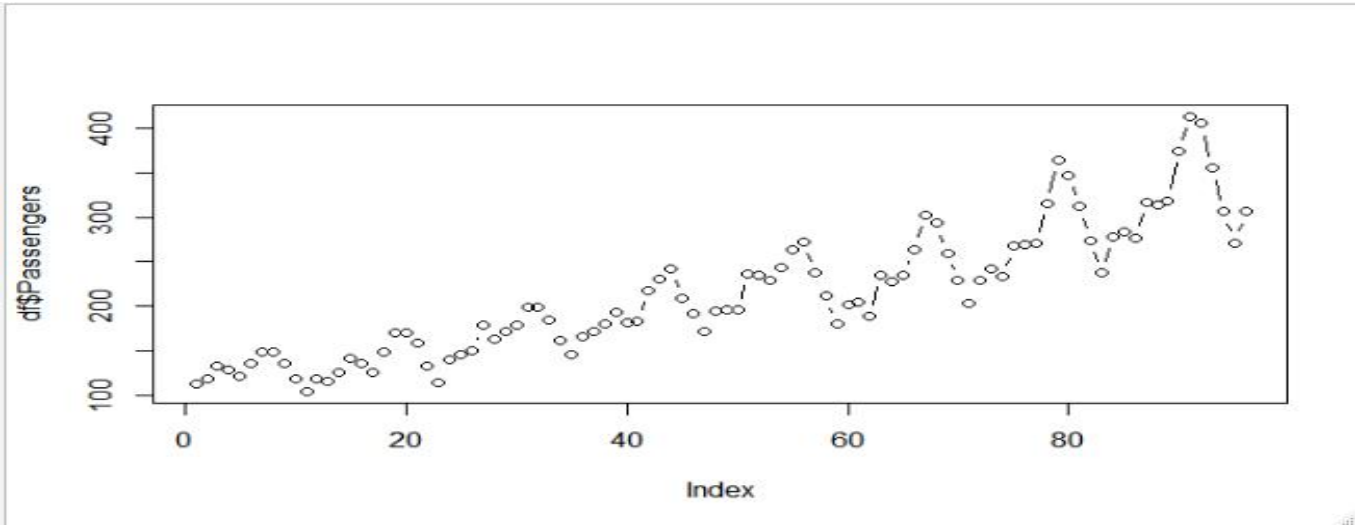
Forecasting

Example- Airline Dataset

In this data set 2 columns as Month and Passengers are available.

Passengers is our target variable.

Data is collected on 1st day of each month from year 1995 to 2002.



From the above plot we can say that increasing trend fluctuate with seasonality changes, also cyclic variation and linear increasing trend of passenger numbers.

Data Preprocessing ➔

I have created 12 dummy variables for 12 months.

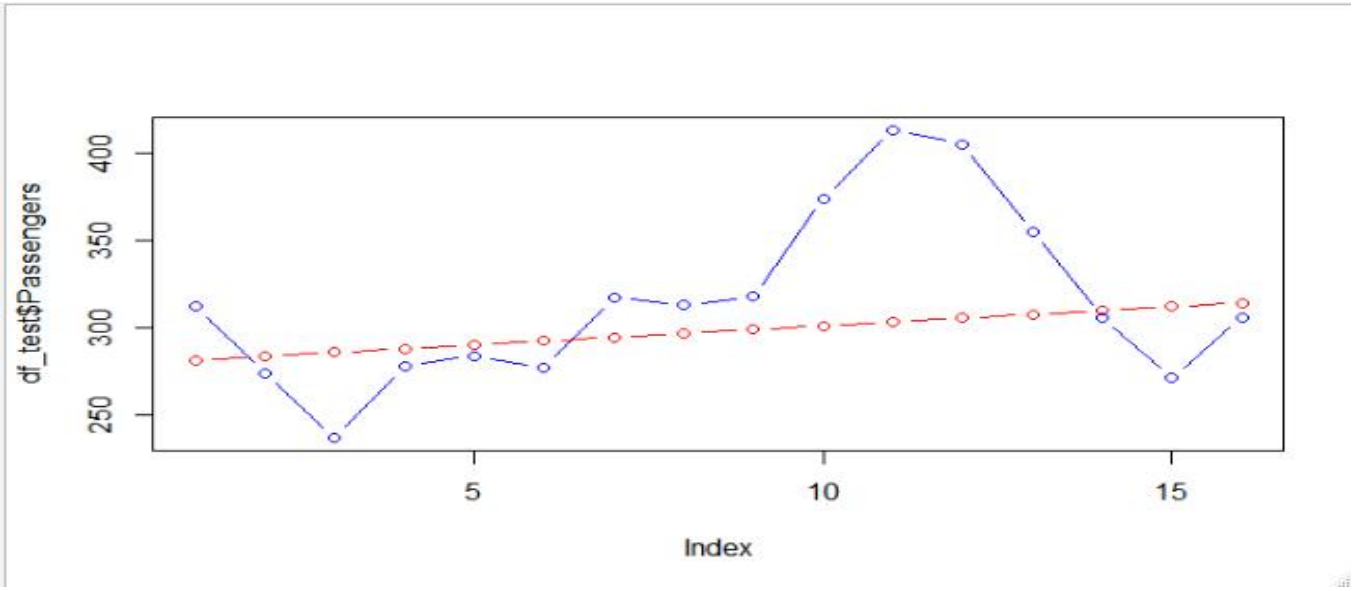
Also created a variable day and we can say that data are collected in the interval of as on average 30 days, so it will represent the rough intervals of data collection days taking initial day as 1.

Out of 96 records, first 80 records considered as train data and rest 16 as test data.

Linear Trend with Variable Day ➔

Multiple R-squared: 0.7829, Adjusted R-squared: 0.780

RMSE ➔ 47.5462

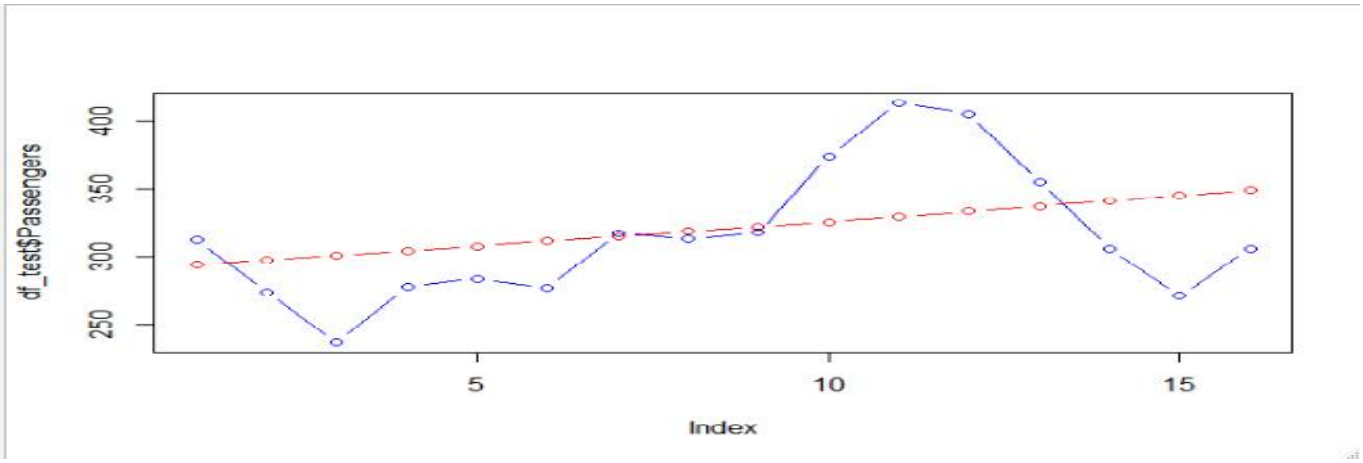


In the above plot, red line is predicted value and blue line actual value.
From the above plot we unable to explain the seasonal variation in our data.

Exponential Model with Variable Day ➔

Multiple R-squared: 0.8181, Adjusted R-squared: 0.8158

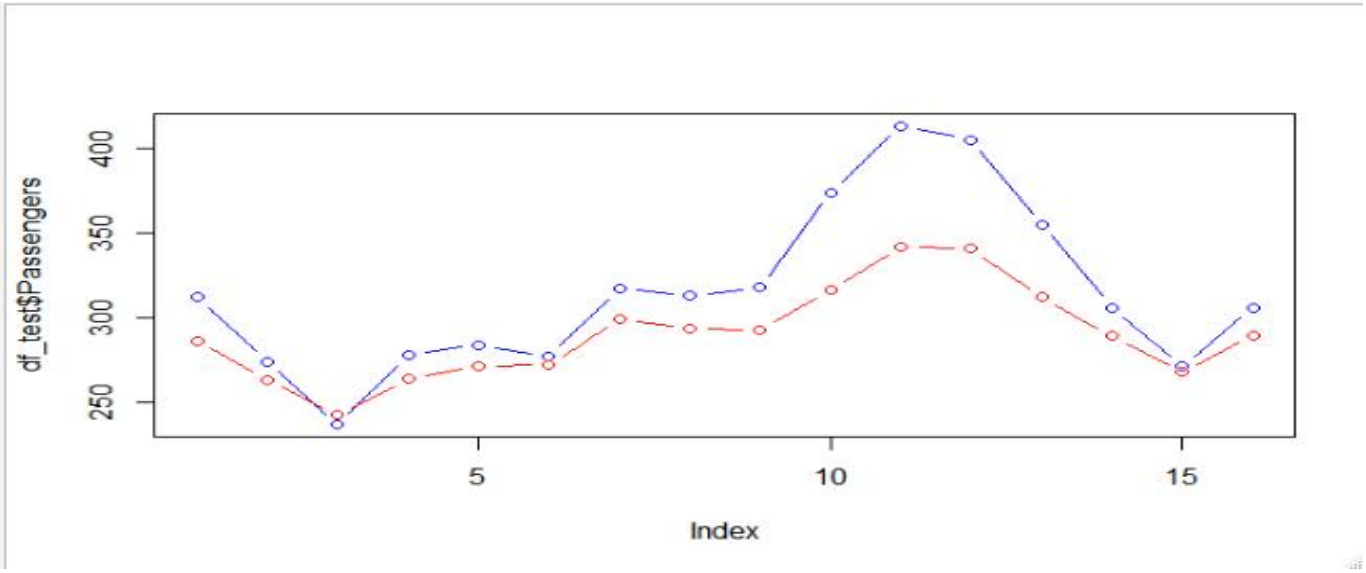
RMSE ➔ 43.79374



Additive Seasonality with Linear Trend with All Data ➔

Multiple R-squared: 0.9533, Adjusted R-squared: 0.9449

RMSE ➔ 33.04



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

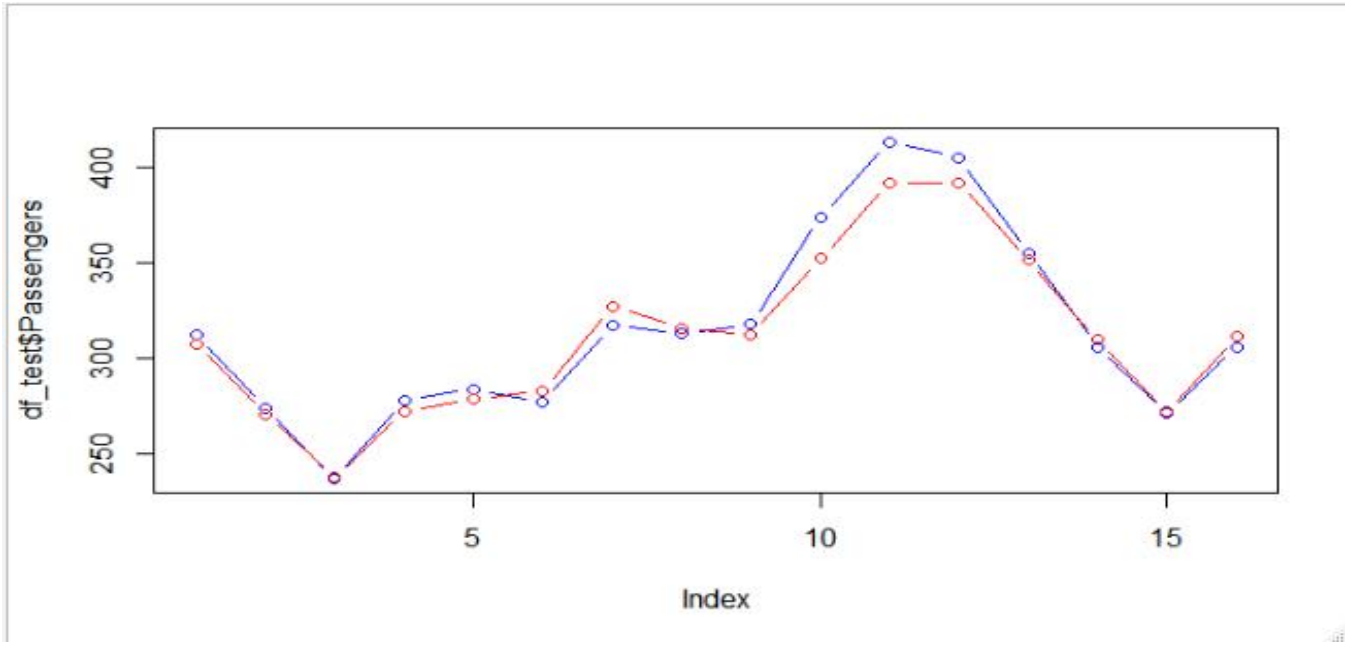
Prediction is good and pattern is matching with original the original data but still unable to explain the extreme values properly.

Multiplicative Seasonality Linear Trend with All the Variables➔

Multiple R-squared: 0.9743, Adjusted R-squared: 0.9696

RMSE ➔ 9.469

In the below plot, predicted values are nearly equal to actual values, So we are getting our desired efficiency.



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

From the above information we can infer that Model 4 is our best model with high R^2 and low RMSE and actual vs prediction graph.