FORECASTING

TABLE OF CONTENTS

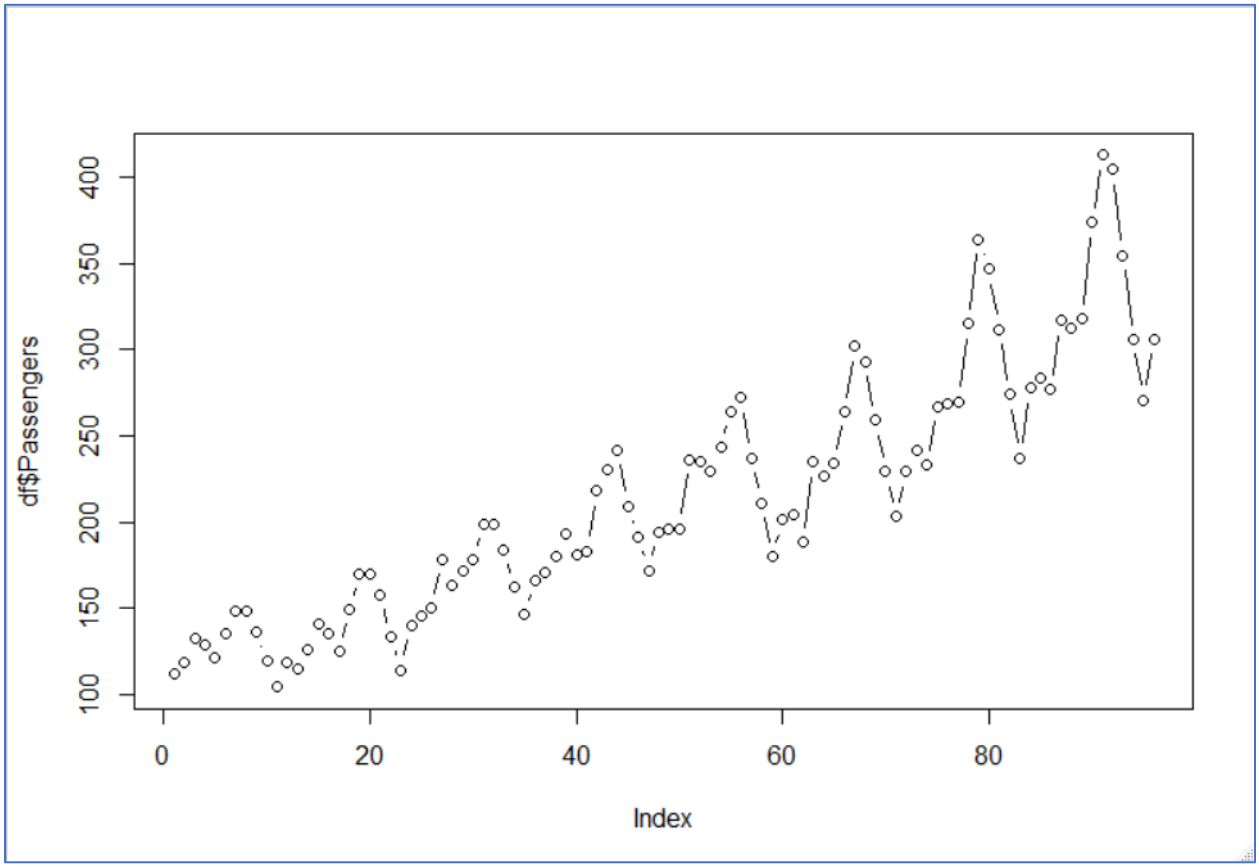
FORECASTING

QUESTION NO 1 FOR PREDICTING PESSENGER IN AIRLINE DATA

In Airline data we have 2 columns

Month (Independent) and Passengers (Target)

Here data is for year 1995 to 2002, each data is collected in 1st day of each month.



Here our passenger list seems to be in an increasing trend as well as fluctuate with seasonality changes. So, we can see cyclic variation in our data as well as linear increasing trend of passenger numbers.

DATA PREPROCESSING

Our Variable Month is in format (yyyy-mm-dd).

First of all I have replicated every year 12 time and assigned it as Year

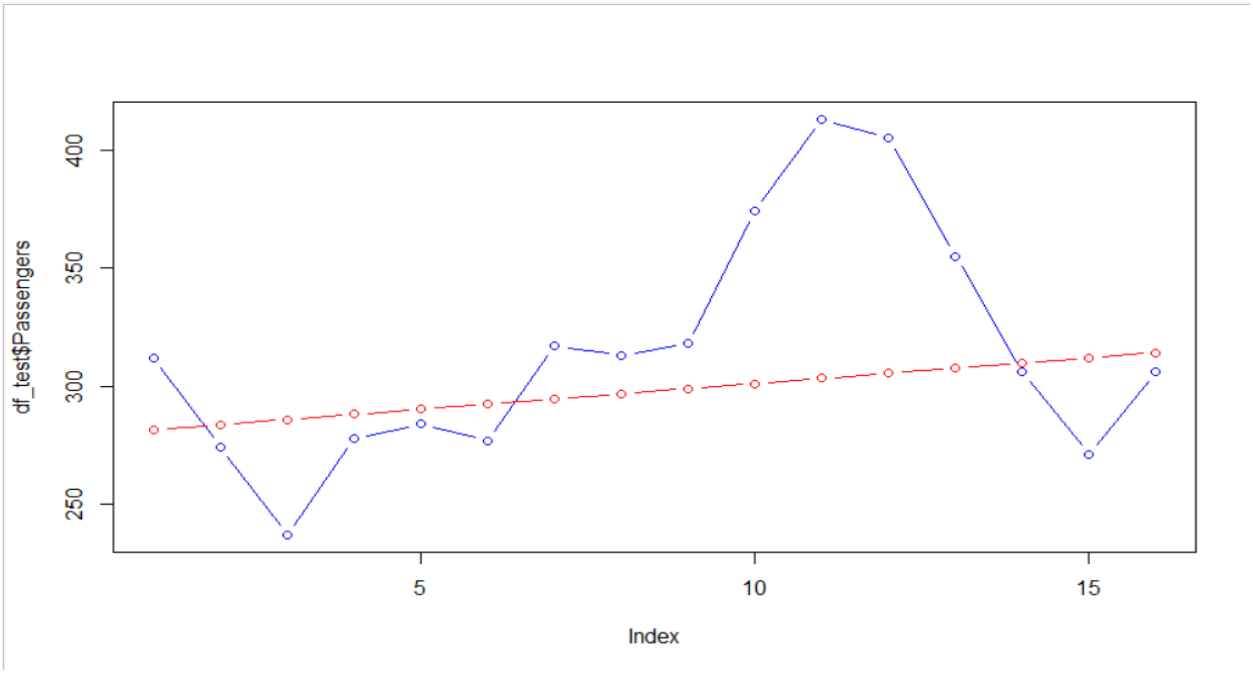
Then I created 12 dummy variables for 12 months

Finally, I created a variable day, as on an average we can say that the data are collected in the interval of 30 days. So, it will represent the rough intervals of data collection days taking initial day as 1.

I have 96 records, out of which I considered first 80 records for train data and rest 16 records for test data

LINEAR TREND WITH VARIABLE DAY

Here I considered my day variable and check whether the linear trend is enough to explain my model or not.



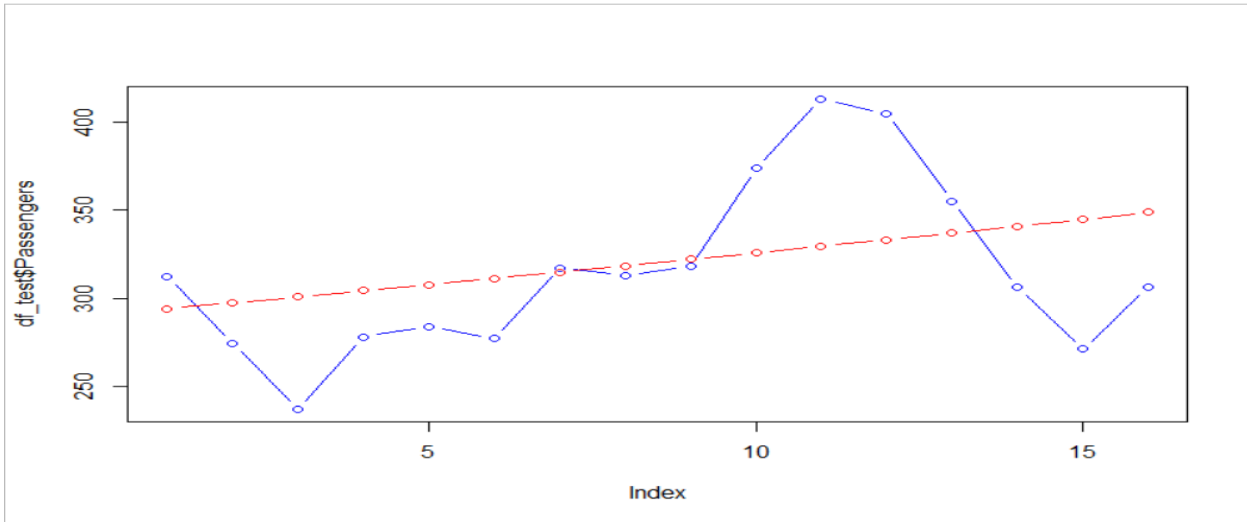
In the above plot red line is my predicted value and blue line is my actual value

Here my linear model is unable to explain the seasonal variation in my data. Which is a lack in my model.

Here my R^2 value is 0.7829 and RMSE value is 47.5462

EXPONENTIAL MODEL WITH VARIABLE DAY

Here in my model I used log in my target variable, which is exactly same as taking exponential in my independent variables.

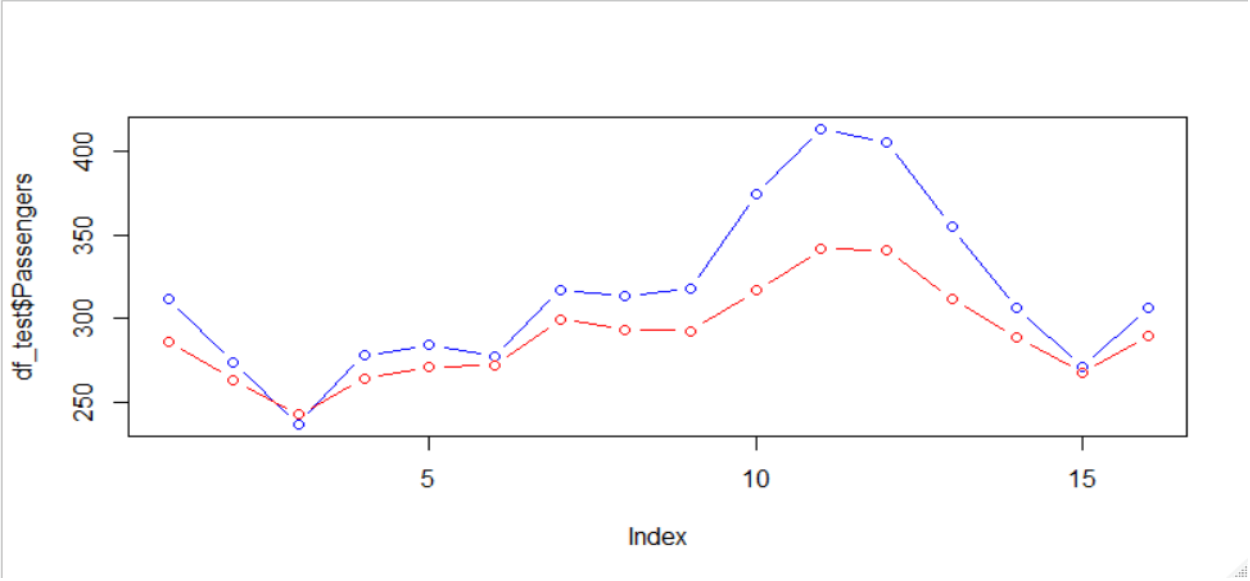


Here in my second model I come up with R^2 as 0.8181 and RMSE as 43.7937

ADDITIVE SEASONALITY AND LINEAR TREND WITH ALL DATA

I have considered all my variables in this model to remove seasonality cyclic problem in my prediction I have introduced month column as a dummy column in my model.

Here My prediction is quite good and the pattern is matching with the original data, still its unable to explain the

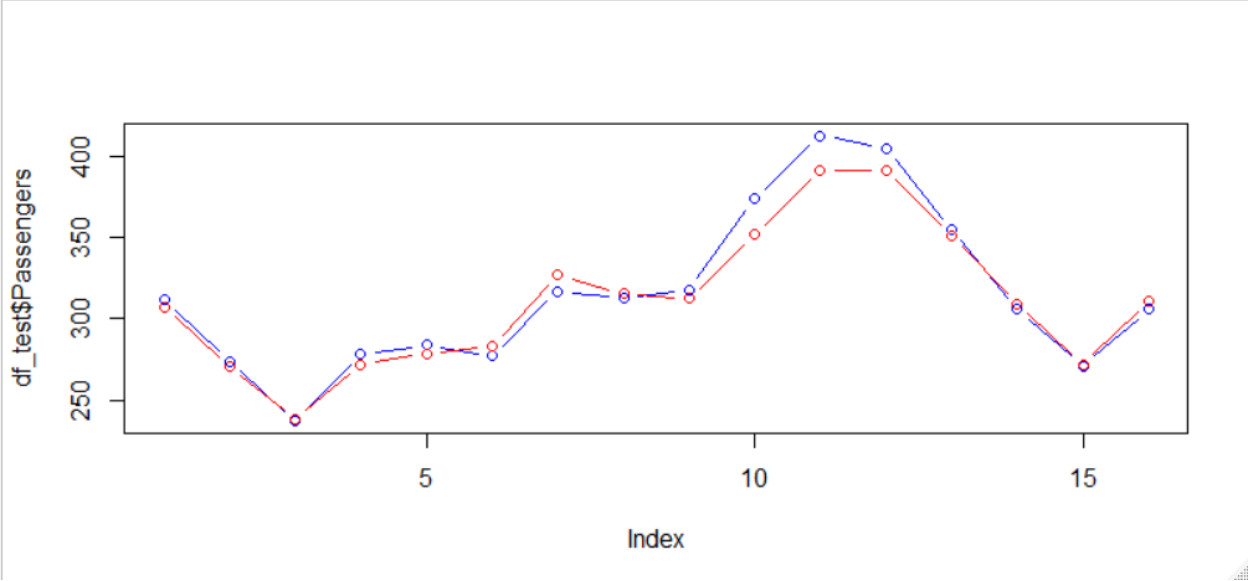


extreme values properly.

Here I come up with R^2 as 0.95 and RMSE is 33.04

MULTIPLICATIVE SEASONALITY LINEAR TREND WITH ALL THE VARIABLES

Applied exponential transformation in my independent variables by taking log in my target variable.



In this plot We can see my predicted values are nearly equal to my Actual values, which seems like we get our desired efficiency. Here my R^2 is 0.97 and RMSE is 9.469 i.e. least among all other models.

TABULATION

Model	R^2	RMSE
Linear Trend	0.7829	47.5462
Exponential Model	0.8181	43.7937
Additive Seasonality and Linear Trend	0.95	33.04
Multiplicative Seasonality Linear trend	0.97	9.469

CONCLUSION:

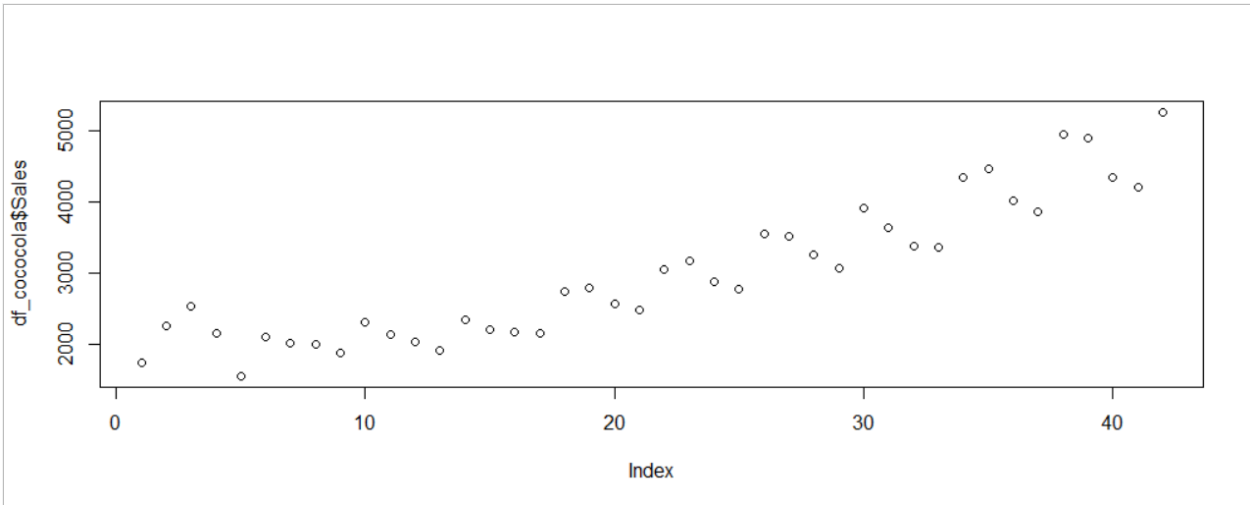
From the above Tabulation we can see that in my 4th model I am getting highest R^2 value as well as lowest RMSE Values, and also looking on the actual and prediction graph, I will consider my 4th model as the final model.

QUESTION NO 2 FOR PREDICTING THE SALES OF COCOCOLA

In my data set I have 2 variables and 42 records.

My records are collected in quarterly basis, i.e. in the gap of 3 months each year. Been collected

Data is collected in between year 1986 to 1996. The data is free from missing values as well as outliers.



Here we can see sales gradually increases from the record 9 i.e. from the third cycle which is representing a linear trend as well as the variation within the year is representing the cyclic variation in our data

DATA PRE PROCESSING

As done above in previous question again I am creating a variable Year

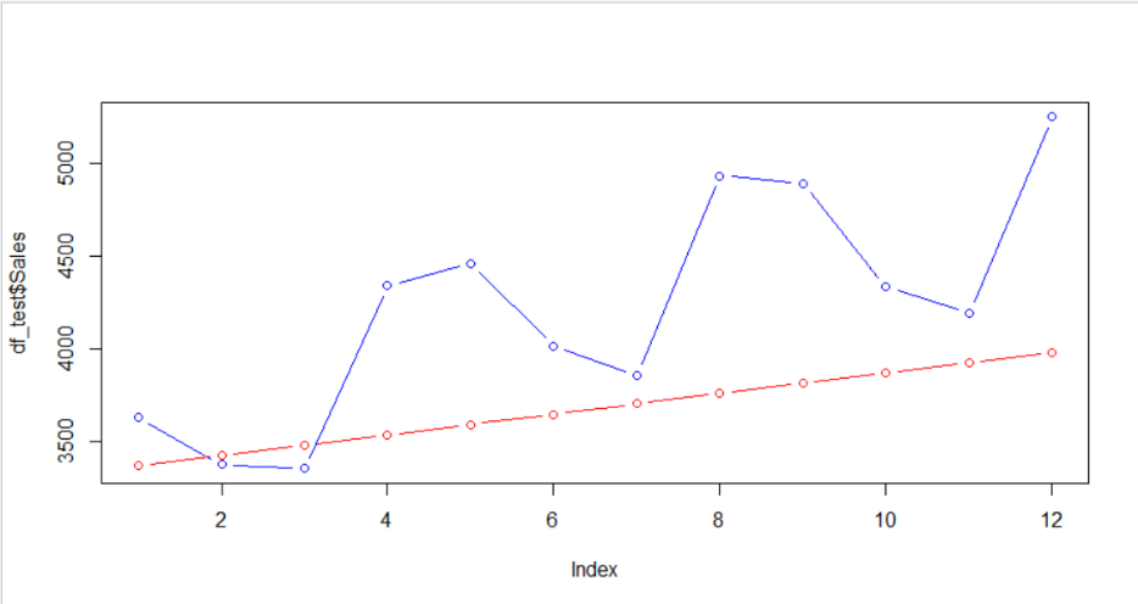
Here as the data is collected in quarterly basis so I have created 4 dummy variables for 4 quarters.

Created a column called “rn” i.e. representing the record number which will help us to follow the trend in my data predictions.

I have considered my first 30 records as my train data and the rest 12 data as my test data

LINEAR TREND WITH VARIABLE RN (RECORD NUMBER)

Here I have considered my record number as my independent variable to check whether the Sales is increasing trend or not.



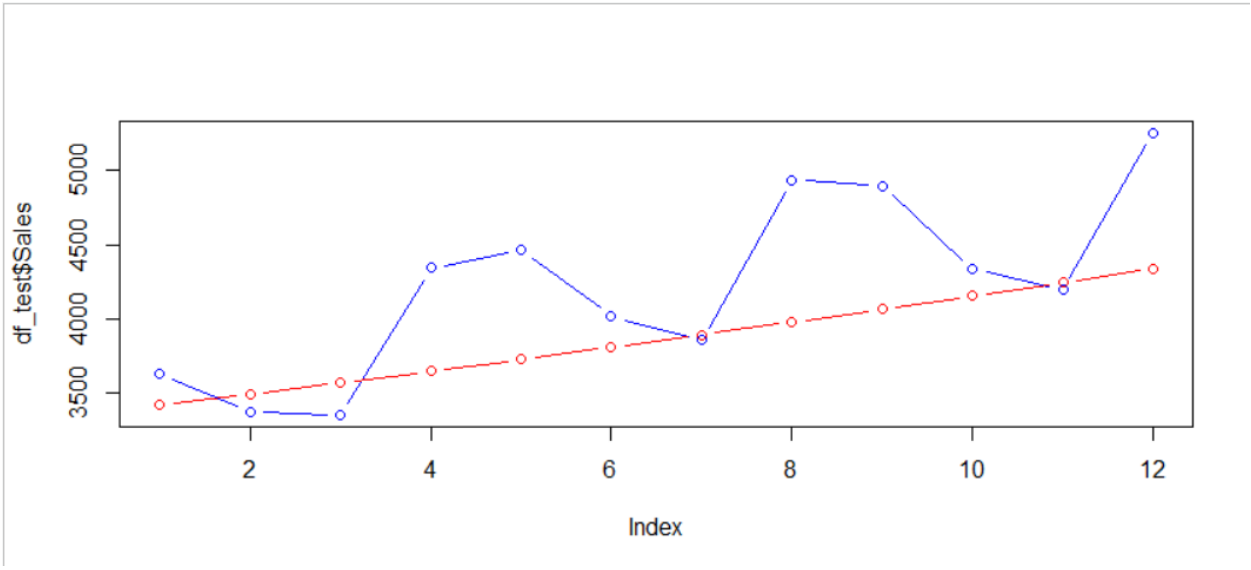
Here my red dot are representing the predicted values and blue dots are representing the Actual Values.

Here its unable to explain the cyclic variation in our data and also due to considering the data points 1 to 8 which is erratic in nature we are facing some issues, which I am going to remove in my future models.

Here in this model I am getting R^2 as 0.7079. and RMSE as 714

EXPONENTIAL MODEL CONDIDERING ONLY RECORD NUMBERS

As described in my first question's answer regarding the exponential model, I have considered log(Sales) as my Target variable and Here I am getting the Actual and Predicted points in my plot.

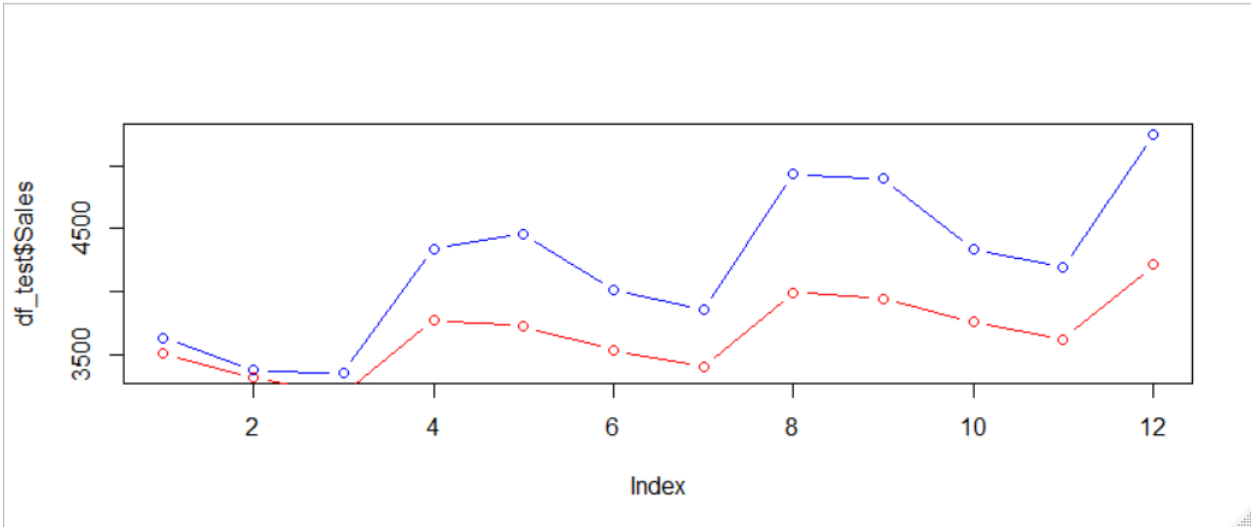


Here in this plot we can see slight increment in my trend line from the last model. So we can say exponential model is quite effective than my linear trend model.

Here I am getting R^2 = 0.7067 although decreased from the above model as well I am getting less RMSE than my previous model i.e. 552.2821 which is good.

ADDITIVE SEASONALITY WITH LINEAR TREND

In my 3rd model I have considered my Dummy variables which is regarding my quartiles (Q1,Q2,Q3,Q4). Also I have considered my "rn" variable for keeping the linear tendency of my prediction.

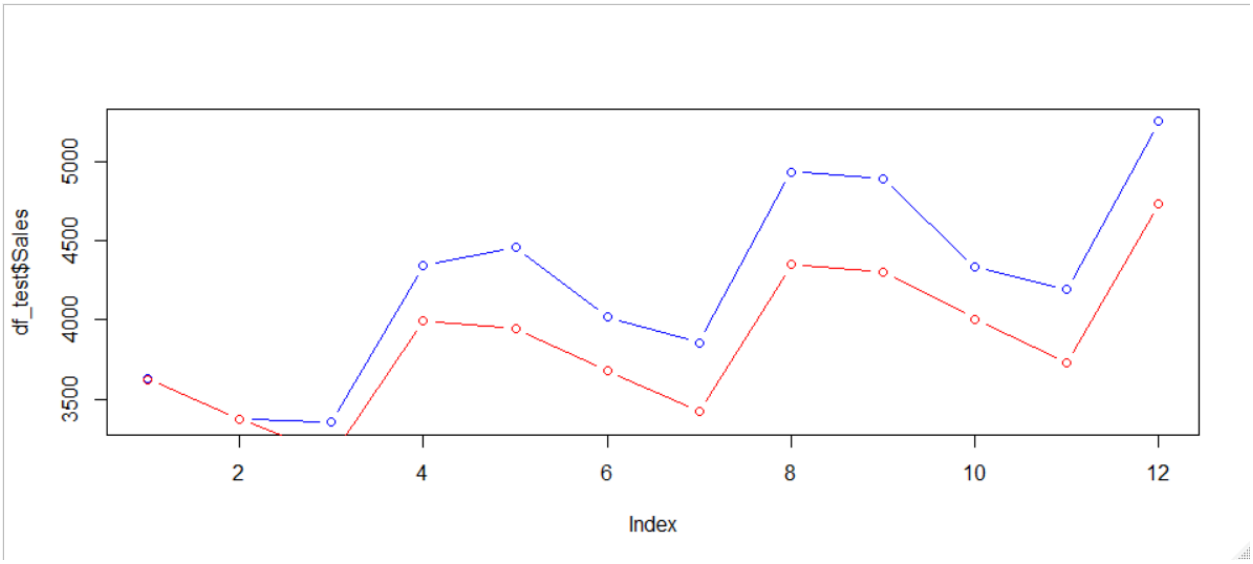


From this plot we can see similar pattern in my Predicted variable as we can see in our Actual variable, Which is quite helpful for us, but still the predicted value is widely separated from the Actual. Which can be treated taking logarithm in my target variable.

Here I am getting R^2 = 0.84 (**increased**) good for us and RMSE = 637 (**increased**) not good.

MULTIPLICATIVE SEASONALITY LINEAR TREND

Here I am doing my exponential transformation in my independent variable as explained in my previous models.

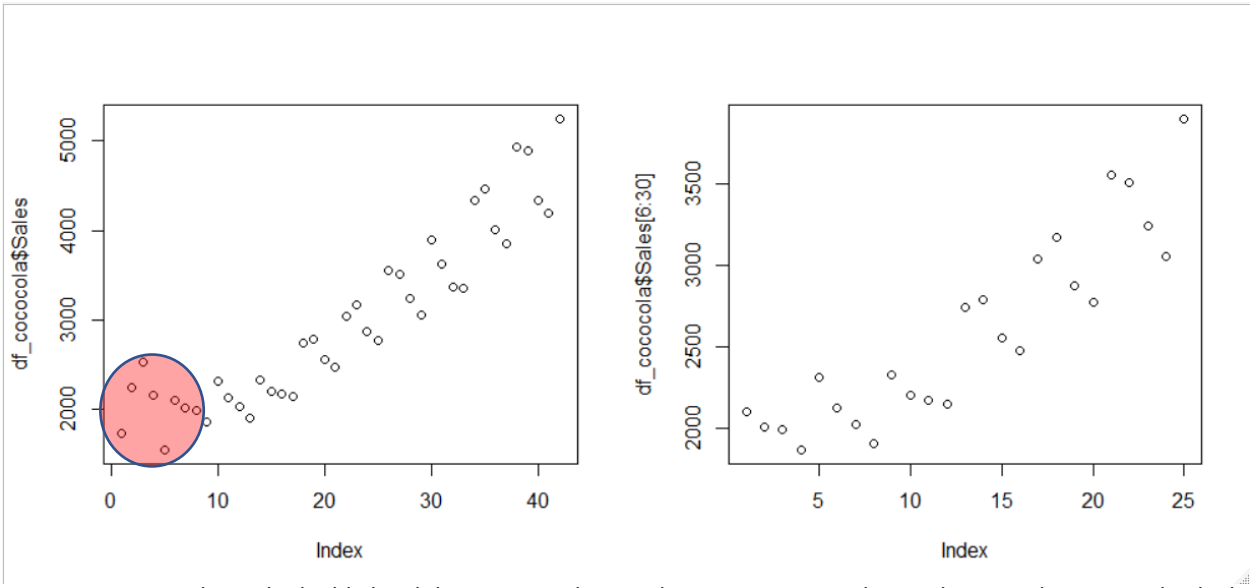


In this plot we can see the separation (gap) between our Actual and predicted model is decreased to some extent.

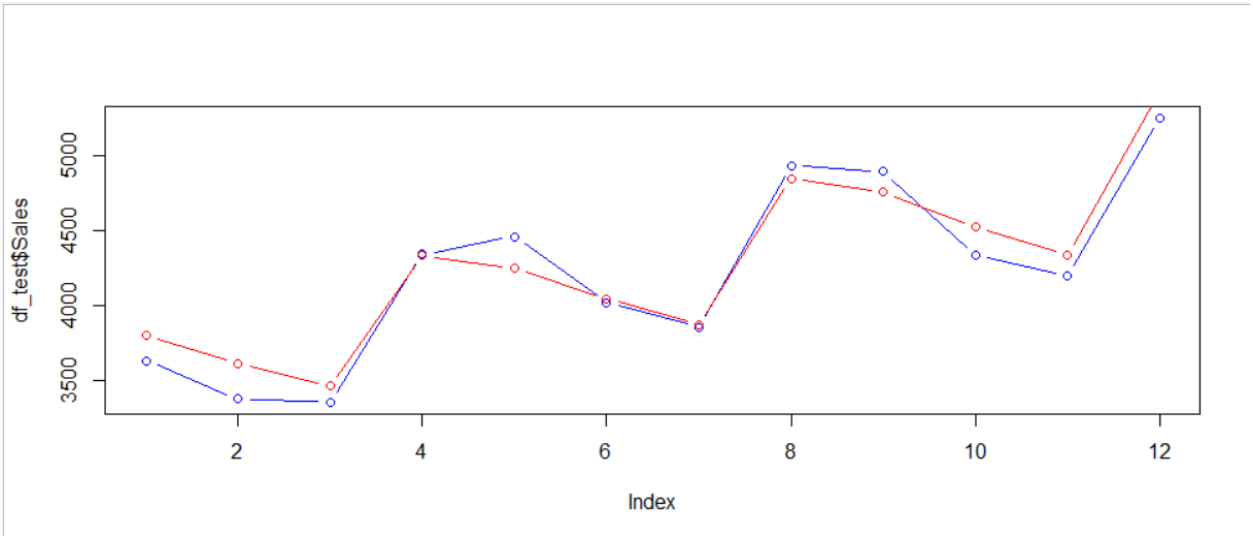
Here I am getting R^2= 0.8586 and RMSE =410 Here our model accuracy is quite increased from our previous model.

But to cover the remaining accuracy, I wish to cut the erratic variation in my data i.e. record 1 to 8 which have no definite pattern.

MULTIPLICATIVE SEASONALITY LINEAR TREND REMOVING THE ERRATIC COMPONENTS



Here I am saying about the highlighted data points, These 8 data points are neither cyclic nor in linear trend, which seems like random, and erratic in nature. So I prefer to remove there 8 records in my 5th model.



Here in this plot I achieved my desired accuracy , the difference between the Actual and predicted data points is negligible in this model prediction.

Here I am getting R^2 = 0.9655 and least RMSE among all i.e. 146.4575

TABULATION

MODEL	R^2	RMSE
LINEAR TREND	0.7079	714.0144
EXPONENTIAL MODEL	0.7067	552.2821
ADDITIVE SEASONALITY AND LINEAR TREND	0.8457	637.9405
MULTIPLICATIVE SEASONALITY LINEAR TREND	0.8586	410.2497
MULTIPLICATIVE SEASONALITY LINEAR TREND*	0.9655	146.4575

CONCLUSION

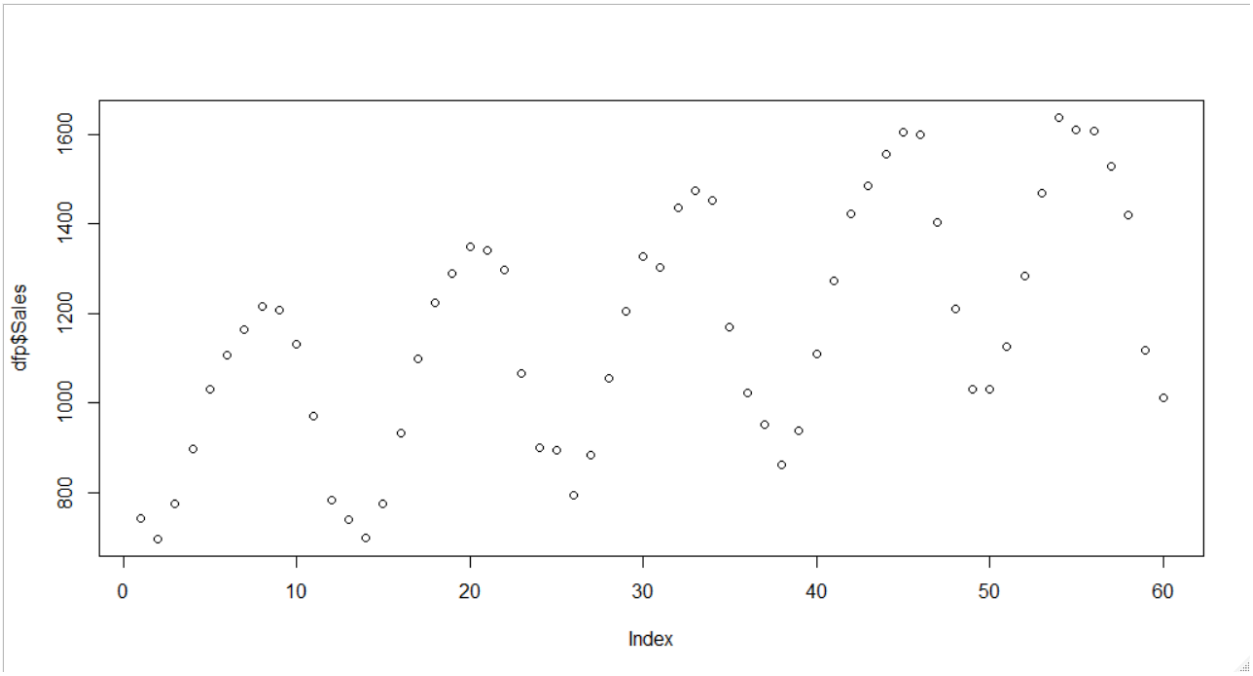
From the above table we can see that our last model (5^{th}) model is with the highest R^2 value as well as least RMSE value among all the models, So I may consider my 5^{th} model as my final model.

Also Same inference can be drawn from the plots too.

QUESTION NO 3 FOR PREDICTING THE SALES OF PLASTIC

In my data set I have 2 variables i.e. Sales (Target variable) and Month (independent variable). Our data contains 60 records, which is recorded monthly basis.

Also our data is free from missing values and outliers.



From this graph we can see that Sales is purely cyclic in nature, with week sign of trendline in my Sales data.

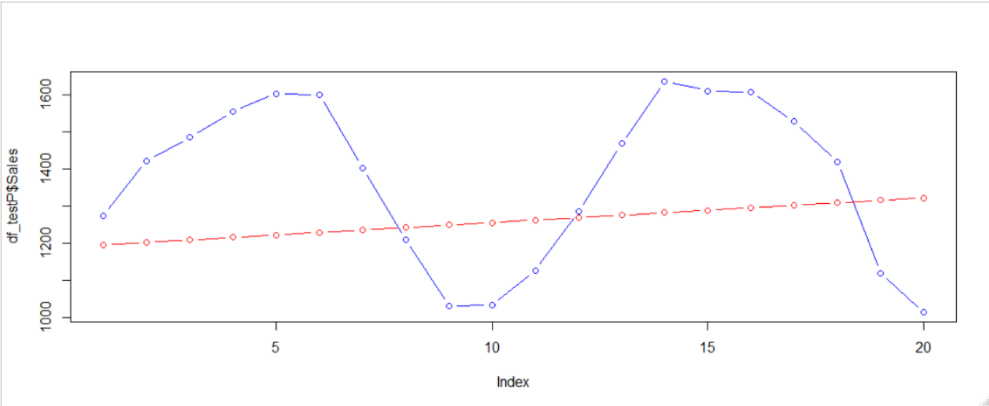
DATA PRE PROCESSING

As done in my 1st Experiment (Preprocessing for my 1st Experiment), here also I have created my year variable from 1949 to 1953, and created dummy variable for all the months.

For training purpose I have considered 40 initial records, and for test purpose I have my last 20 records.

LINEAR TREND MODEL

In my linear model I have considered my record number for analyzing the presence of linear trend in my data.



In the above plot we can see that, the predicted value is not capable for representing the cyclic (Seasonal Variation) in our data. So we may face serious issue considering my linear trend here.

By the way, I am getting my R^2 = 0.1211 i.e. very less and RMSE as 248

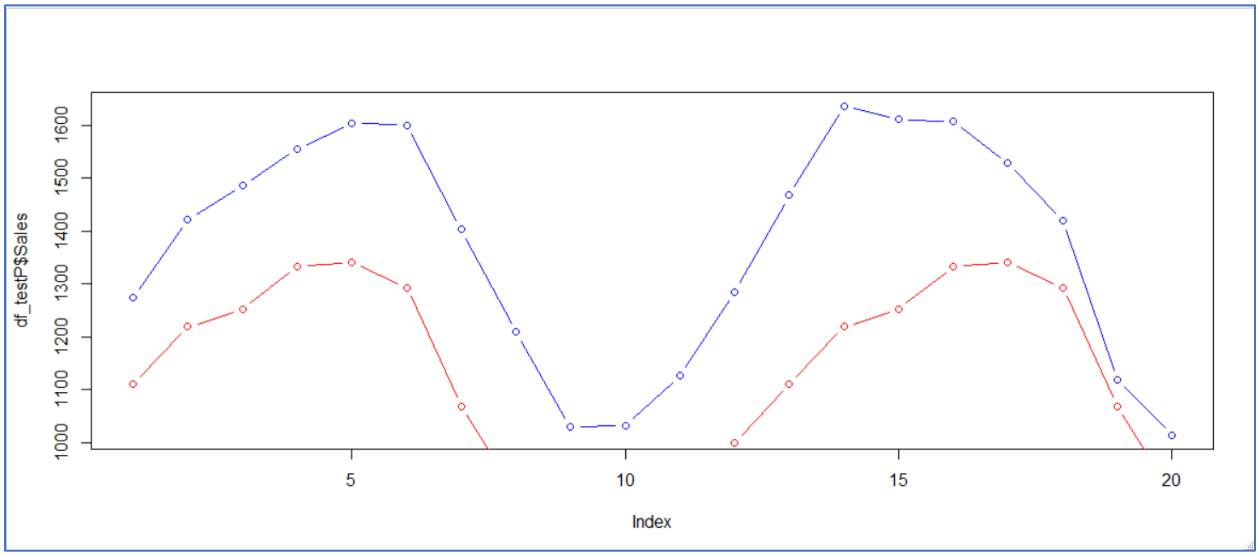
EXPONENTIAL MODEL

In my exponential model I am considering my target variable with log transformation as explained in my previous exponential transformation methods.

Here no changes from the above model is seen, We are getting the same results for the variable record number(rn). So its not recommended to go for this model in particular this data set.

MODEL TO CAPTURE SEASONAL VARIATION:

In this model I have considered all my dummy variables only.



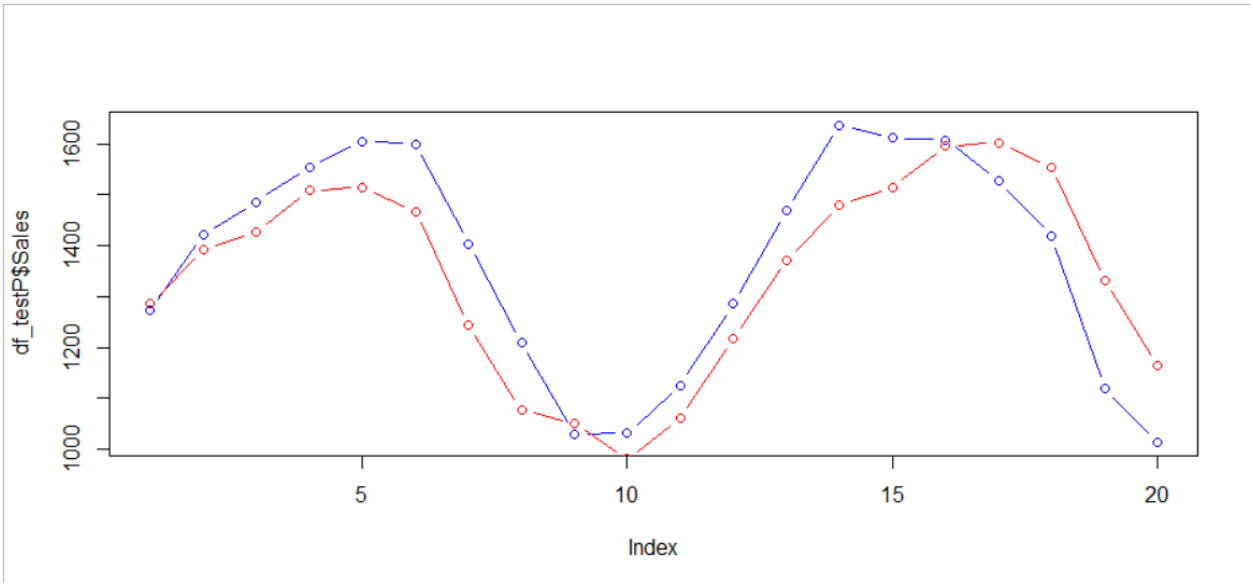
Here from the above plot we can see that, the pattern of Actual value is being followed by our predicted values.

Here I come up with R^2 = 0.8405 and RMSE = 263.2362

From this we assured that the Sales is totally dependent upon the Seasonal changes.

ADDITIVE SEASONALITY WITH LINEAR TREND

In this model I have considered the Year column with all the Dummy variables i.e. (Months)

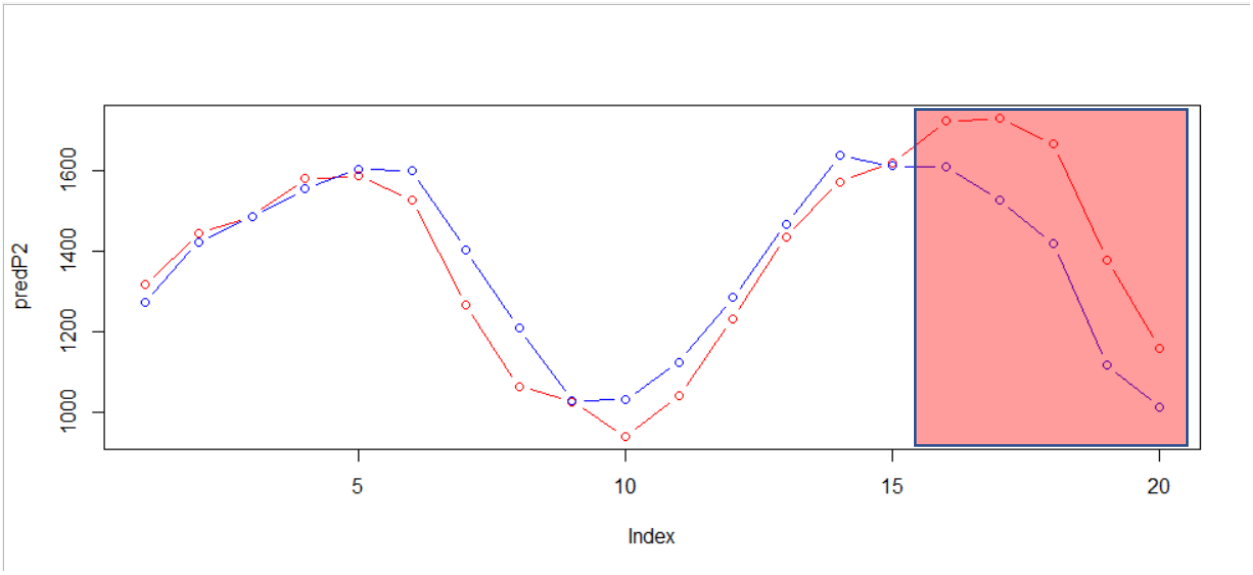


From this plot we can see that the gap i.e difference between the Actual and Predicted variable is compressed to some extent as compare our previous model.

Here I come up with R^2 = 0.9791 and RMSE = 105.2468, improvement from our first model.

MULTIPLICATIVE SEASONALITY

This model is modification from the previous model. Here the only thing is exponential transformation to our independent variables and removal of record number (rn) from the model.



This plot is fitting well with our Actual data but slight problem regarding the last 5 records, here the difference is quite noticeable.

Here I am getting R^2 = 0.9848 (same as previous) RMSE = 117.115 (increased from previous model)

TABULATION

MODEL	R^2	RMSE
LINEAR TREND	0.1211	248
EXPONENTIAL MODEL	0.1211	248
MODEL TO CAPTURE SEASONAL VARIATION	0.8405	263.2362
ADDITIVE SEASONALITY WITH LINEAR TREND	0.9791	105.2468
MULTIPLICATIVE SEASONALITY	0.9848	117.115

CONCLUSION:

From the above tabulation and Looking at all the Actual and Predicted Plots, I can say that my 4th model "Additive Seasonality with Linear trend" model is perfect with less RMSE value as well as a good R^2 value.