Retail Giant Sales Forecasting

Submitted by Kratika Sharma

Problem Statement I - Data Preparation

Business problem : Global Mart is an online supergiant store that has worldwide operations. This store takes orders and delivers across the globe and deals with all the major product categories — consumer, corporate and home office.

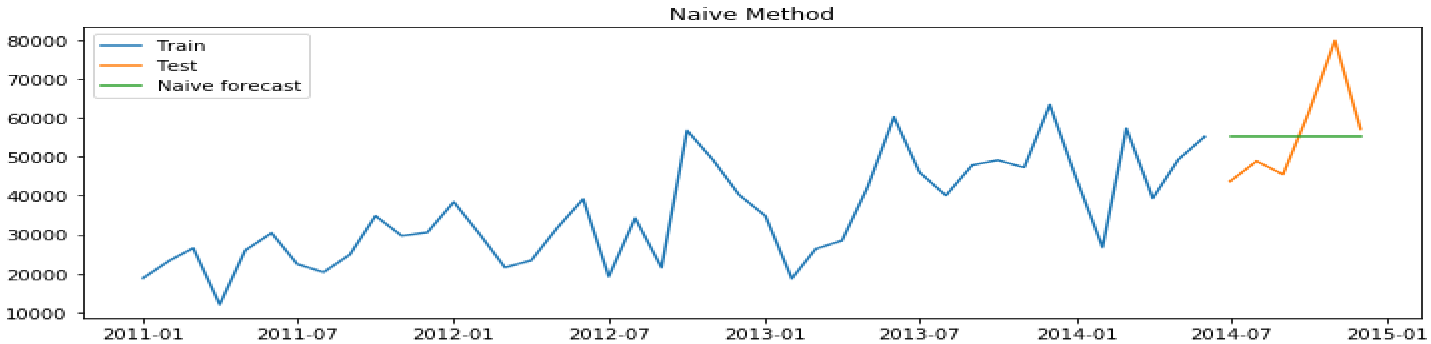
As a sales manager for this store, you have to forecast the sales of the products for the next 6 months, so that you have a proper estimate and can plan your inventory and business processes accordingly.

Data Preparation Steps :

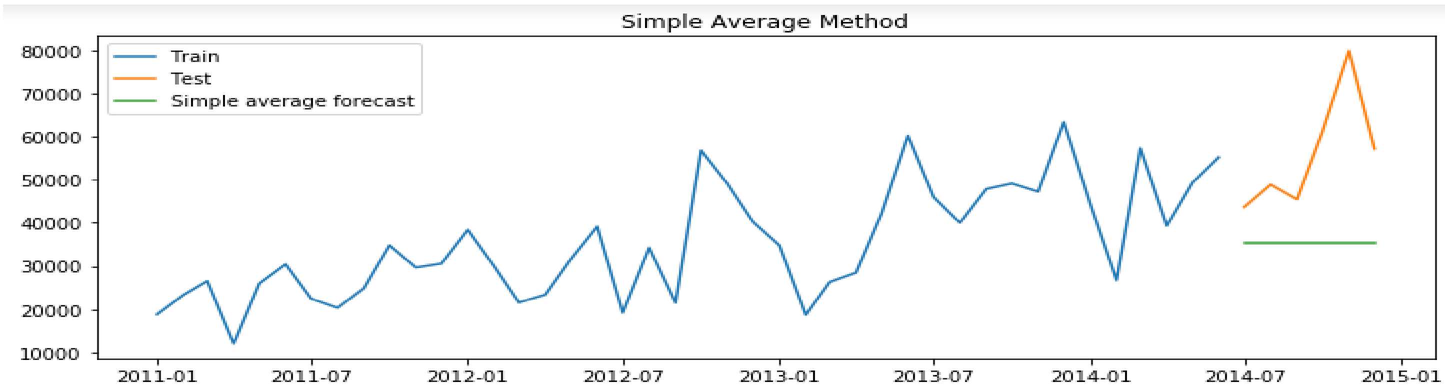
- Create a new column Market-Segment
- Find coefficient of variation for different market segments
- Choose the most consistently profitable market- segment using cov value
- Filter entire data for most profitable market segment.
- Aggregate the data into monthly level by taking sum
- Divide data into test and train
- Train all models on train data.

Problem Statement II-Model Building and Evaluation

- **Naive Method :** We can observe in below plot that the forecast value using naïve forecast is similar to the last value of train data. Hence it is under forecasting the value.
- MAPE : 17.47 and RMSE : 12355.97

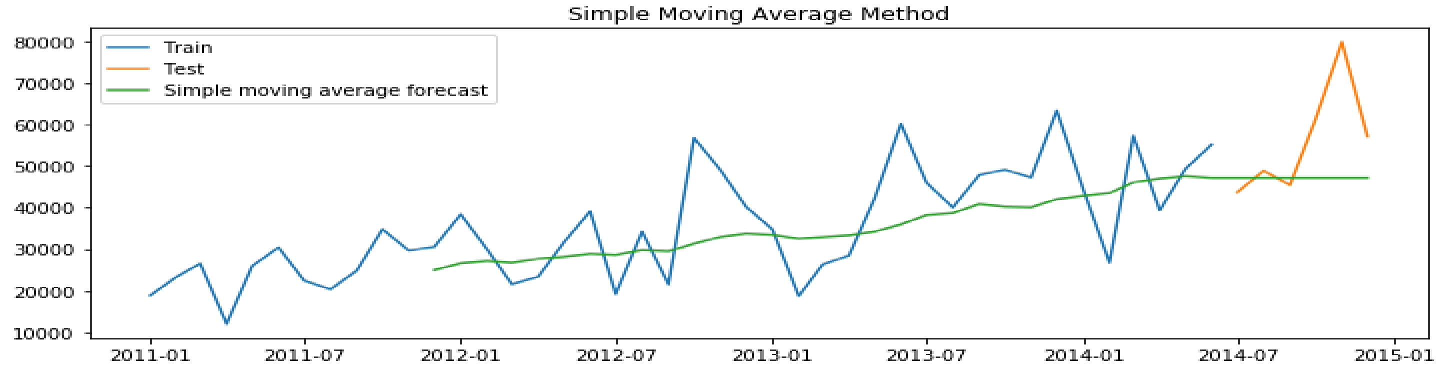


- **Simple average method :** We noticed that our data has some trend and because of that our forecast value is underestimated here.
- MAPE : 34.34 and RMSE : 24146.06

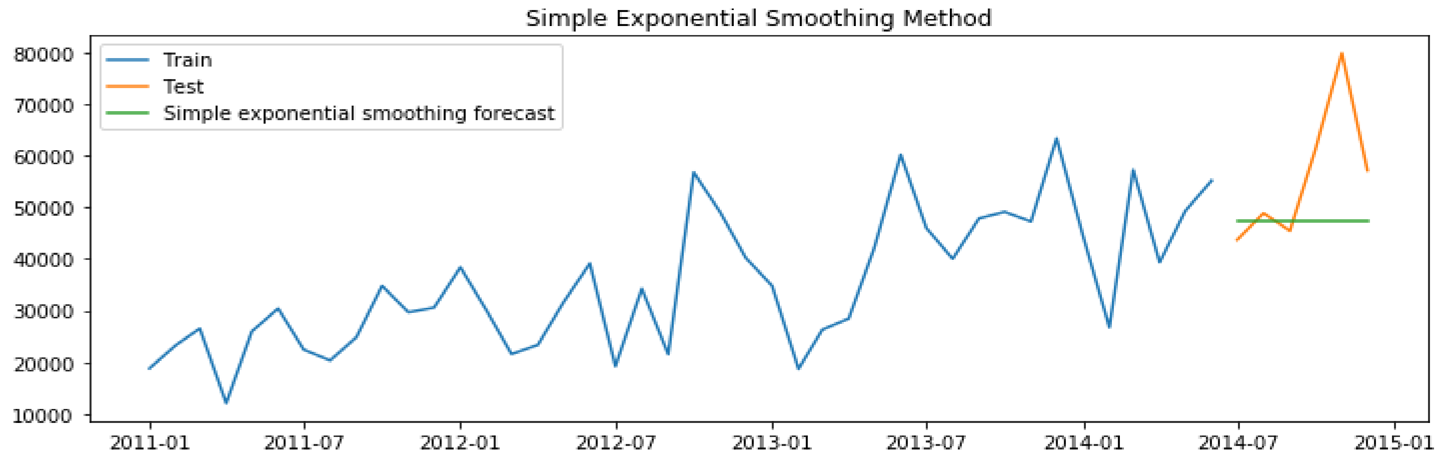


Problem Statement II-Model Building and Evaluation

- **Simple moving average method :** We have set moving average window to be 12 months and here we able to notice that we are getting better forecast here as compare to naïve and simple average method. We can also achieve little bit of seasonality by reducing the moving average window.
- **RMSE :** 15192 , **MAPE** : 16.10

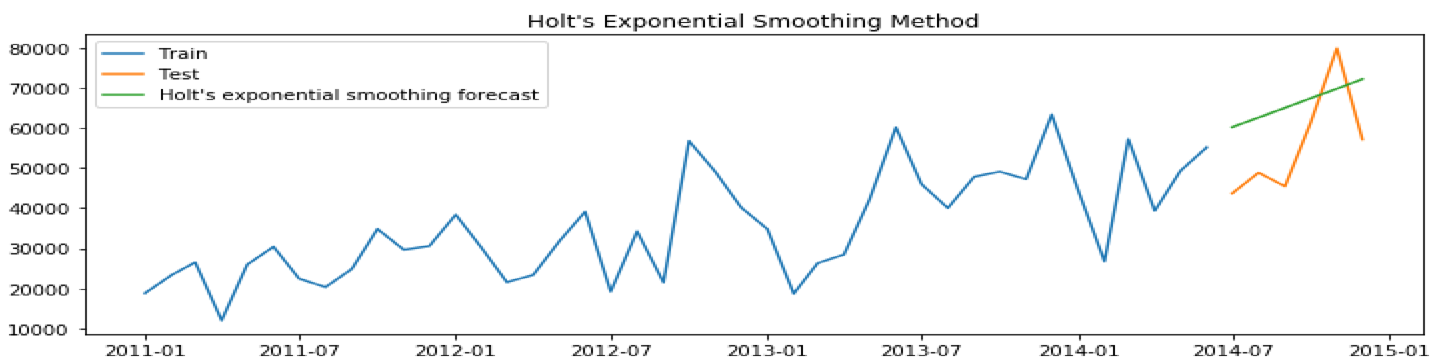


- **Simple Exponential Smoothing Technique :** We can observe here that we are capturing level in simple exponential smoothing technique. It does not include trend or seasonality.
- **RMSE:** 15011.49 **MAPE:** 15.99

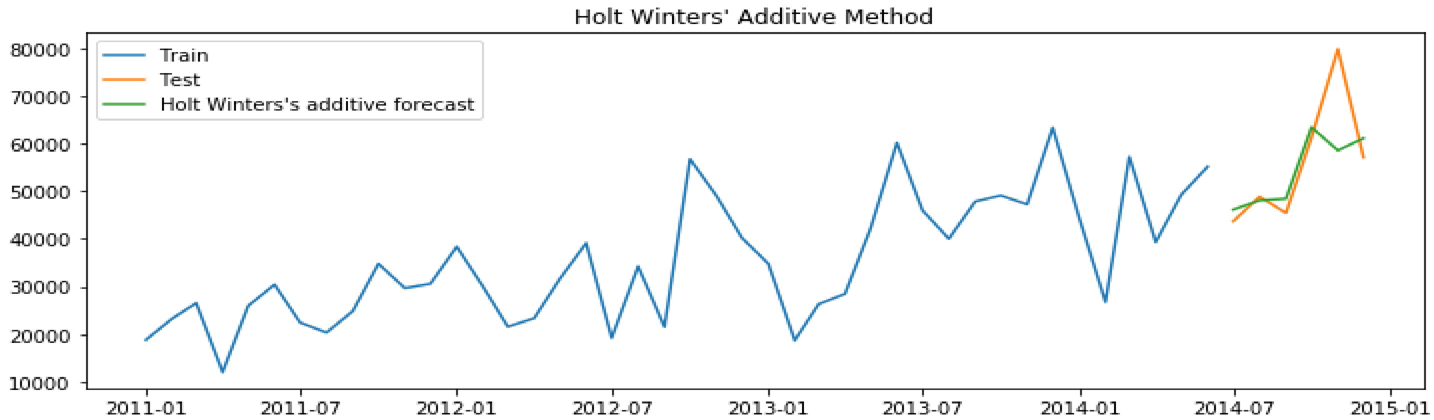


Problem Statement II-Model Building and Evaluation

- **Holt's method :** Here we gave smoothing level as 0.2 that means we are adding more weight to the last observation and smoothing slope as 0.2 that means if we calculate delta between current and one previous observation. Then we assigning 0.02 weight to the last one. Hence It doesn't capture seasonality.
- **RMSE :** 18976.35, **MAPE** : 34.57

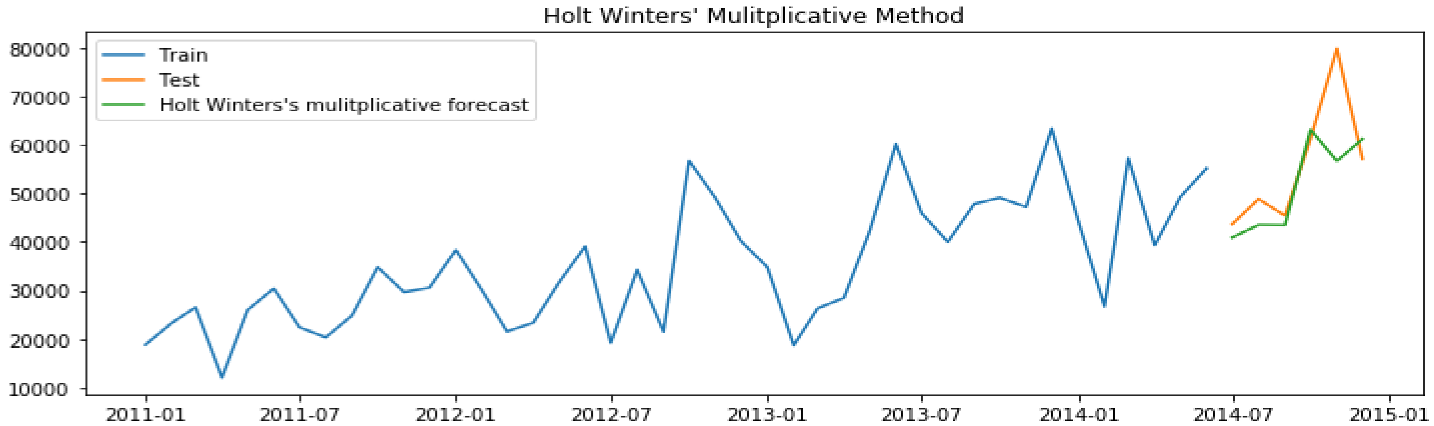


- **Holt Winters' additive method:** This method has smoothing level is 1.12, smoothing slope is 1.12, smoothing seasonal is 0.0 and seasonal periods=12. Here we are able to capture the seasonality as well.
- **RMSE :** 9037.44, **MAPE** : 8.52



Problem Statement II-Model Building and Evaluation

- **Holt Winter's multiplicative method:** Here we add seasonal parameter as mul. Using this method we are able to capture Level, Trend and Seasonality component.
- **RMSE :** 9976.32, **MAPE** : 10.12



- **Stationary Tests:** Taking threshold as 0.01. We used two method for stationary test :-
  - **Augmented Dickey-Fuller (ADF) test :** Here the P-value is greater then 0.01 hence we cannot reject the null hypothesis and series is not stationary.

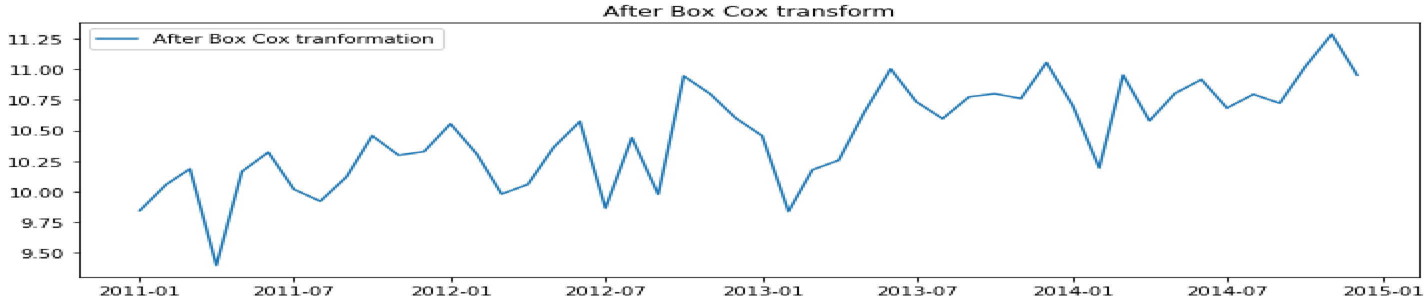
```
ADF Statistic: -3.376024
Critical Values @ 0.05: -2.93
p-value: 0.011804
```

  - **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test :** Here the P-value is greater then 0.01 hence we can reject the null hypothesis and series is not stationary.

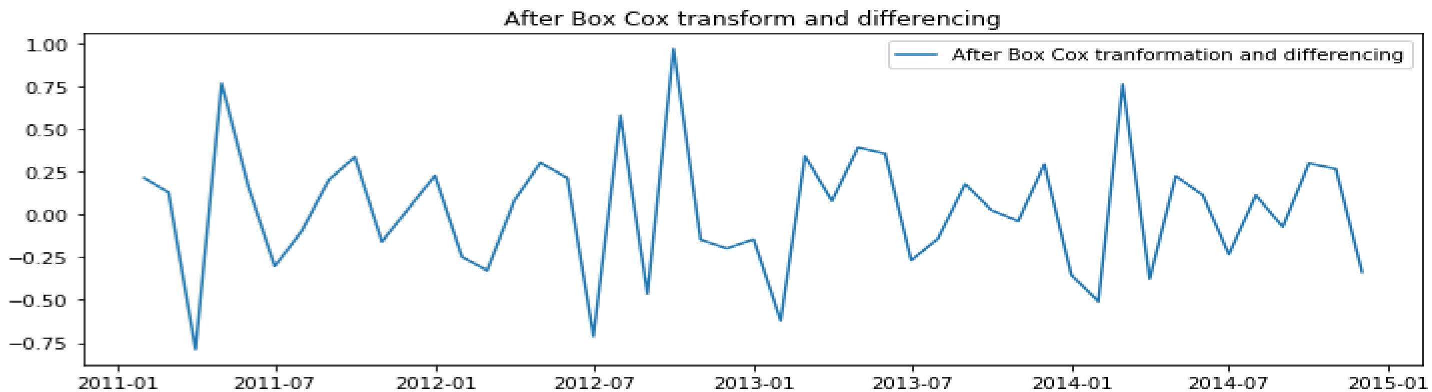
```
KPSS Statistic: 0.577076
Critical Values @ 0.05: 0.46
p-value: 0.024720
```

Non-Stationary to Stationary

- **Box Cox transformation to make variance constant:** By using box cox transformation we are able to make variance constant in series.



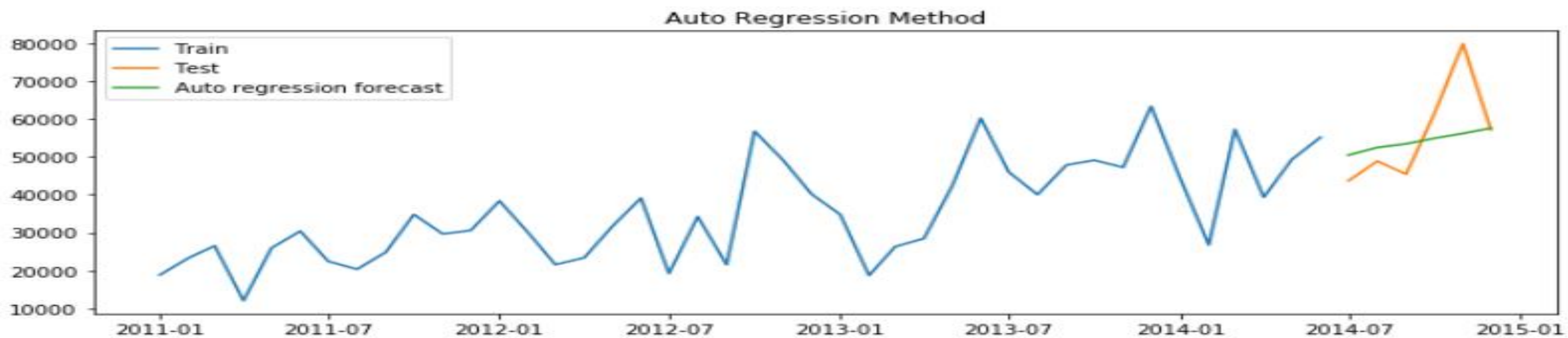
- **Differencing to remove trend :** We are able to remove trend from series using differencing.



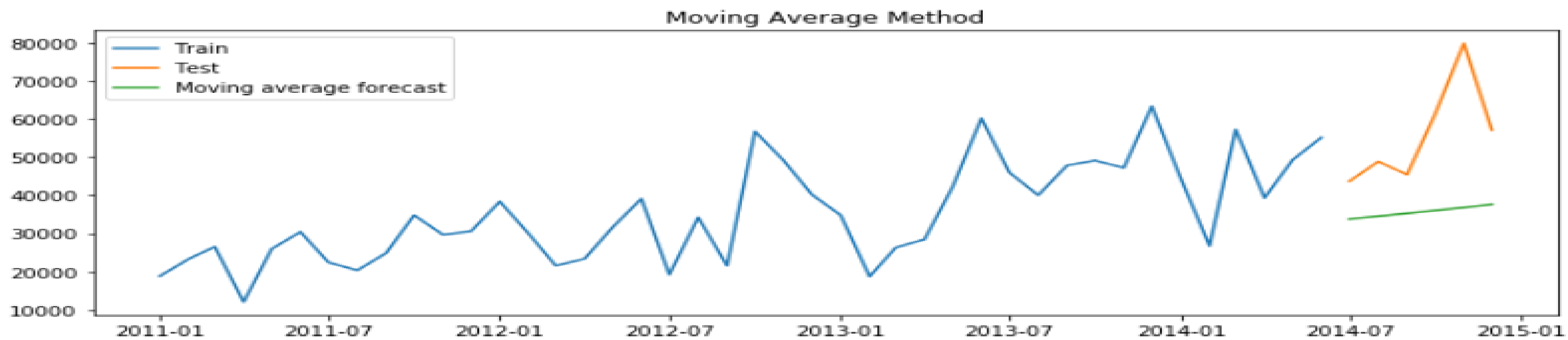
- After Box Cox transformation and Differencing we are getting p value from Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test respectively is 0.000170 and 0.100000 hence we can say that now the series is stationary.

Auto Regressive models

- **Auto regression method (AR) :** We are using order as 1,0,0 where lag order(p) is equal to 1. We are able to capture the trend of the data here.
- **RMSE :** 10985.28, **MAPE** : 13.56

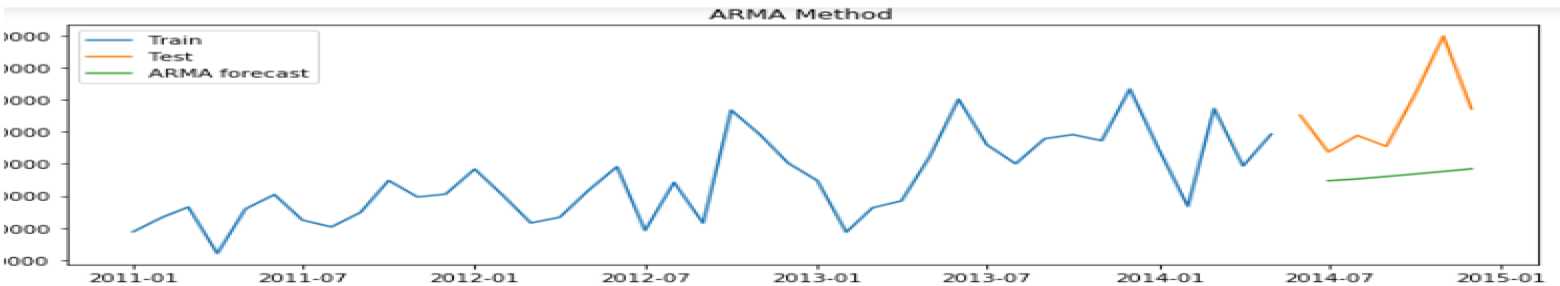


- **Moving average method (MA):** We are using order as 0,0,1 where window size(q) is equal to 1. Here the model under predict the values
- **RMSE :** 23360.02, **MAPE** : 33.93

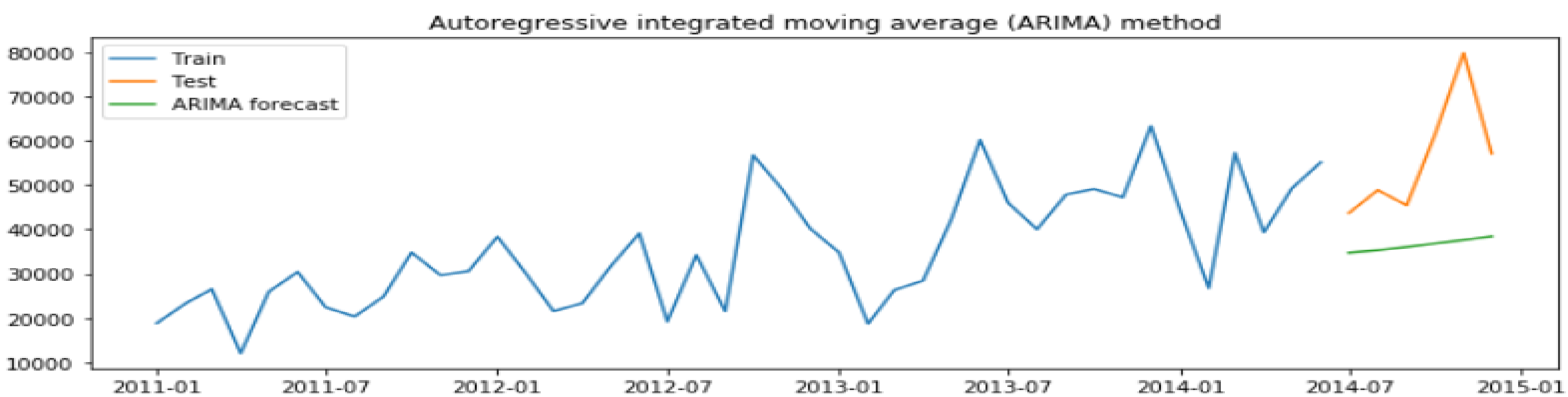


Auto Regressive models

- Auto regression moving average method (ARMA) : It has two parts auto regressive and moving average. We are passing order 1, 0, 1 where p = 1 and q =1. We can observe here that it under predict the value of series
- **RMSE :** 10985.28, **MAPE** : 13.56



- **Auto Regressive Integrated Moving Average (ARIMA) :** It has three part Auto regression, Intergraded, with Moving average. We are passing order 1,1,1 where p=1,q=1,d=1. Here we get the similar output as ARMA because we are using d = 1. So basically we are using the same data.
- **RMSE :** 10985.28, **MAPE** : 13.56



Auto Regressive models

- **Seasonal auto regressive integrated moving average (SARIMA) :** We pass order as 1,1,1 where p =1 , q= 1,d = 1 and it also include seasonal order parameter which is 1,1,1,12. Here we observe that it is able to capture level , trend and seasonality quit well.
- **RMSE :**9617.18, **MAPE** : 12.88