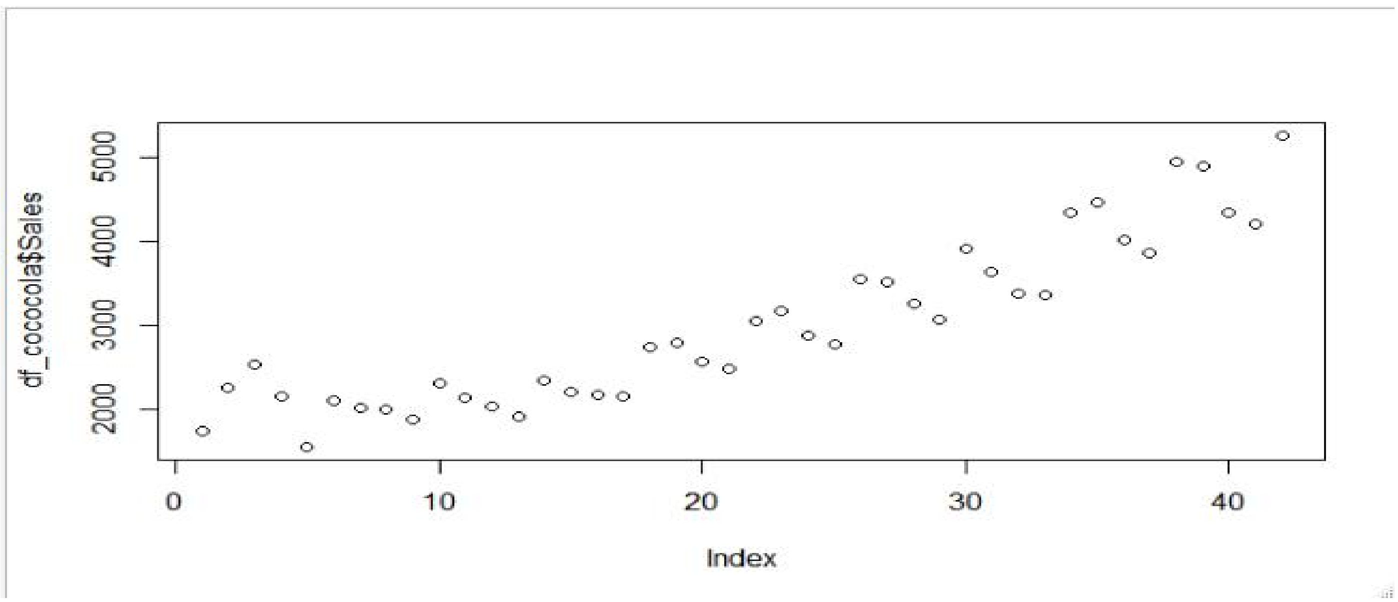
Forecasting

Example-Coca-Cola Dataset

Dataset having 42 records and 2 variables and its collected on Quarterly basis between year 1986 to 1996 and its free from missing values and outliers.



From the above plot, sales are gradually increasing from third cycle i.e. from record 9 which is representing a linear trend as well as the variation within the year is representing the cyclic variation in our data.

Data Preprocessing ➔

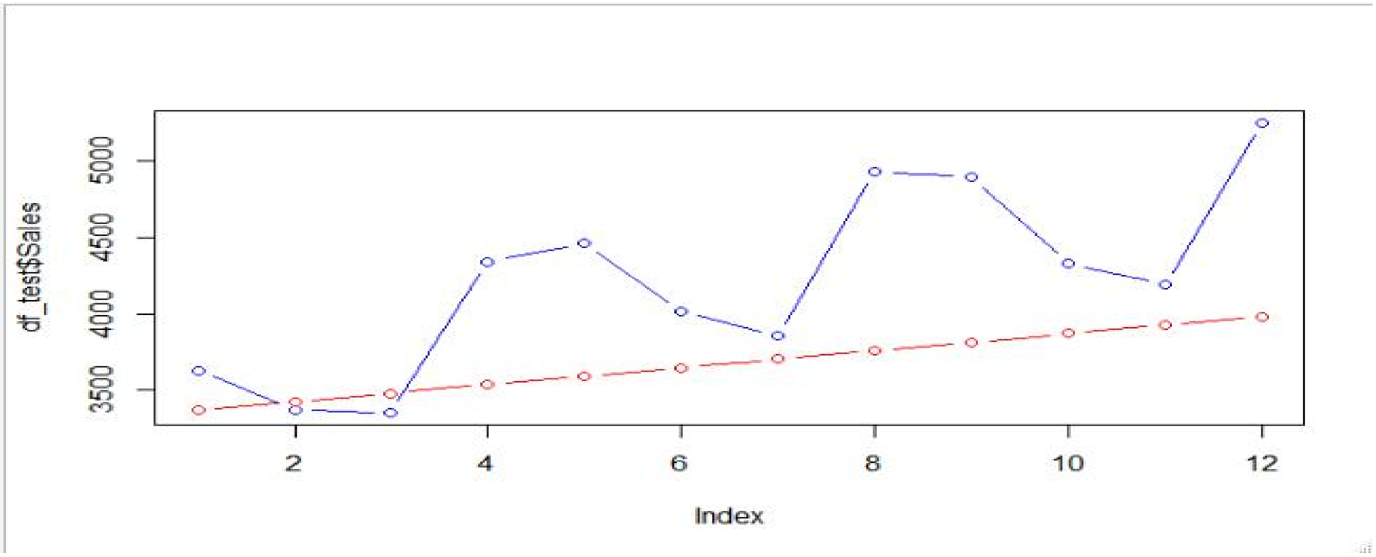
Data is collected quarterly basis so creating 4 dummy variables for 4 quarters.

From the dataset first 30 records considered as train data and balance 12 records as test data.

Linear Trend with Record Number (RN) ➔

`Multiple R-squared:  0.7079,   Adjusted R-squared:  0.6975`

RMSE ➔714.0144

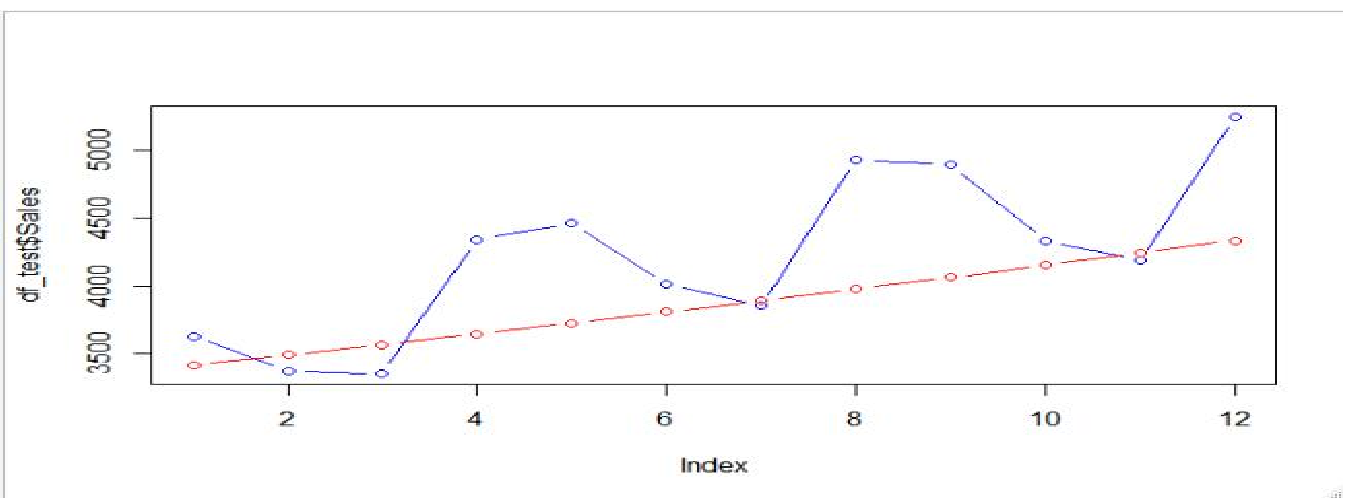


In the above plot, red line is predicted values and blue line is actual values. We unable to explain the cyclic variation due considering the data points 1 to 8 which is erratic nature. So, we will remove in future models.

Exponential Model Considering Only Record Numbers ➔

`Multiple R-squared:  0.7067,   Adjusted R-squared:  0.6962`

RMSE ➔ 552.2821

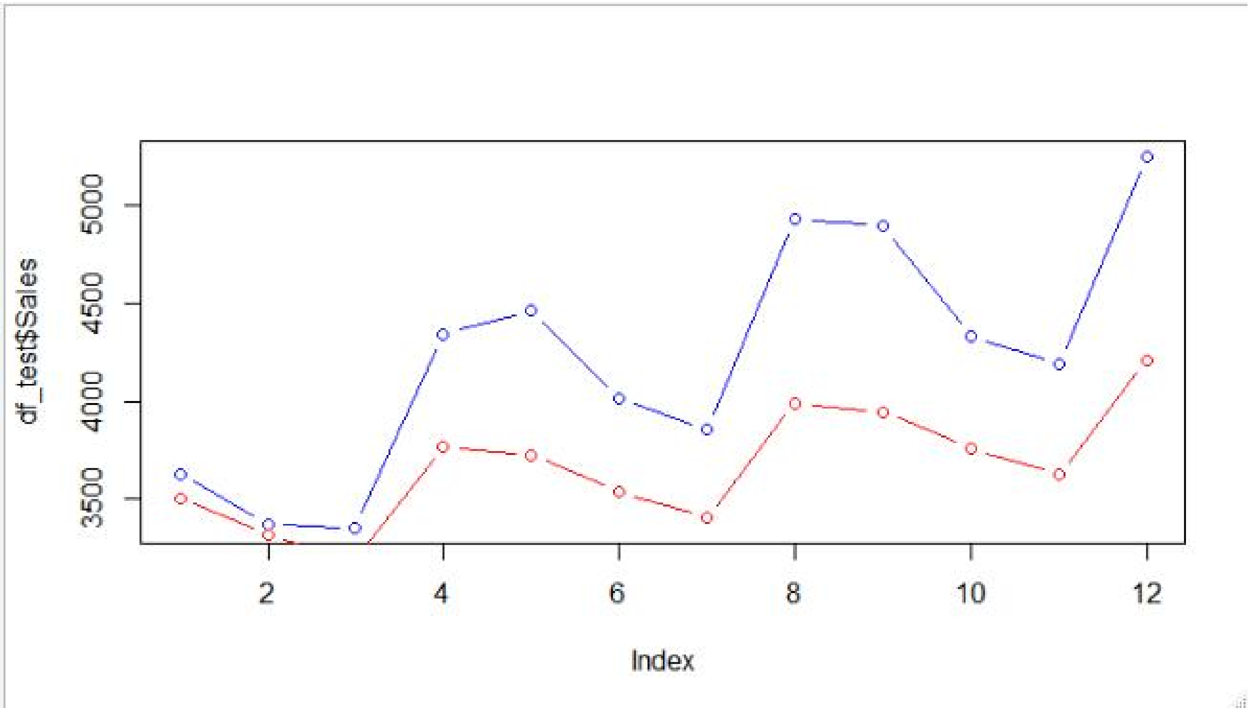


Slightly increment in trend line as well as R^2 than previous model and lesser RMSE. So, we can say that exponential model quite effective than previous linear model.

Additive Seasonality with Linear Trend ➔

```
Multiple R-squared:  0.8457,  Adjusted R-squared:  0.821
```

RMSE ➔ 637.9405



In this model considered dummy variables of Quarters and rn variable to keep linear tendency of prediction.

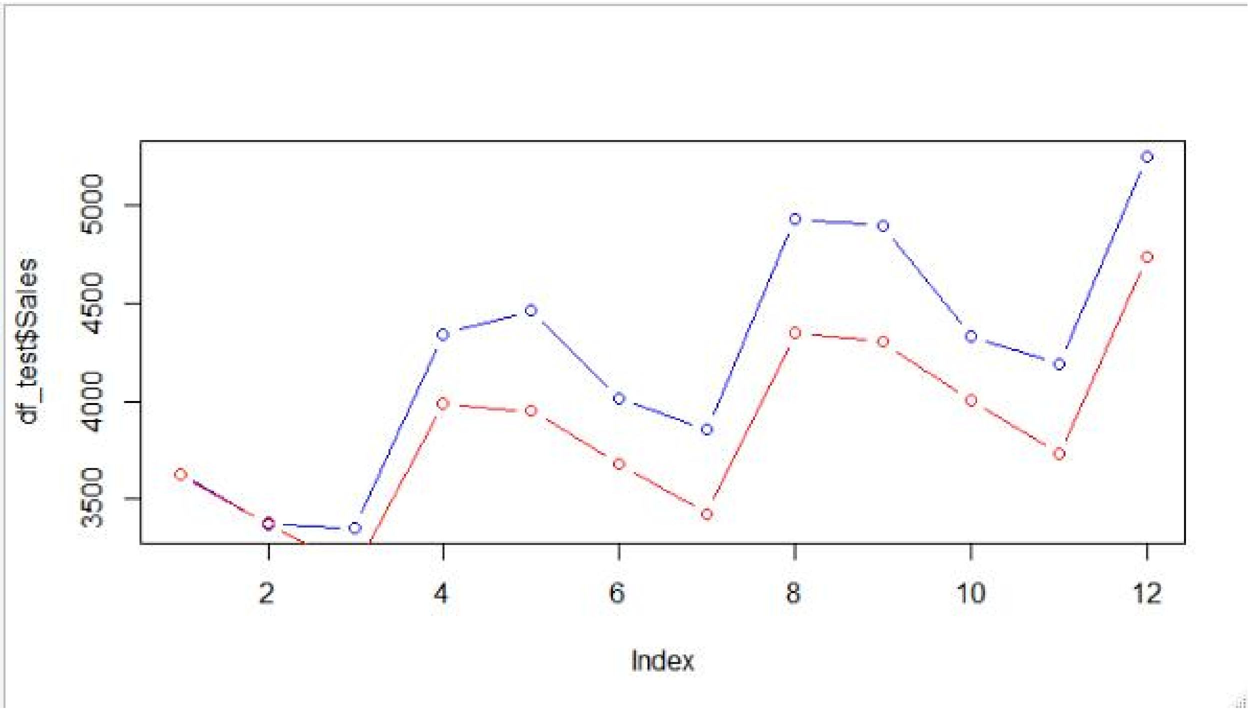
From this plot, we found similar pattern between predicted and actual values but it is still widely separated from actual. So, we will take log of our target variable.

Also, in this model R^2 increased but RMSE also increased which is not good for our model.

Multiplicative Seasonality Linear Trend →

`Multiple R-squared:  0.8586,   Adjusted R-squared:  0.8359`

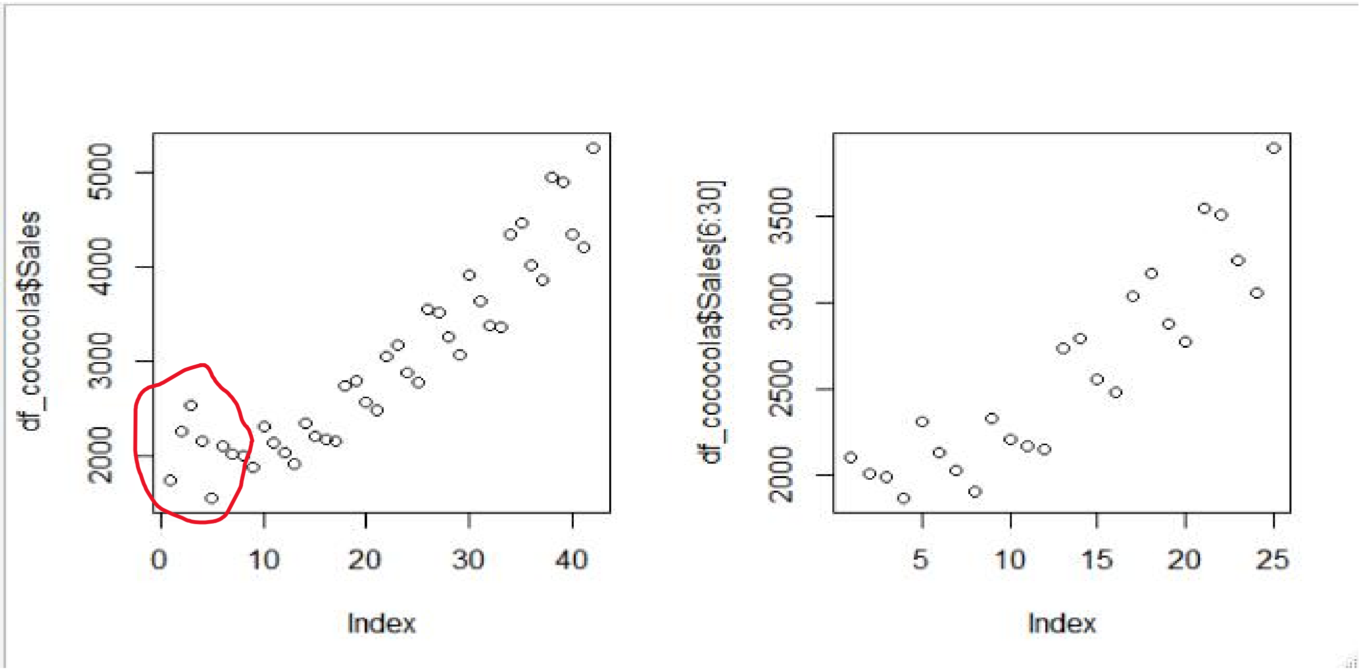
`RMSE → 410.2497`



In this plot gap between predicted and actual values is somewhat Decreased and accuracy also improved than previous model.

Now we will remove erratic variation from our data to improve accuracy i.e. rec ord 1 to 8 which have no definite pattern.

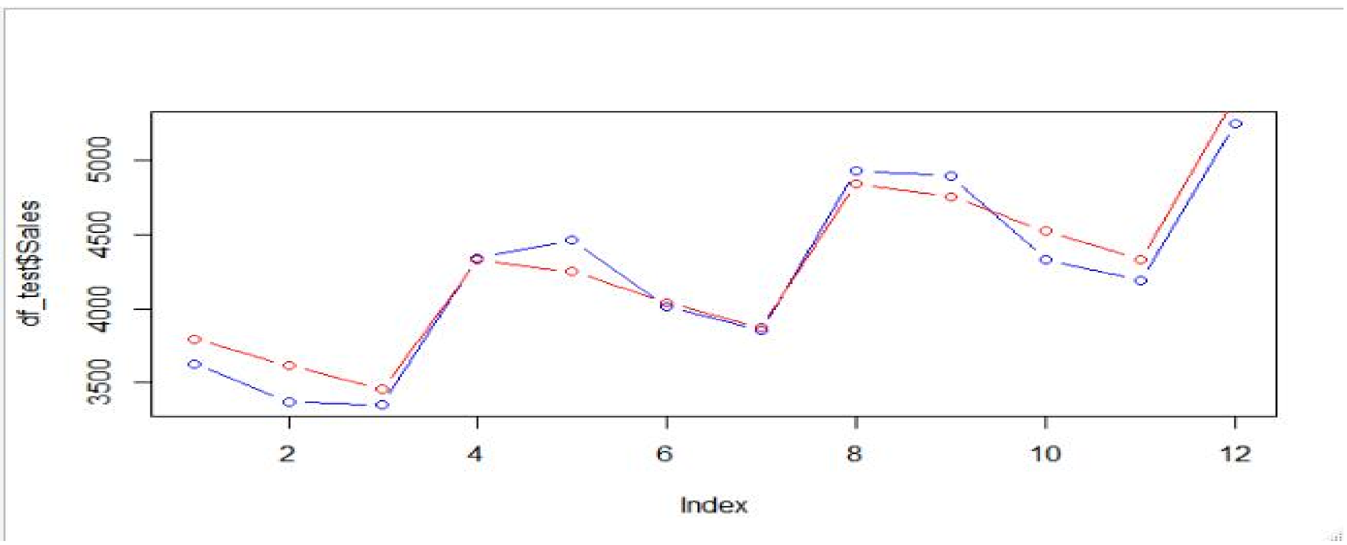
Multiplicative Seasonality Linear Trend with Removing Erratic Components ➔



In the above plot circled data points are neither cyclic nor in linear trend which looks like random and erratic in nature, so I have removed it in my model.

```
Multiple R-squared:  0.9655,   Adjusted R-squared:  0.9578
```

RMSE ➔ 146.4575



Difference between predicted and actual values is negligible in this model.

Model	R^2	RMSE
Linear Trend	**0.7079**	**714.0144**
Exponential Model	**0.7067**	**552.2821**
Additive Seasonality with Linear Trend	**0.8457**	**637.9405**
Multiplicative Seasonality with Linear Trend	**0.8586**	**410.2497**
Multiplicative Seasonality Linear Trend with removing erratic components	**0.9655**	**146.4575**

From the above information we can infer than Model 5 is our final best model with high R^2 and least RMSE.