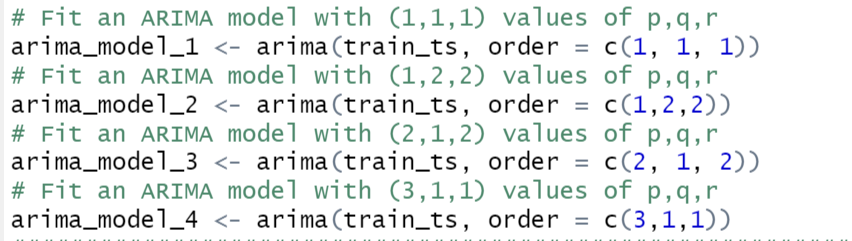
The pollution dataset named ‘**pollutionData209960.csv**’was collected from this [website](http://iot.ee.surrey.ac.uk:8080/datasets/pollution/).We have splitted the data into training and test datasets by manual percentage split of 97% percentage for training and 3% percentage split for the testing data.The reason I selected 3 percent gives us about 527 rows.As per description of task I tended to use the carbon\_monoxide attribute to use in my arima model .I have set the frequency to 288 in code ‘train\_ts <- ts(train\_data, frequency = 288)’ because there are 288 5-minute intervals for every 24 hours.I have considered the 4 types of ARIMA models for training and prediction purposes with (p,d,q) values (1,1,1); (1,2,2);(2,1,2); (3,1,1).All these values are considered based on the ranges of these (p,d,q) values can afford. The arima function is part of the stats library in R, which is a built-in library that comes with the base R installation.So all the different models are trained on train data and then we used it to test it on test data.The forecast function is used to predict the values for next **n** intervals.This gives us mean, upper 95% confidence interval bound, lower 95% confidence interval bound, upper 80% confidence interval bound, lower 80% confidence interval bound of forecasted values.So Mean Absolute Percentage Error Values are considered for all the 4 different models are calculated and plotted in the graph below which can be used to find the best performing model.

# The MAPE values calculated on different bounds of different models

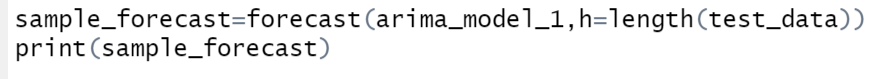
So upper 95, and upper 80 bounds performed well.

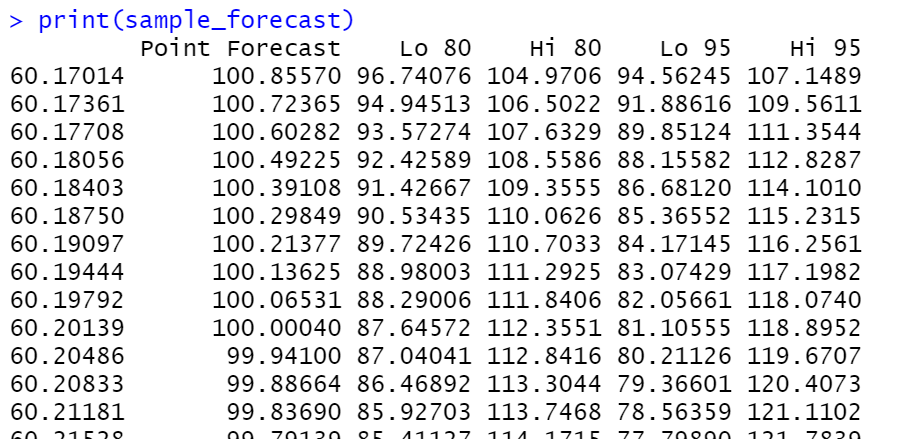
But in upper 80% Confidence Interval bound the model 3 performed well with (p,d,q) values-->(2, 1, 2).So model 3 is most suitable for the prediction purposes

**Code for model training:**

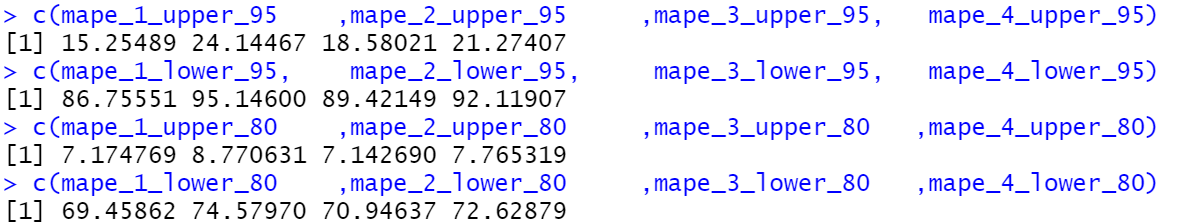


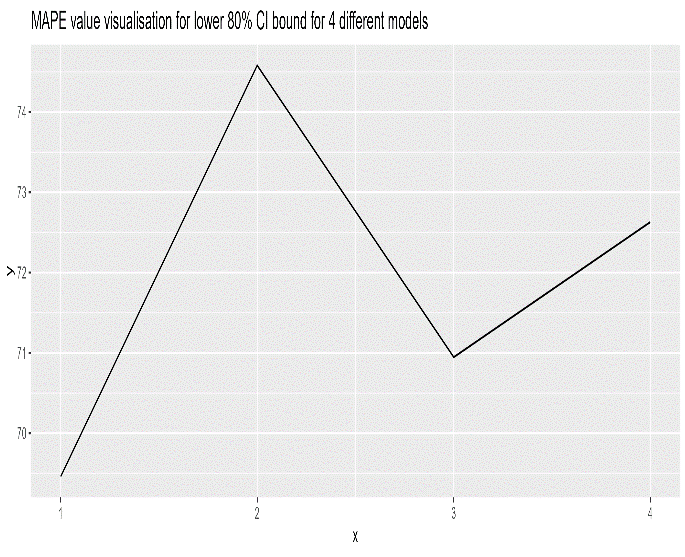
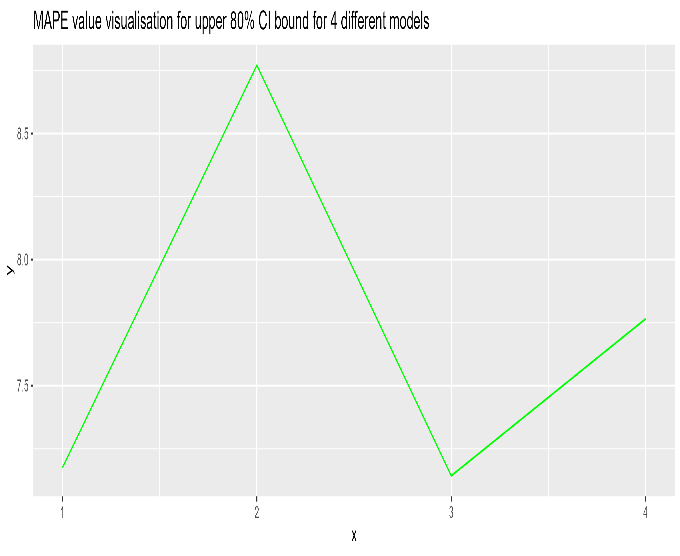
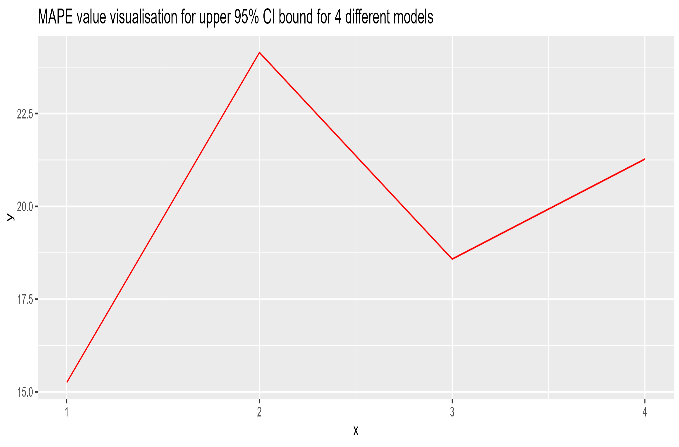
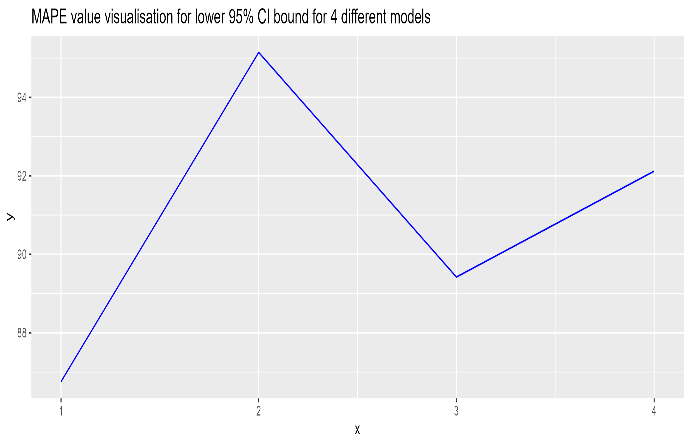
**Model prediction for test\_data for all bounds:**





**Model comparison with different bounds:**

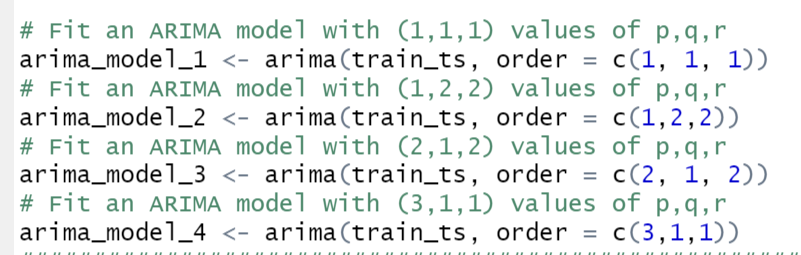




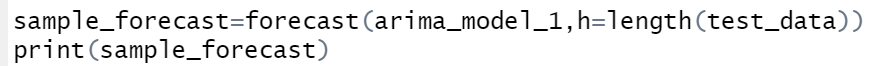
For GROUP1pollutionData209907.csv dataset modelling:

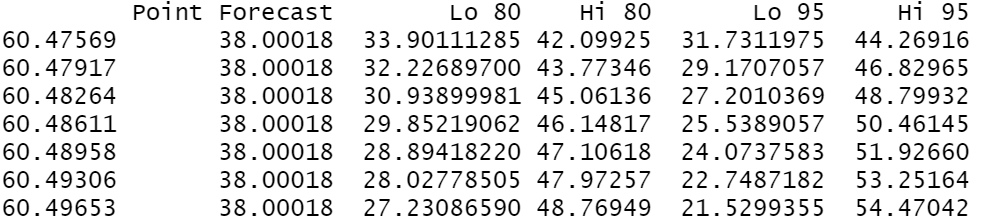
The data was split into training and testing datasets through a manual percentage split, with 98% allocated for training and 2% for testing. The selection of 2% was based on obtaining approximately 527 rows. The attribute "carbon\_monoxide" was used in the ARIMA model, and the frequency was set to 288 using the code "train\_ts <- ts(train\_data, frequency = 288)" to account for the 288 5-minute intervals in every 24 hours. Four types of ARIMA models were considered for training and prediction purposes, with (p,d,q) values of (1,1,1), (1,2,2), (2,1,2), and (3,1,1), selected based on their respective allowable ranges. These models were trained on the training data and tested on the testing data using the forecast function to predict values for the next n intervals. The resulting forecasted values included mean, upper and lower 95% confidence interval bounds, as well as upper and lower 80% confidence interval bounds. Mean Absolute Percentage Error Values were calculate for all four models, and a graph was plotted to identify the best-performing model.*so lower 80 bounds performed well,in lower 80 bound the model 3 performed well with (p,d,q) values-->(2, 1, 2)*

**Code for model training:**



**Model prediction for test\_data for all bounds:**





**Model comparison with different bounds:**

