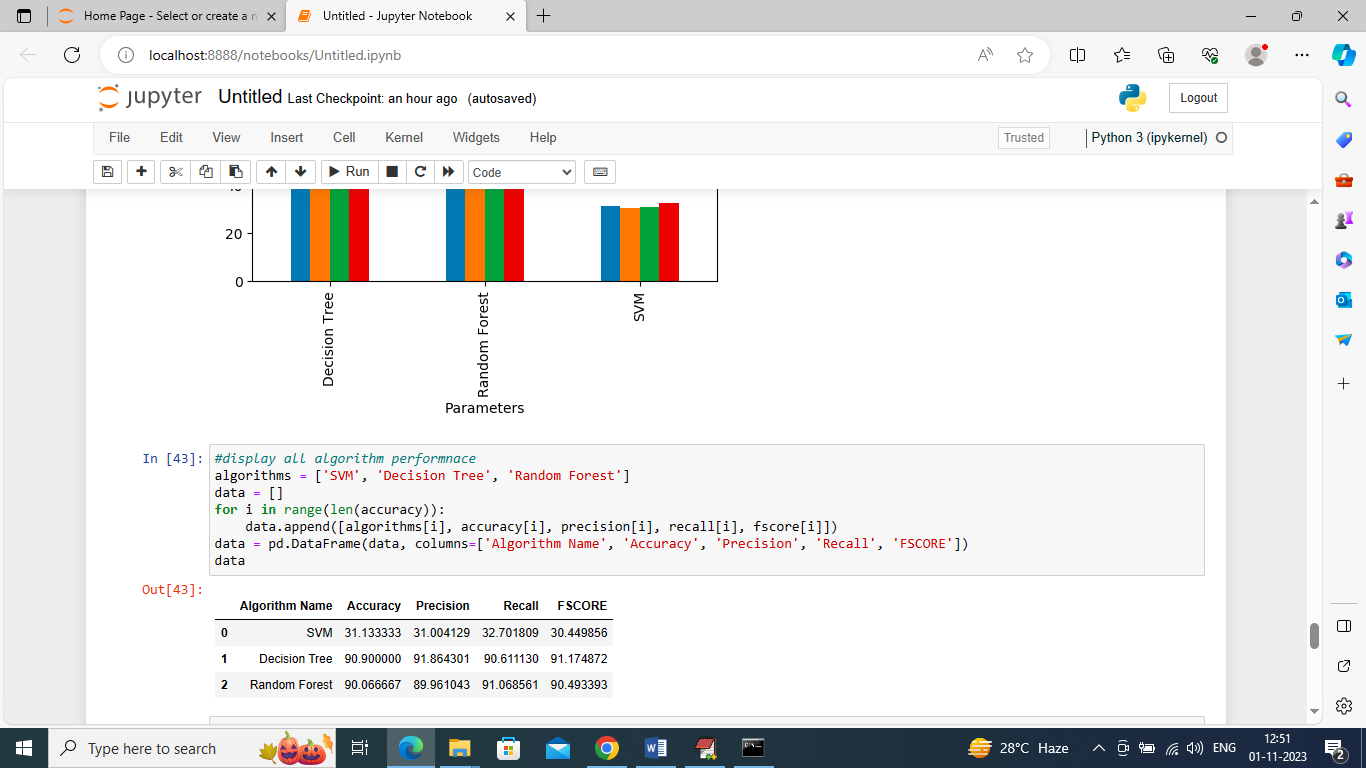
In above sceen training decision tree and it got 90.9% accuracy and can see other metrics and results graph



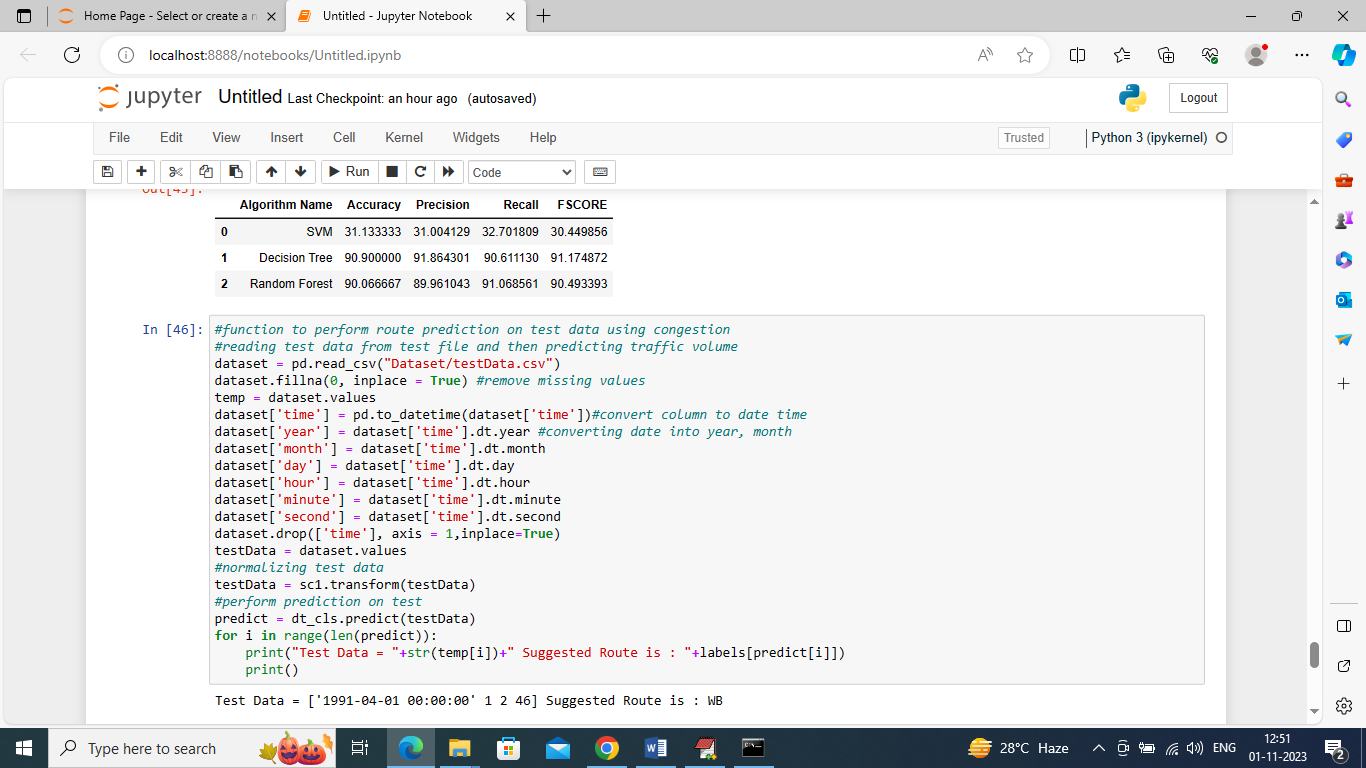
In above screen training Random Forest and it got 90.06% accuracy and can see other metrics also and in above confusion matrix graph in diagnol we can see many records are correctly predicted and in all blue boxes only few are incorrectly prediction. In ROC graph also we can see only few predictions are incorrect



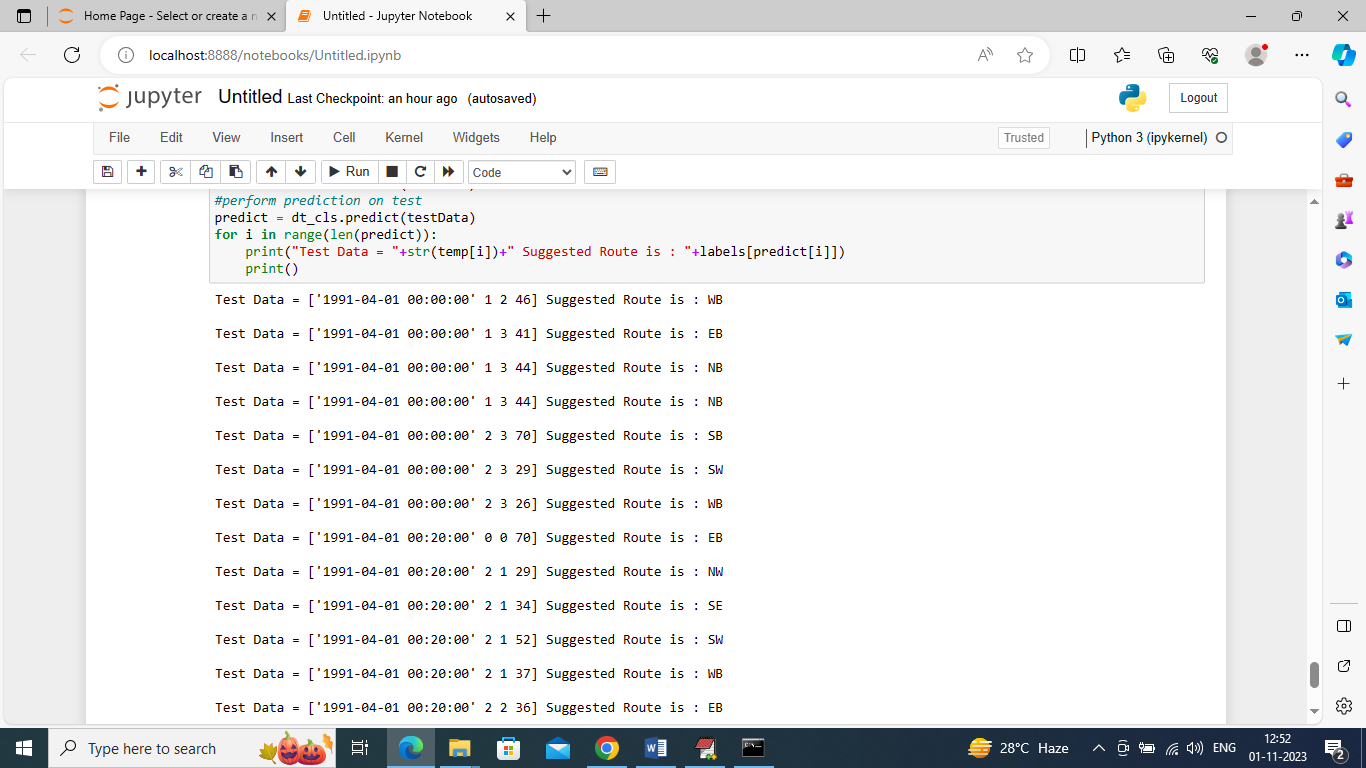
In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different color bars and in all algorithms Random Forest and Decision Tree work best



In above screen can see all algorithm performance in tabular format



In above screen we are defining function to read TEST data and then perform route prediction on test and after execution above block will get below output



In above predictions in square bracket we can see the TEST data where last value is traffic congestion and based on that congestion we can see suggested Route as WB or NB or SE etc.

Note: the direction of travel of the roadway. EB indicates "eastbound" travel, for example, while SW indicates a "southwest" direction of travel.