**Traffic Flow Prediction Dataset – Group 51**

**Data analysis and Pre-processing:**

We have analyzed data and we have received data from 36 sensors. Each sensor is reporting reading at every 15 minutes i.e., Qtr. hour level which means we will have 96 observations in a day.

We have used “tra\_Y\_tr” in our prediction work which is training set output data from traffic flow for 36 locations in 1261 contiguous quarter-hours.

We were given .mat file and we have loaded this data in python using library SciPy and created data frame using pandas.

Data was already clean and it was ranging between 0 to 1 and all null values were already handled as we have checked data wfor null values data.isna().sum()

**Model building:**

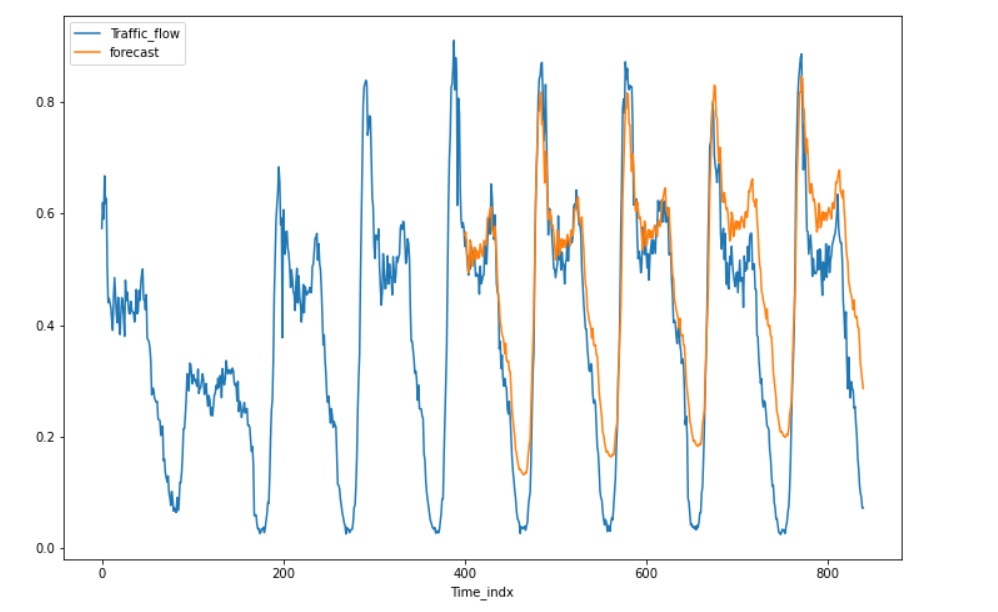
**Time Series data:**

a time series is a series of data points indexed in time order

**Forecasting on Time Series Data:**

In forecasting, we have many models that help us make predictions and forecast the values to fulfil our future aspects according to the situation’s demand. The examples of models can be **AR, MA, ARIMA, SARIMA, VAR, SARIMAX** etc.

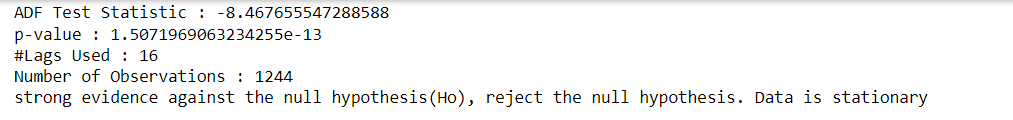
We have loaded data in data frame and plotted it. From graph we can see that seasonality is present in data, we have chosen SARIMAX model and ran it on one of sensor. From below on prediction, we can see predicted observations are best suited so we have decided to use this model in our prediction of traffic flow



Here in the graph, we can see the results: the forecasting line is almost lying on the given values for this model. We didn’t even require the differencing method. Using this model now, we can predict the future values too.

We have also run adfuller test to check data is stationary data or not.

A stationary data is one where mean and variance at any time remains constant.

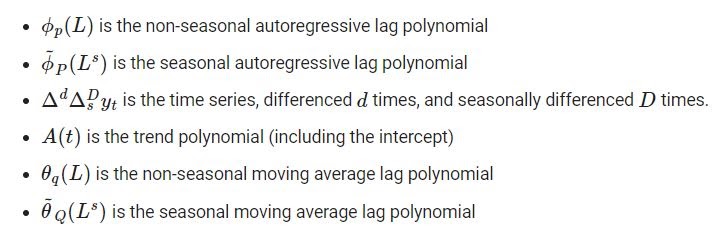


**SRIMAX Model**: In any time-series model, accurate prediction depends on how we have handled Factors like Trend, seasonality and any unpredicted events (i.e., war, pandemic etc.)

SRIMAX stands for *Seasonal Auto-Regressive Integrated Moving Average with exogenous* factors, is the updated version of ARIMA Model. We can say that SRIMAX is equivalent to SARIMA and auto ARIMA with additional feature of handling external effects

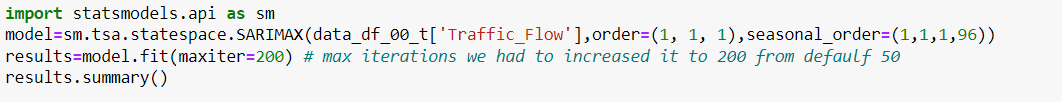
Mathematical representation of SARIMAX model is





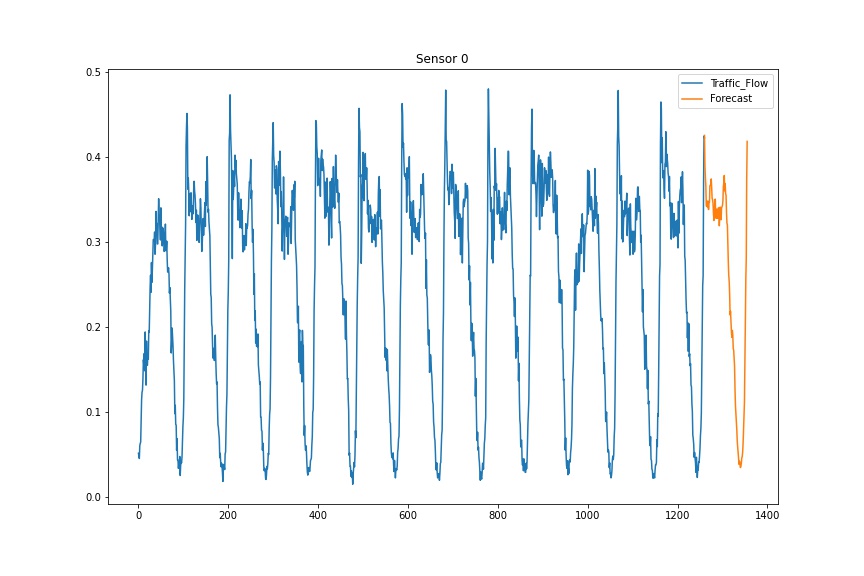
We have chosen order (p, d, q) as (1,1,1) based on best fit values.

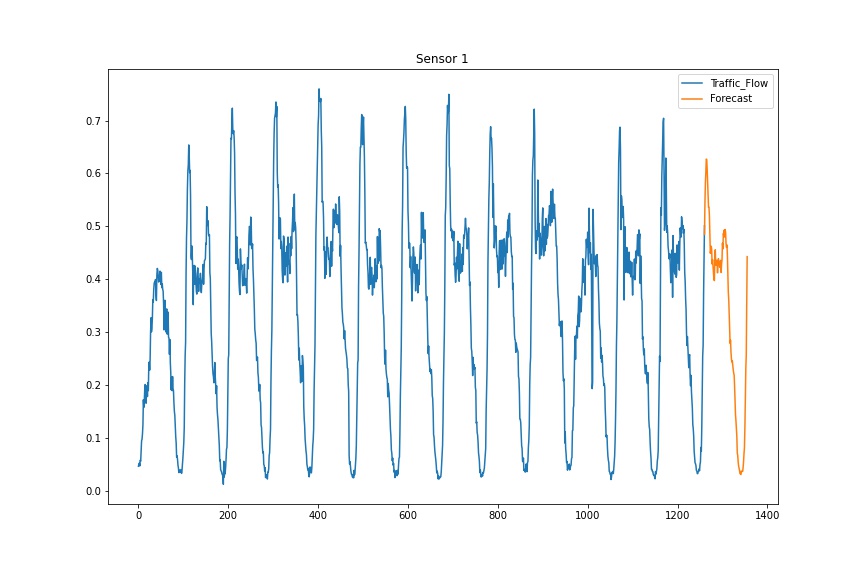
Where p is the number of autoregressive terms, d is the number of nonseasonal differences needed for stationarity, and. q is **the number of lagged forecast errors in the prediction equation**

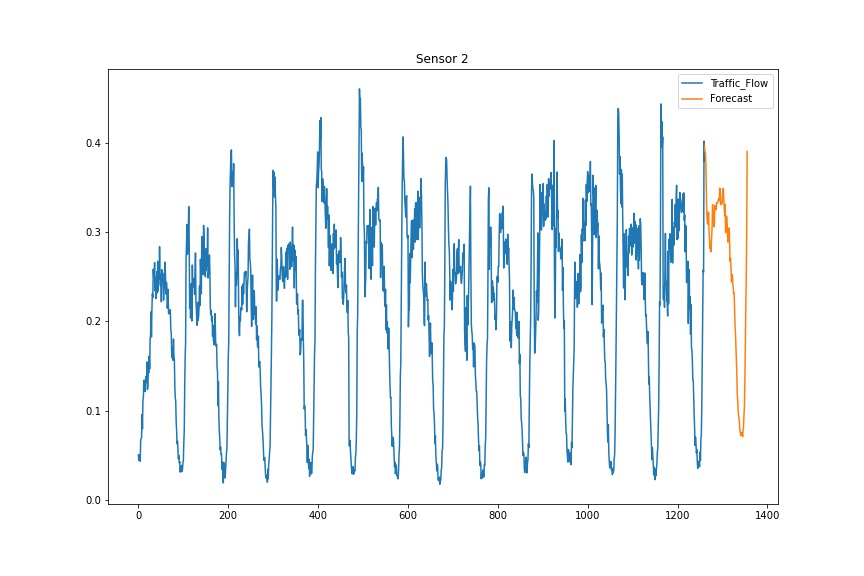


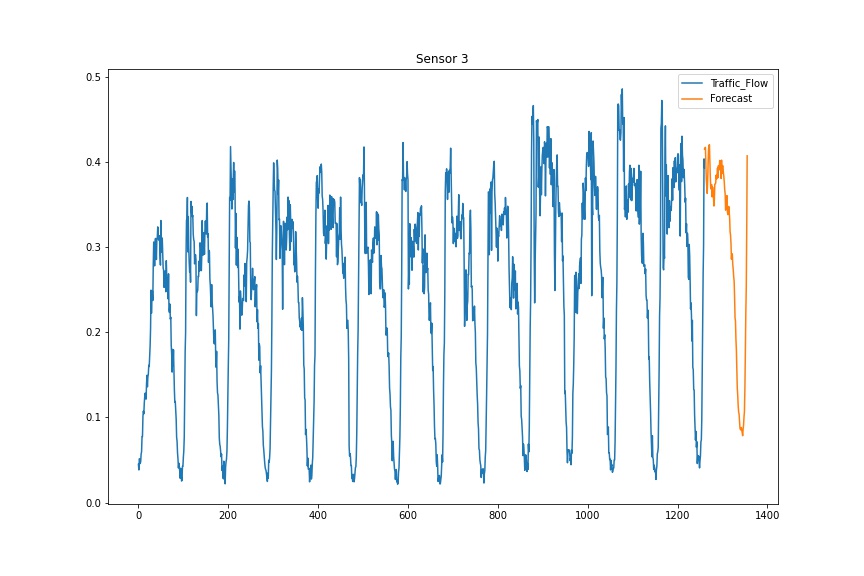
**Report/presentation of Outcomes:**

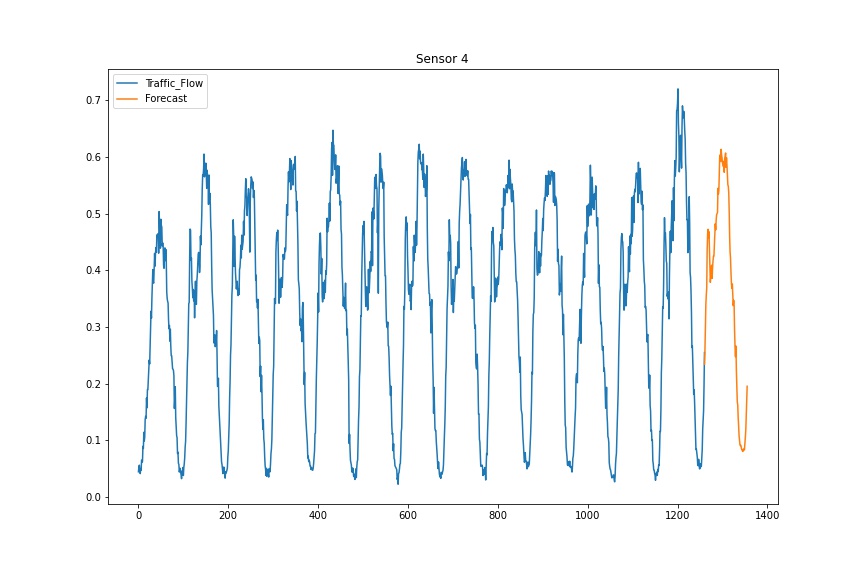
We have forecasted for next 24 hours and we can see that our forecast is almost matching with past data and hence we can see that our forecasting is good for use

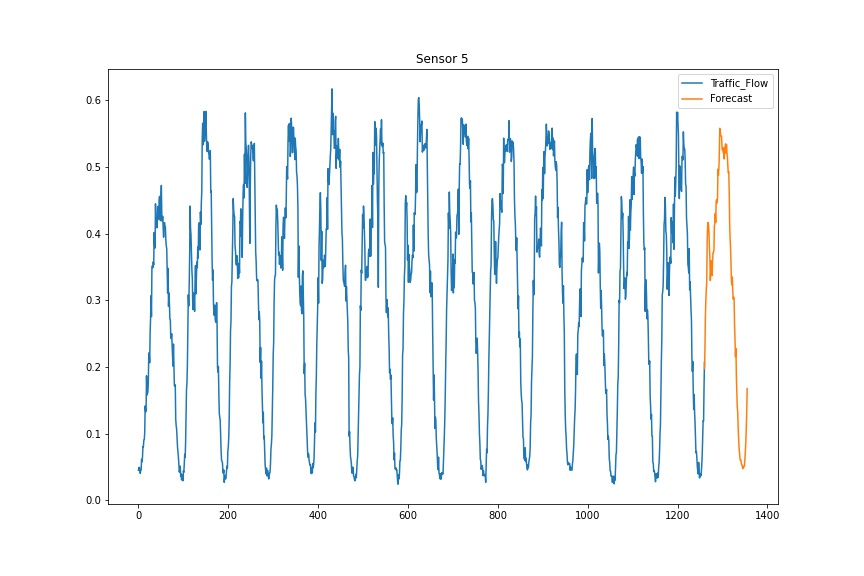


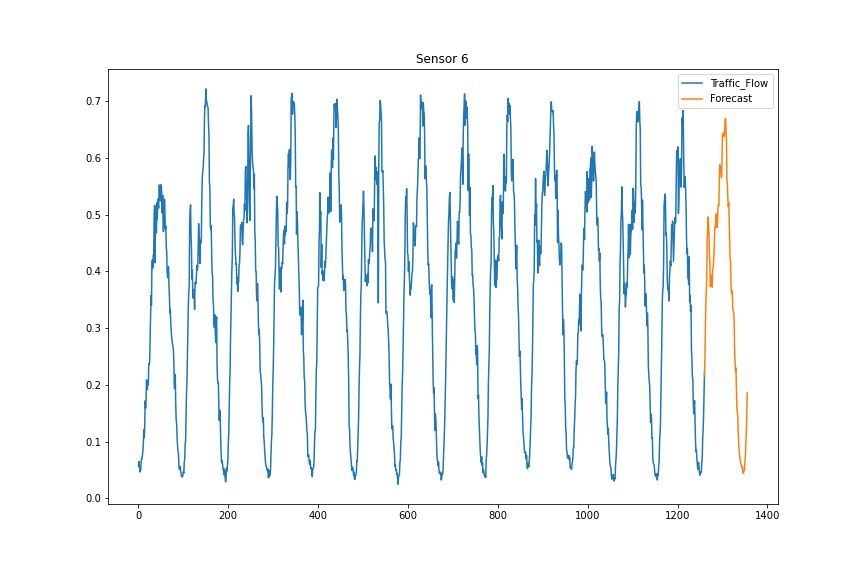


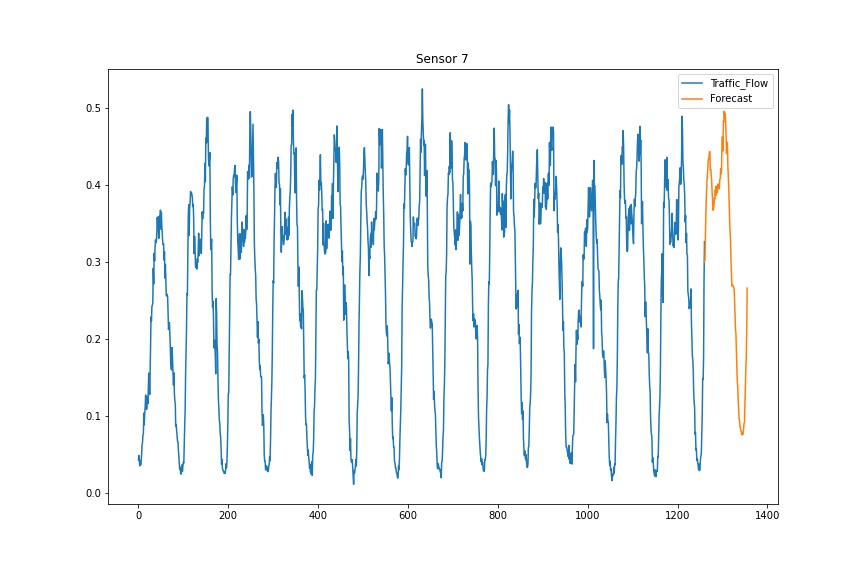


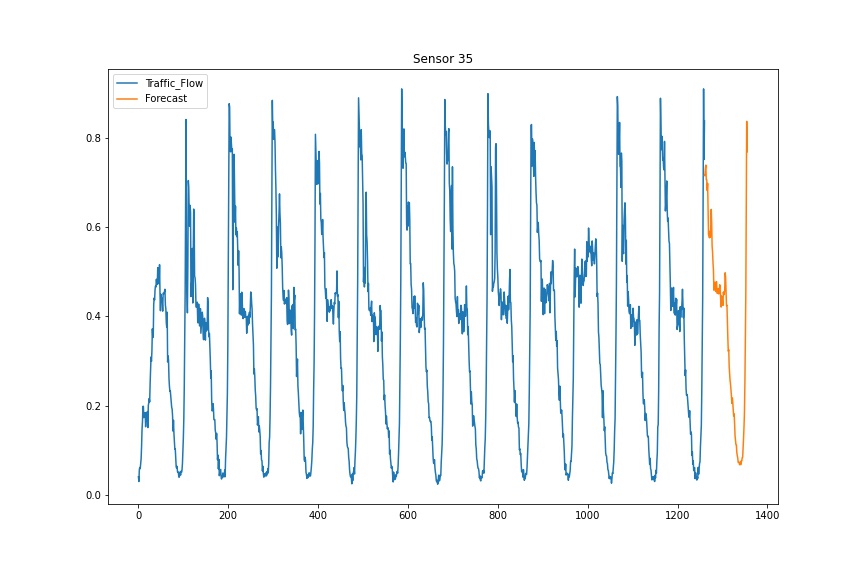
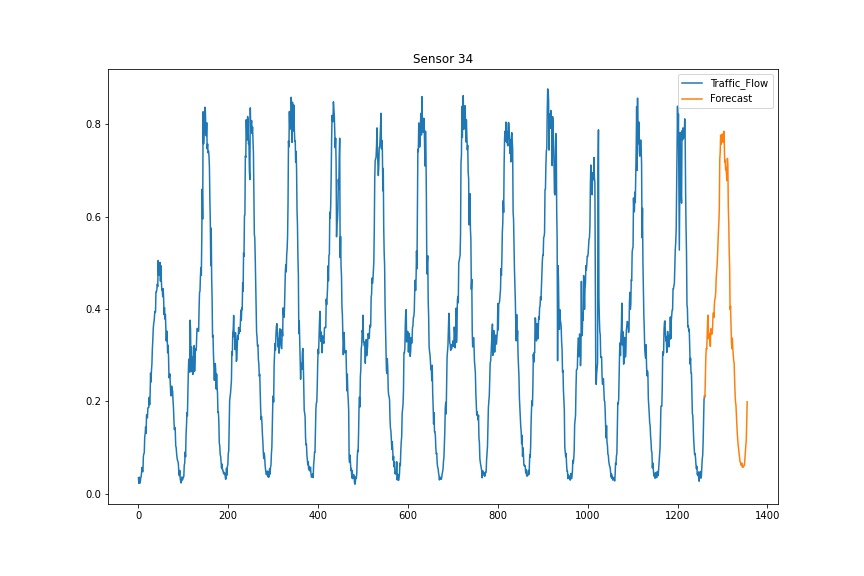
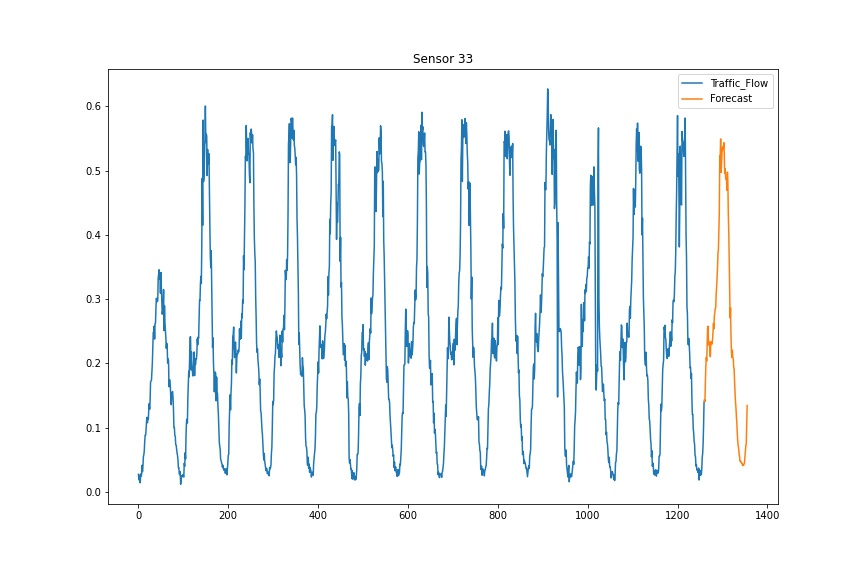
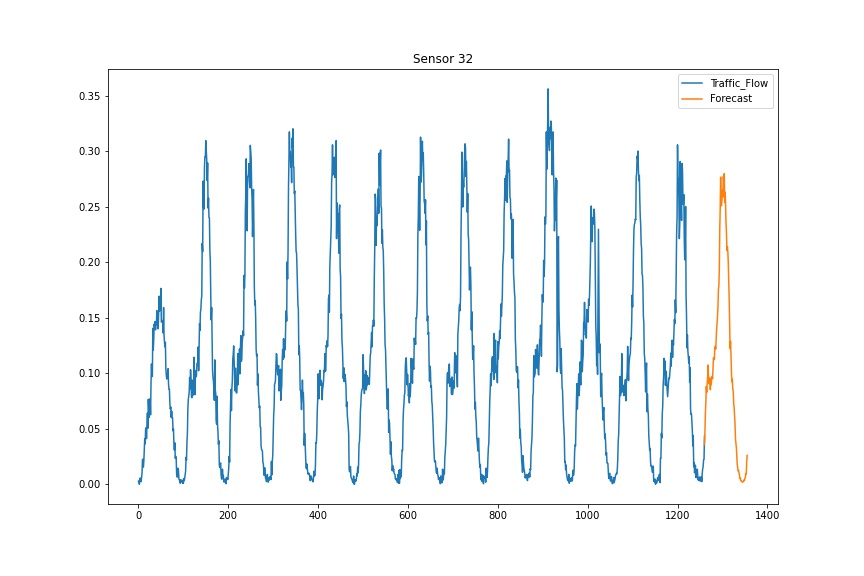
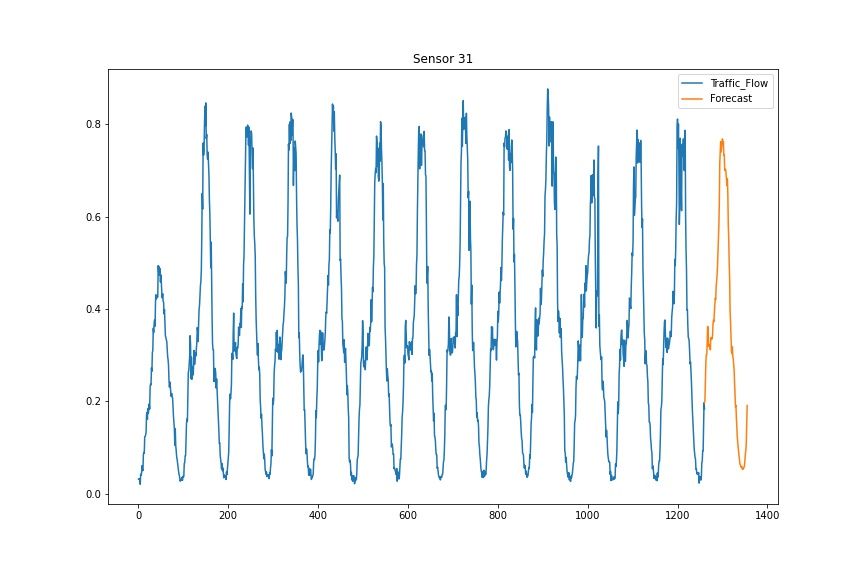
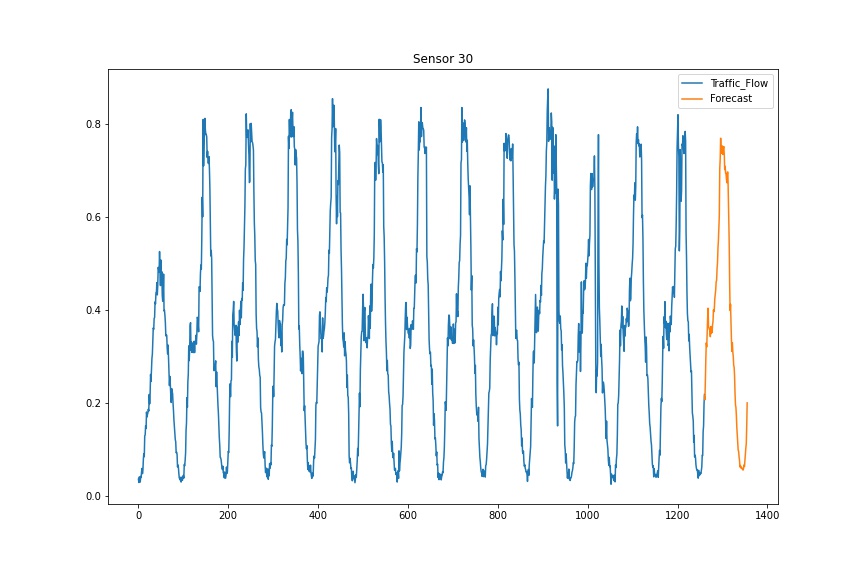
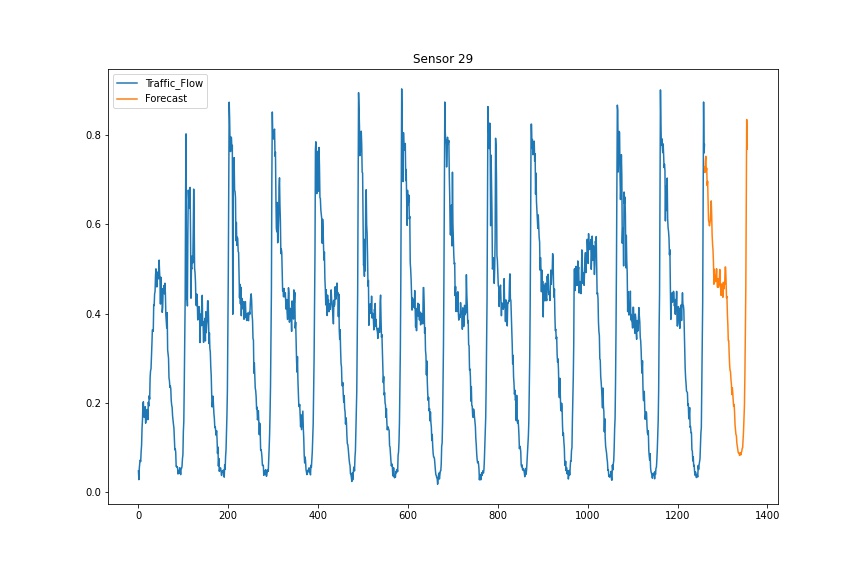
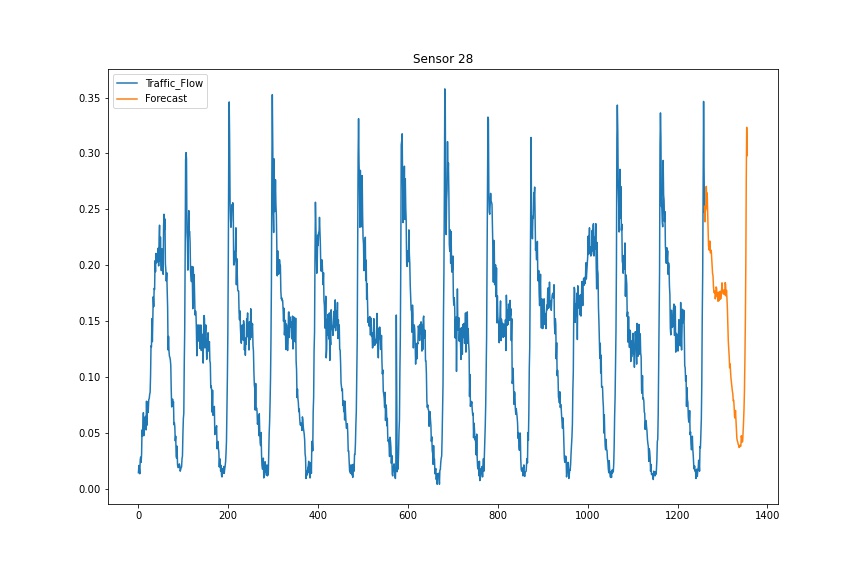
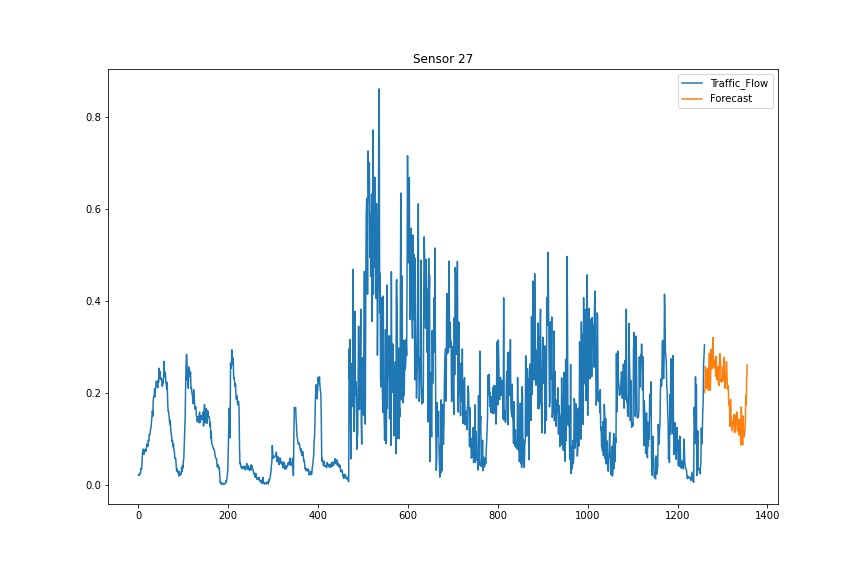
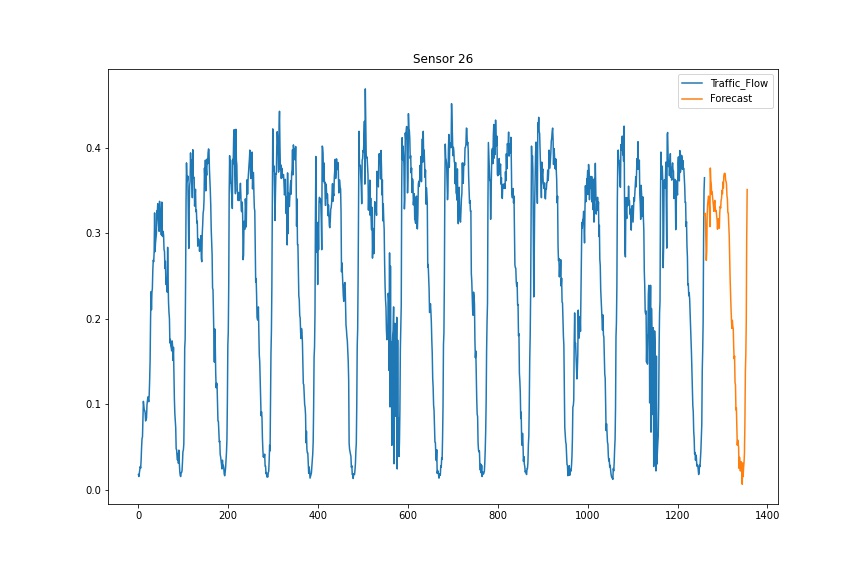
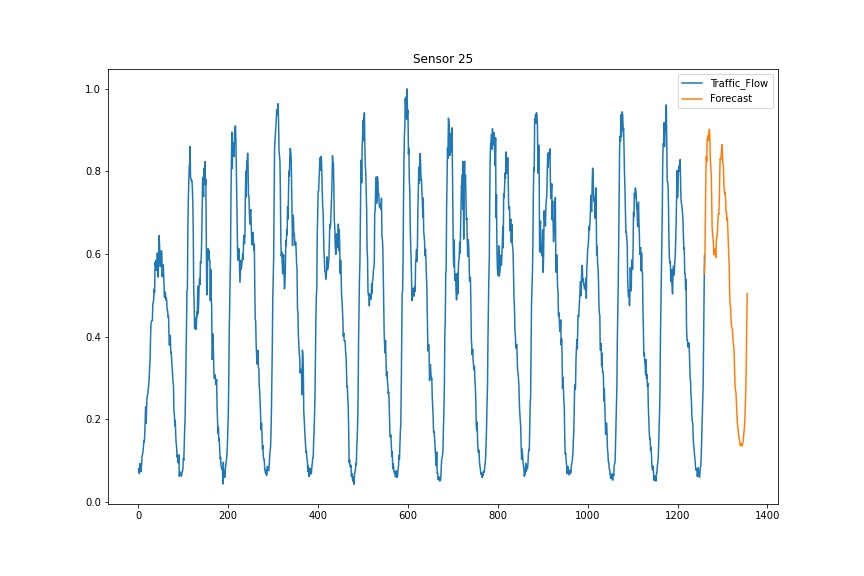
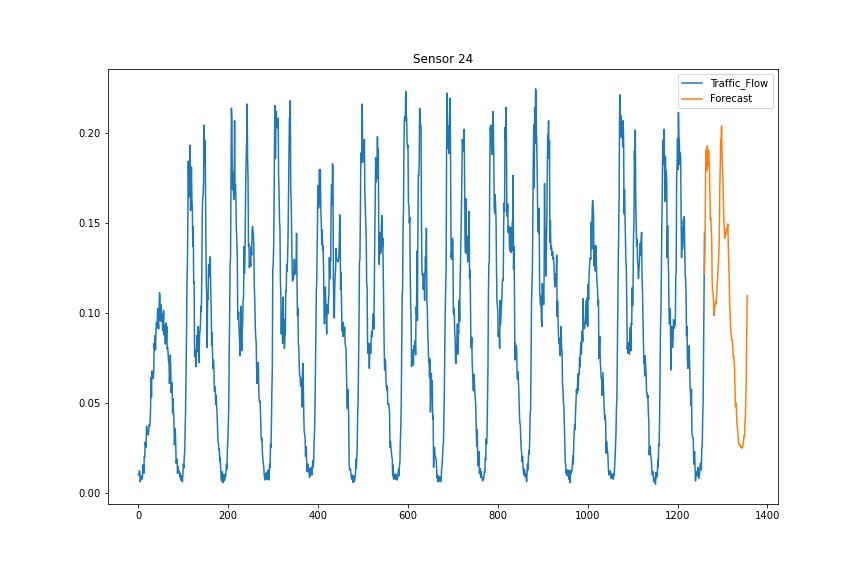
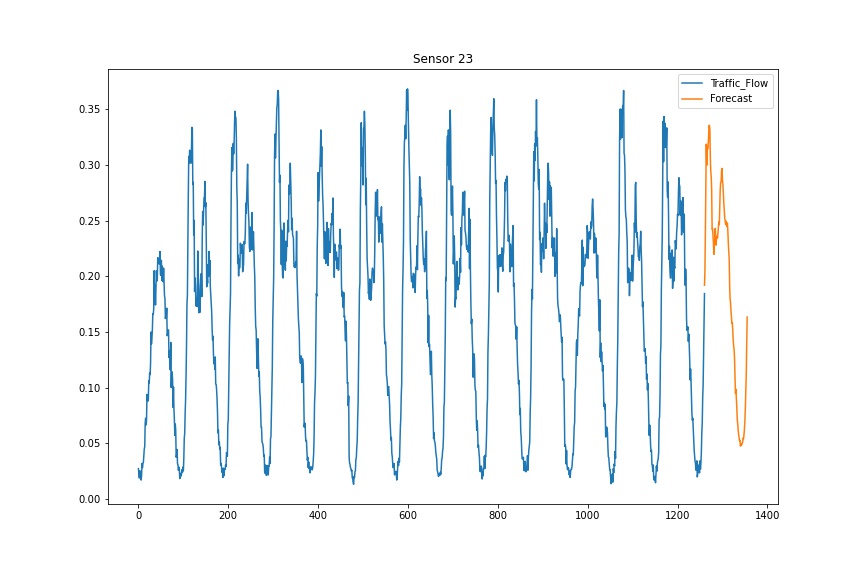
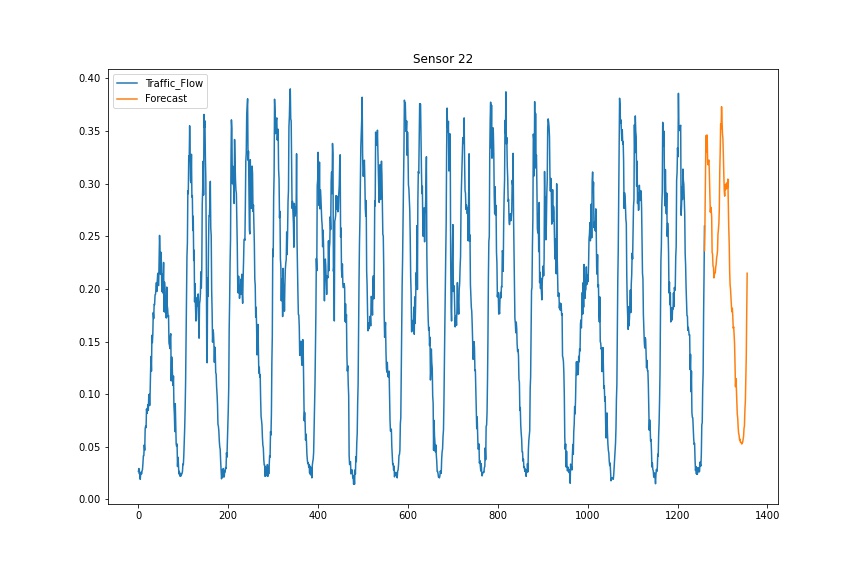
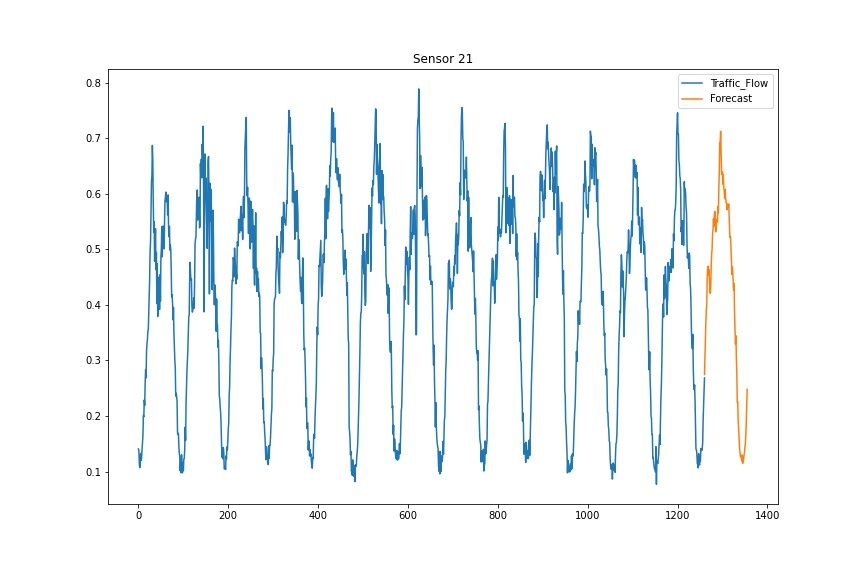
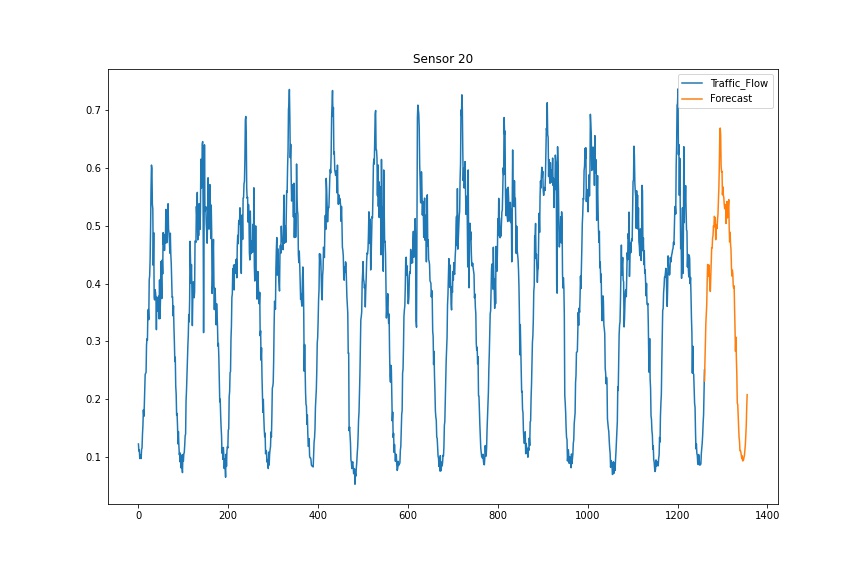
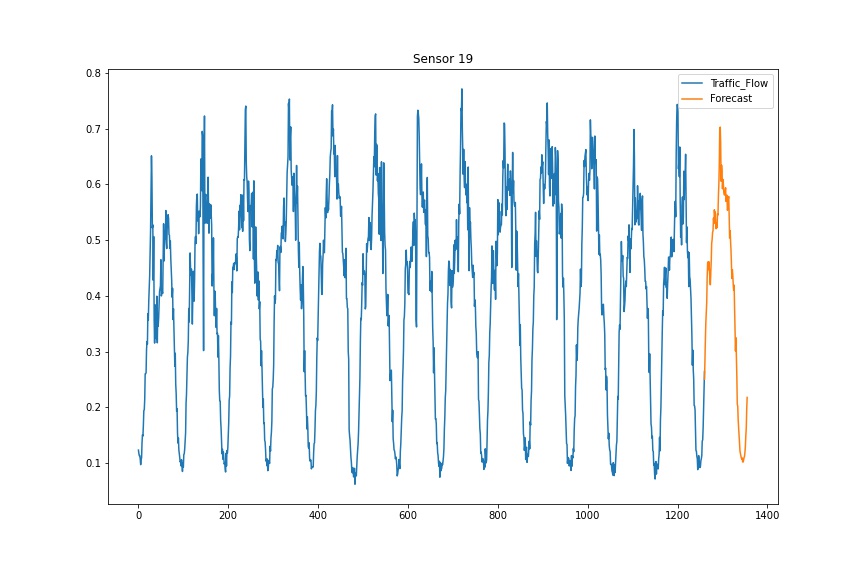
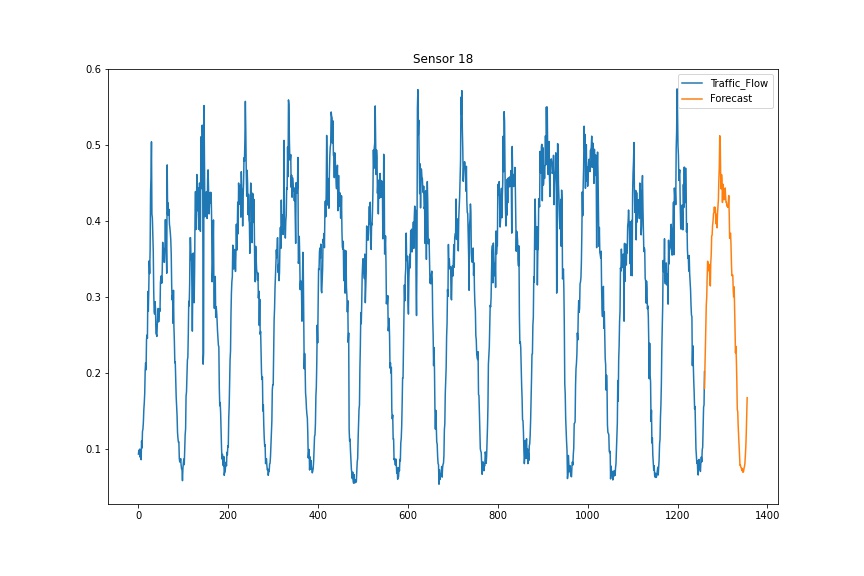
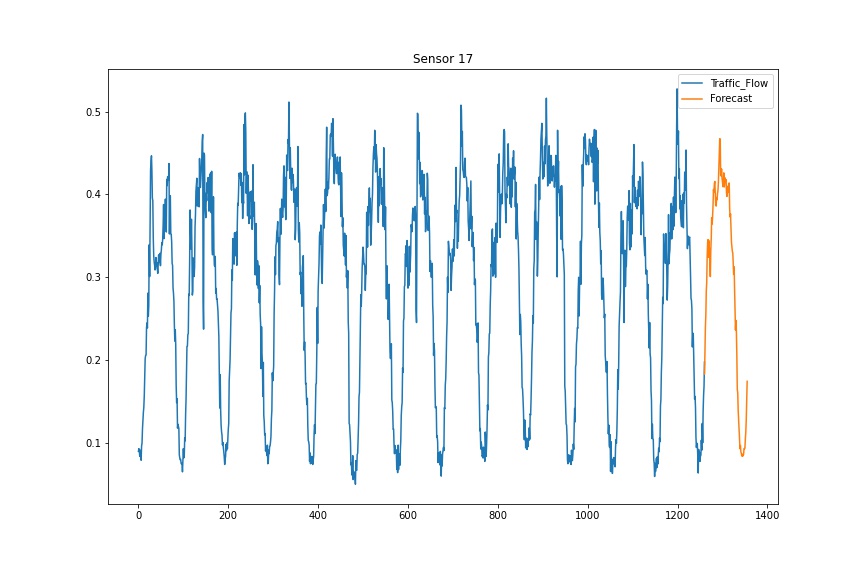
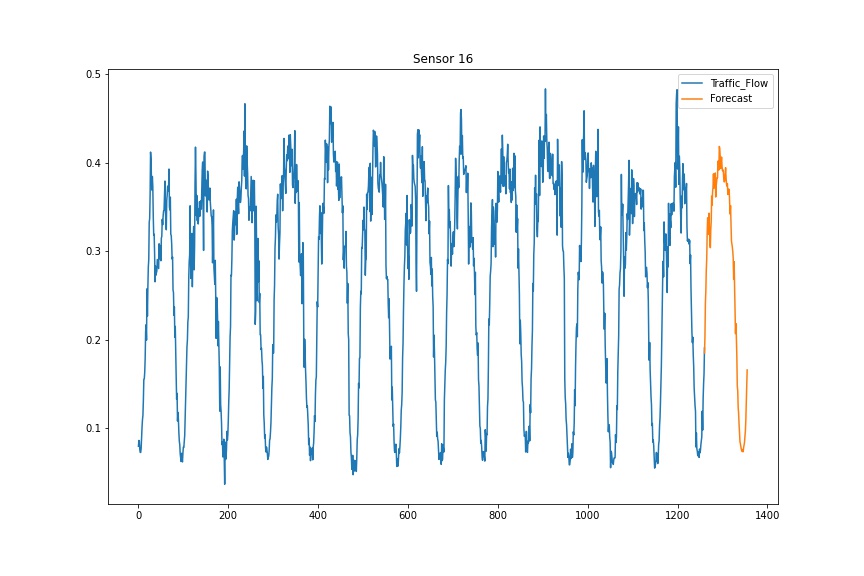
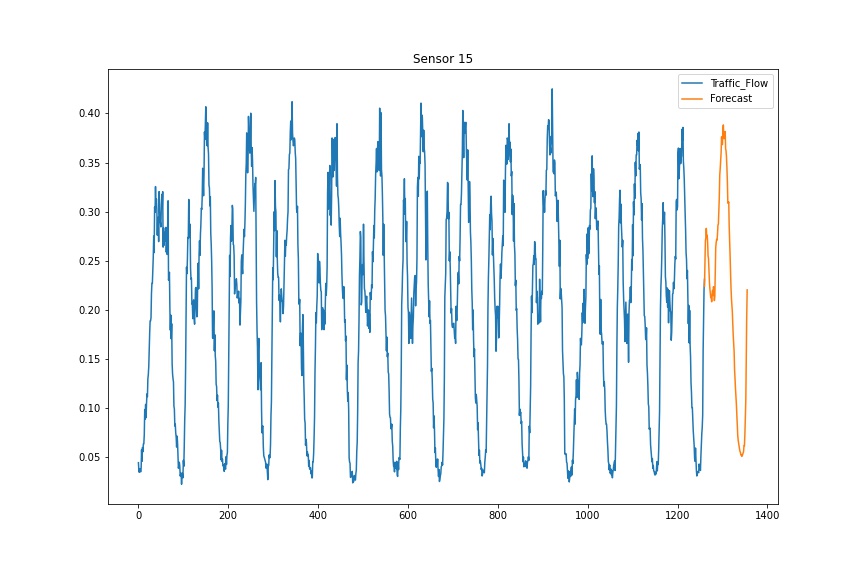
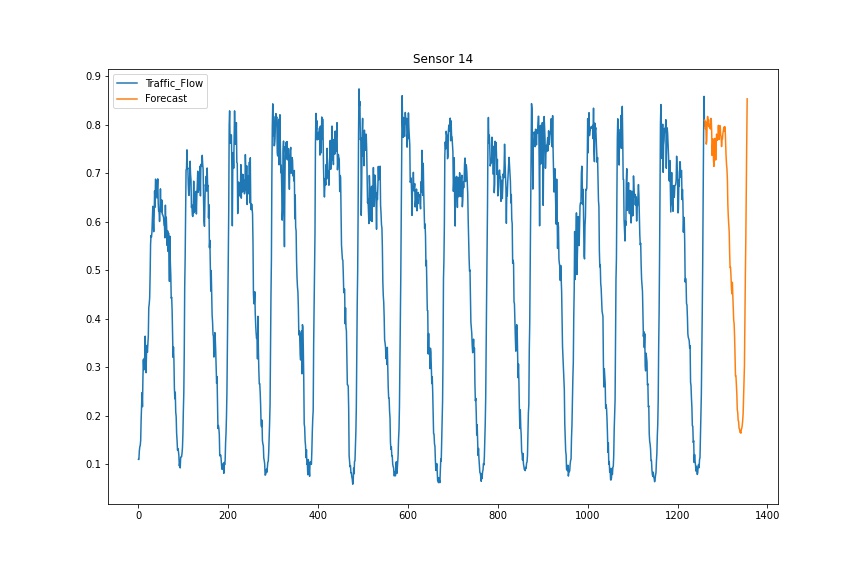
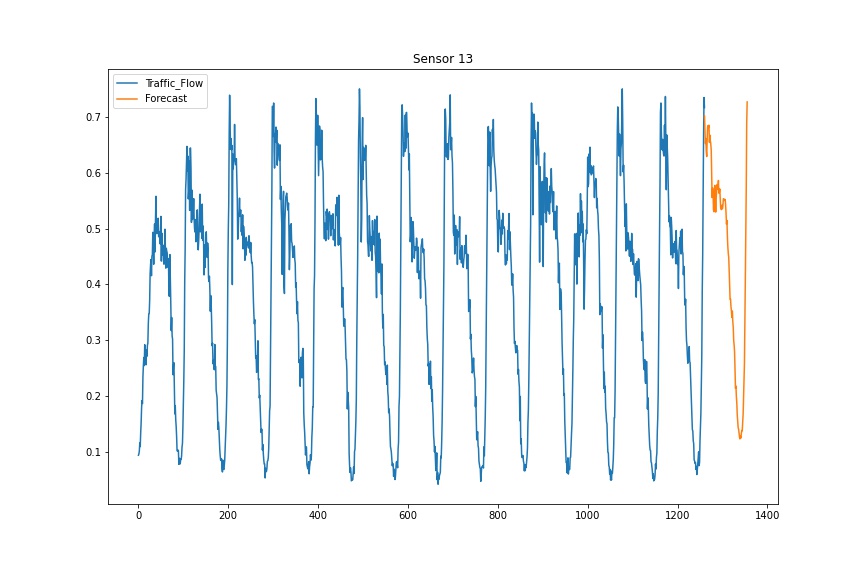
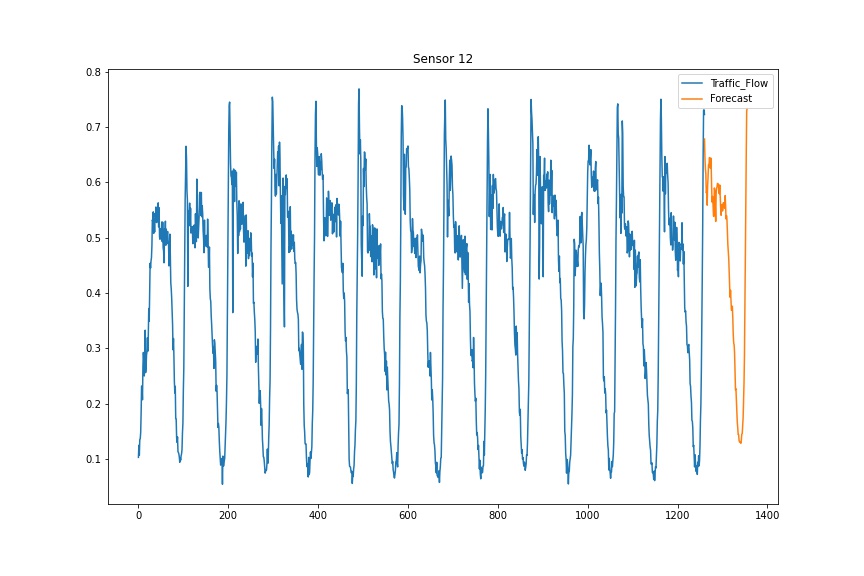
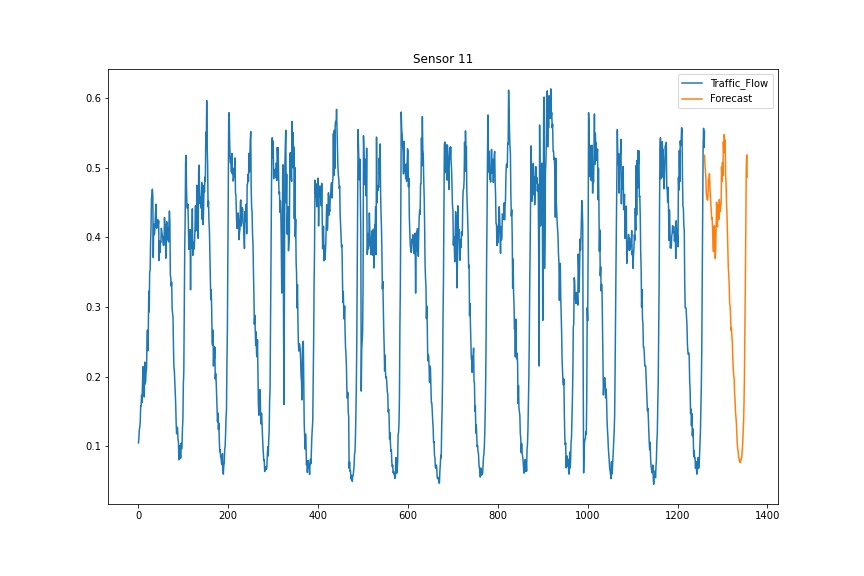
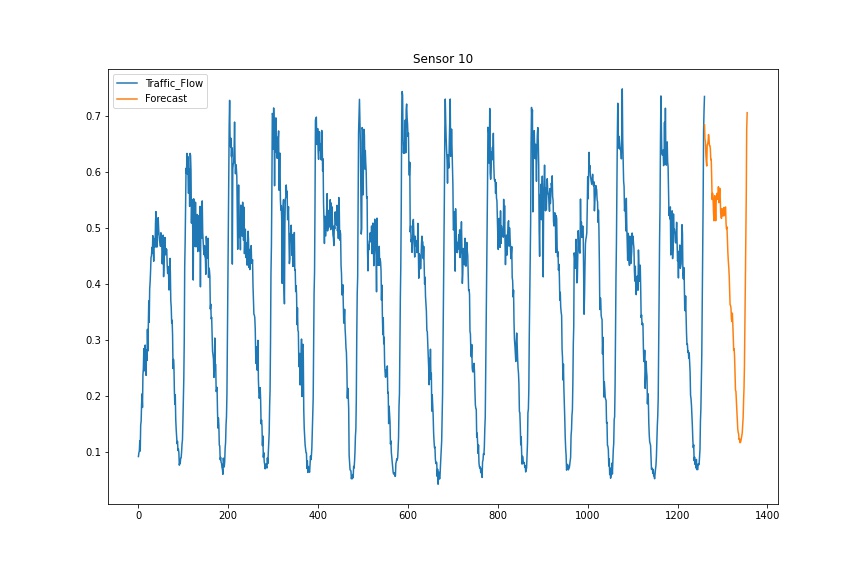
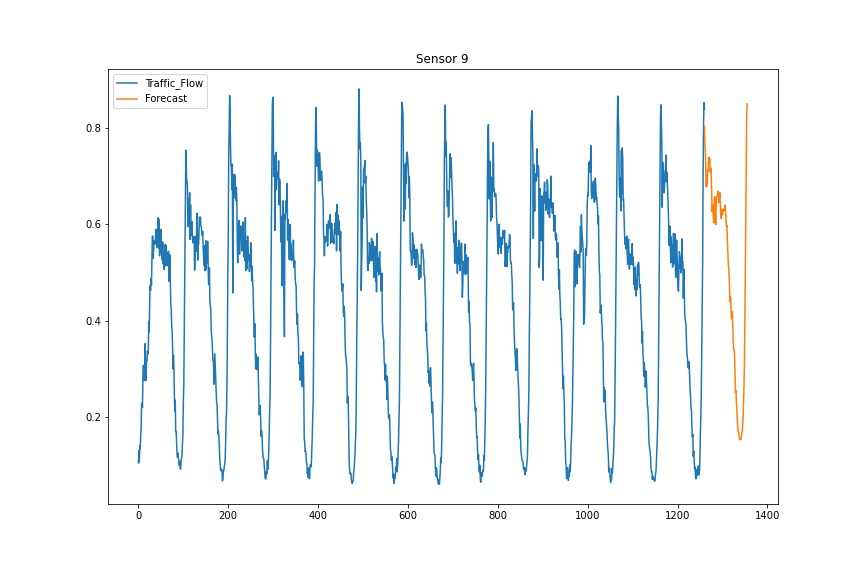
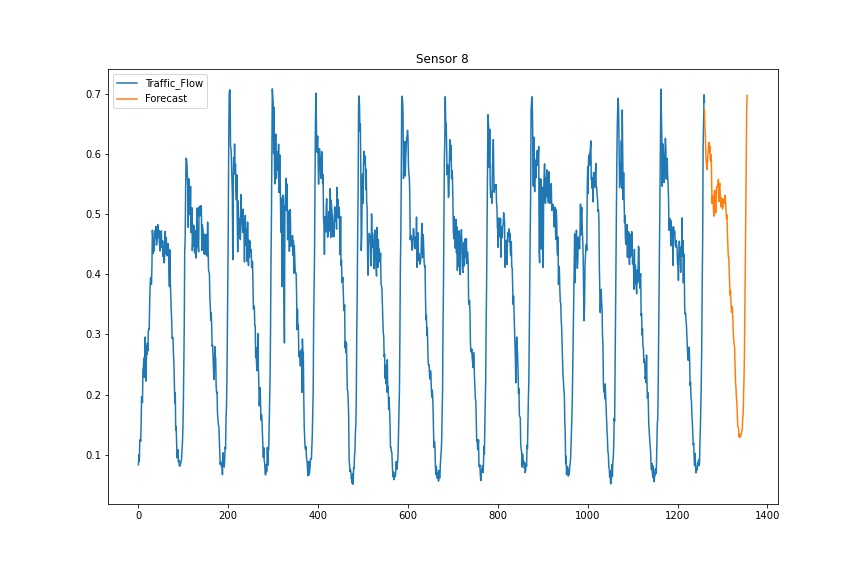






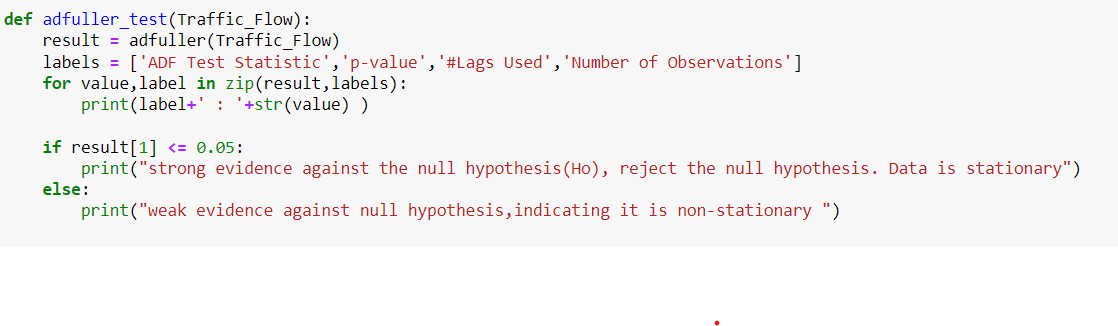




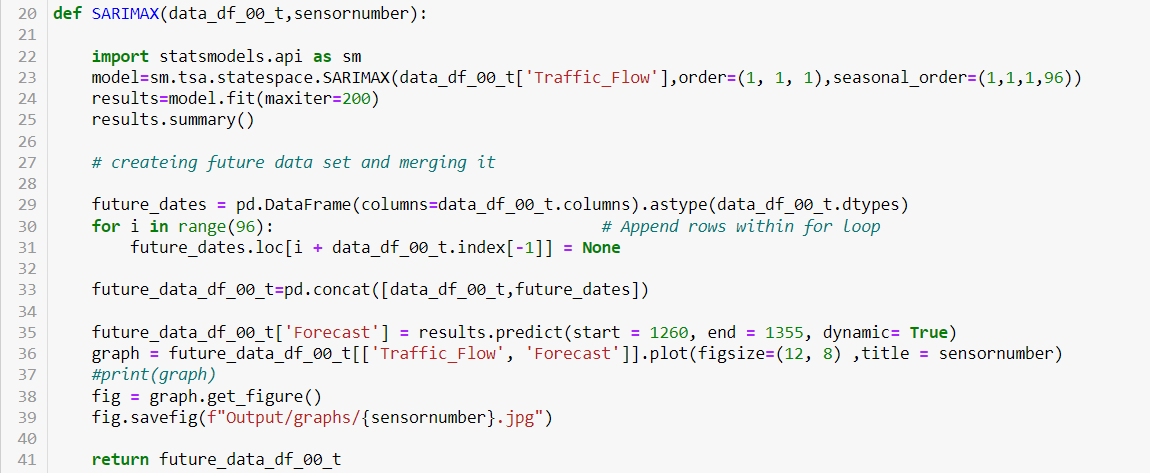


**Code Explanation:** We have written our program in Python using Scipy, matplotlib, pandas and statsmodels with proper comments indicating each step

We have added one function to check if data is stationary or not using adfuller test



Prediction is done using function SARIMAX where we will pass each sensor’s data and sensor number. Further this function will forecast data and plot it and save data in output folded with file name as “sensor <number>.jpg”



Processing starts with loading matlab file and then creating data frame



Output is generated by invoking SARIMAX function

