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

# mape\_1\_upper\_95 mape\_2\_upper\_95 mape\_3\_upper\_95 mape\_4\_upper\_95

# 15.25489 24.14467 18.58021 21.27407

# mape\_1\_lower\_95 mape\_2\_lower\_95 mape\_3\_lower\_95 mape\_4\_lower\_95

# 86.75551 95.14600 89.42149 92.11907

# mape\_1\_upper\_80 mape\_2\_upper\_80 mape\_3\_upper\_80 mape\_4\_upper\_80

# 7.174769 8.770631 7.142690 7.765319

# mape\_1\_lower\_80 mape\_2\_lower\_80 mape\_3\_lower\_80 mape\_4\_lower\_80

# 69.45862 74.57970 70.94637 72.62879

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

