Retail Giant Time Series Forecast

FORECASTING THE MOST PROFITABLE MARKET SEGMENT SALES

RENU SINGH THAKUR

Problem Statement

- 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.
- Usage of COV to get the most profitable market segment to forecast sales for next 6 months

Approach

- Review data
- Aggregate data for 48 months for each Market Segment
- Prepare of train test data
- Calculate COV - Coefficient of Variance
- Apply all the time series forecast methods on Most Profitable market segment
- Store all RMSE & MAPE forecasting errors in data frame to get the best method for train data

Market Segments

APAC-Consumer
LATAM-Consumer
US-Consumer
EU-Consumer
APAC-Corporate
EU-Corporate
LATAM-Corporate
US-Corporate
EMEA-Consumer
Africa-Consumer
APAC-Home Office
LATAM-Home Office
US-Home Office
EU-Home Office
EMEA-Corporate
Africa-Corporate
EMEA-Home Office
Africa-Home Office
Canada-Consumer
Canada-Corporate
Canada-Home Office

Data Set

Total Dataset size – 51290 rows

	Order Date	Segment	Market	Sales	Profit
0	31-07-2012	Consumer	US	2309.650	762.1845
1	05-02-2013	Corporate	APAC	3709.395	-288.7650
2	17-10-2013	Consumer	APAC	5175.171	919.9710
3	28-01-2013	Home Office	EU	2892.510	-96.5400
4	05-11-2013	Consumer	Africa	2832.960	311.5200

Dataset after adding new columns

	Order Date	Segment	Market	Sales	Profit	Market_Segment	OrderMonthYear
0	31-07-2012	Consumer	US	2309.650	762.1845	US-Consumer	2012-07
1	05-02-2013	Corporate	APAC	3709.395	-288.7650	APAC-Corporate	2013-02
2	17-10-2013	Consumer	APAC	5175.171	919.9710	APAC-Consumer	2013-10
3	28-01-2013	Home Office	EU	2892.510	-96.5400	EU-Home Office	2013-01
4	05-11-2013	Consumer	Africa	2832.960	311.5200	Africa-Consumer	2013-11

Aggregated data for 21 market segments for 48 months

Market_Segment	APAC-Consumer	APAC-Corporate	APAC-Home Office	Africa-Consumer	Africa-Corporate	Africa-Home Office	Canada-Consumer	Canada-Corporate	Canada-Home Office	EMEA-Consumer	...	EMEA-Home Office
OrderMonthYear												
2011-01-01	15711.7125	3374.2098	3973.6623	7909.083	1760.046	2071.770	314.22	16.29	NaN	2790.456	...	299.490
2011-02-01	12910.8588	18157.2654	5869.0272	4886.136	1087.899	2942.562	56.91	NaN	440.52	1287.510	...	1271.310
2011-03-01	19472.5632	8769.7386	4817.5392	2656.830	1073.934	163.680	1405.26	NaN	174.96	9696.108	...	1235.982
2011-04-01	15440.3046	8985.6765	5739.2580	4004.082	3767.901	2710.446	286.08	NaN	NaN	1769.001	...	1364.640
2011-05-01	24348.9723	20841.3672	1909.3983	5011.614	1210.308	487.476	752.01	NaN	NaN	3716.592	...	2338.068

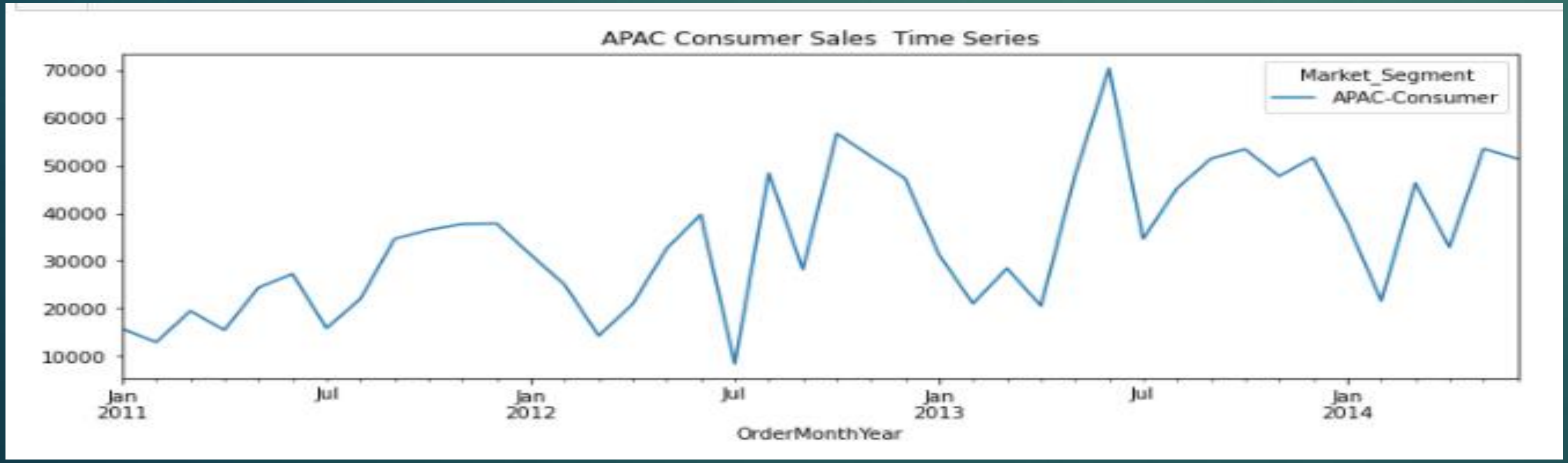
Data Split in Train & Test

```
# Getting 42 months of data as train and 6 months as test
train_len=42
train_sales=df[:train_len]
test_sales=df[train_len:]
train_sales.shape , test_sales.shape
```

((42, 21), (6, 21))

Coefficient of Variance (COV)

- Ratio of Standard Deviation to Mean for each Market Segment
- Lower the COV , lesser the variation in data and we can consider that market segment as most profitable
- For each market segment calculated the ratio of standard deviation and mean
- Got the lowest COV as 0.42 for APAC-Consumer
- Created new train & test dataset based on APAC-Consumer



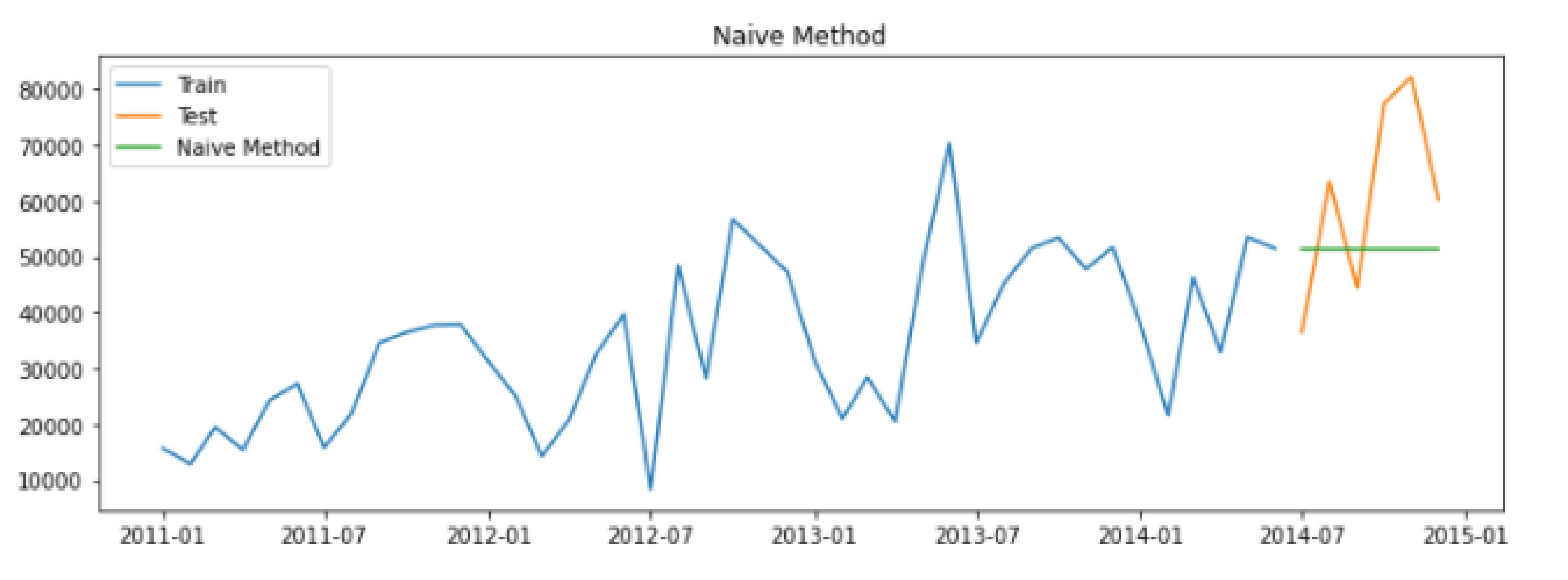

```
[('APAC-Consumer', 0.42),
 ('EU-Consumer', 0.46),
 ('APAC-Corporate', 0.48),
 ('LATAM-Consumer', 0.48),
 ('EU-Corporate', 0.51),
 ('APAC-Home Office', 0.54),
 ('Africa-Consumer', 0.56),
 ('EMEA-Corporate', 0.56),
 ('EU-Home Office', 0.59),
 ('US-Consumer', 0.59),
 ('LATAM-Home Office', 0.6),
 ('US-Corporate', 0.6),
 ('EMEA-Consumer', 0.62),
 ('LATAM-Corporate', 0.62),
 ('Africa-Corporate', 0.73),
 ('Africa-Home Office', 0.74),
 ('US-Home Office', 0.78),
 ('EMEA-Home Office', 0.84),
 ('Canada-Corporate', 0.94),
 ('Canada-Consumer', 1.14),
 ('Canada-Home Office', 1.65)]
```

```
# Preparing new train & test for most profitable market segment
train=train_sales[['APAC-Consumer']]
test=test_sales[['APAC-Consumer']]
test.shape , train.shape
```

```
((6, 1), (42, 1))
```

Smoothing Technique : Naive Method

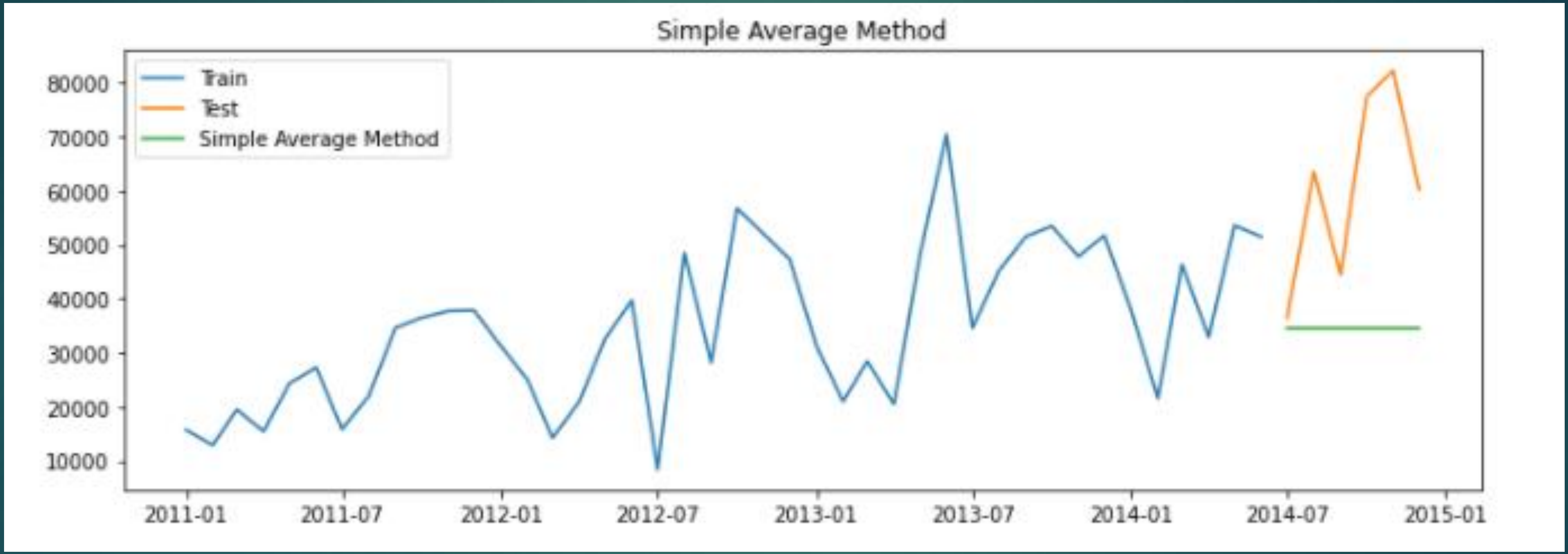
- Naïve method is getting the last value from forecasted attribute , in this case Sales values for the last month of the data set



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86

Smoothing Technique : Simple Average Method

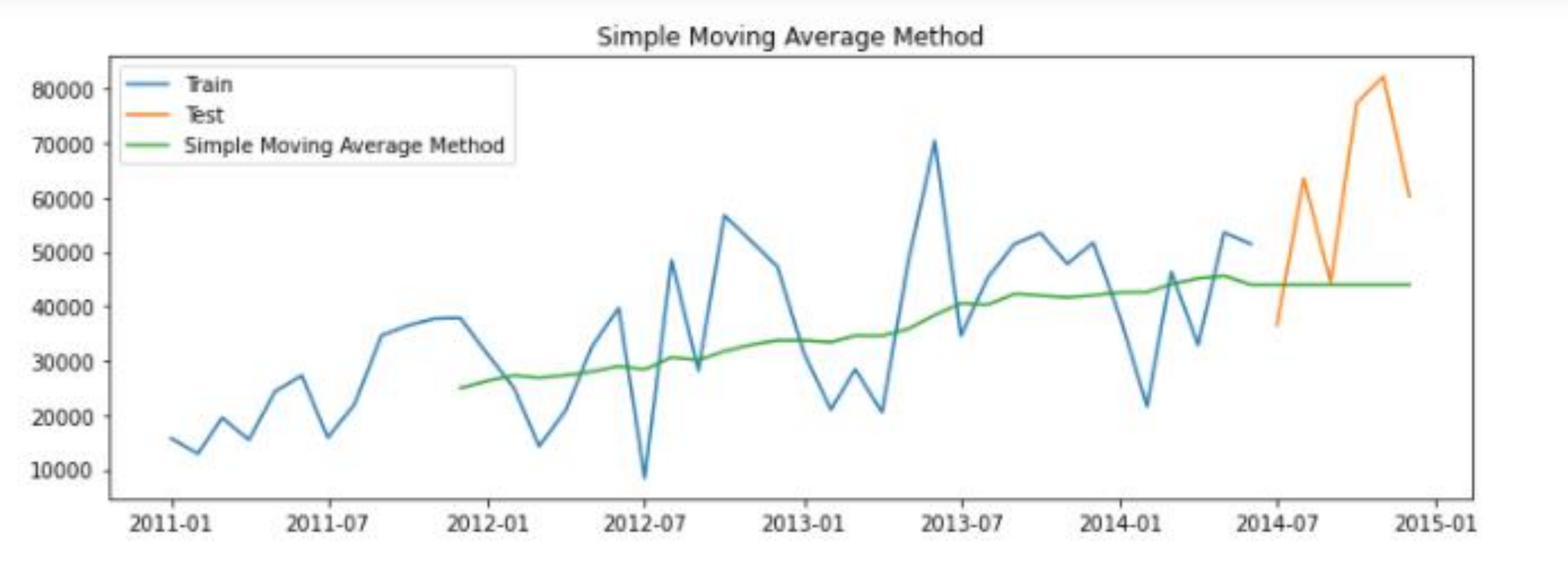
- Simple average method is to get the mean of the forecasted value .



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18

Smoothing Technique : Simple Moving Average Method

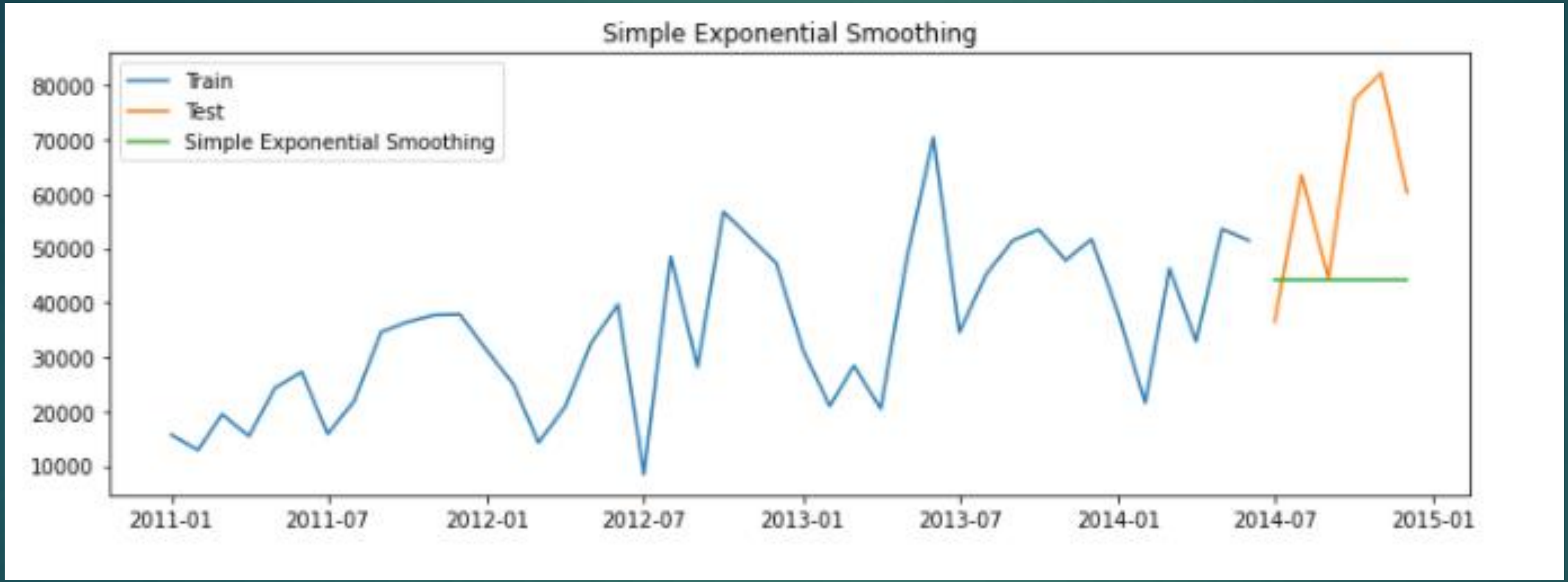
- In Simple moving average the forecasts are calculated using the average of the time-series data in the moving window considered. Window Size = 12



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15

Smoothing Technique : Simple Exponential Smoothing

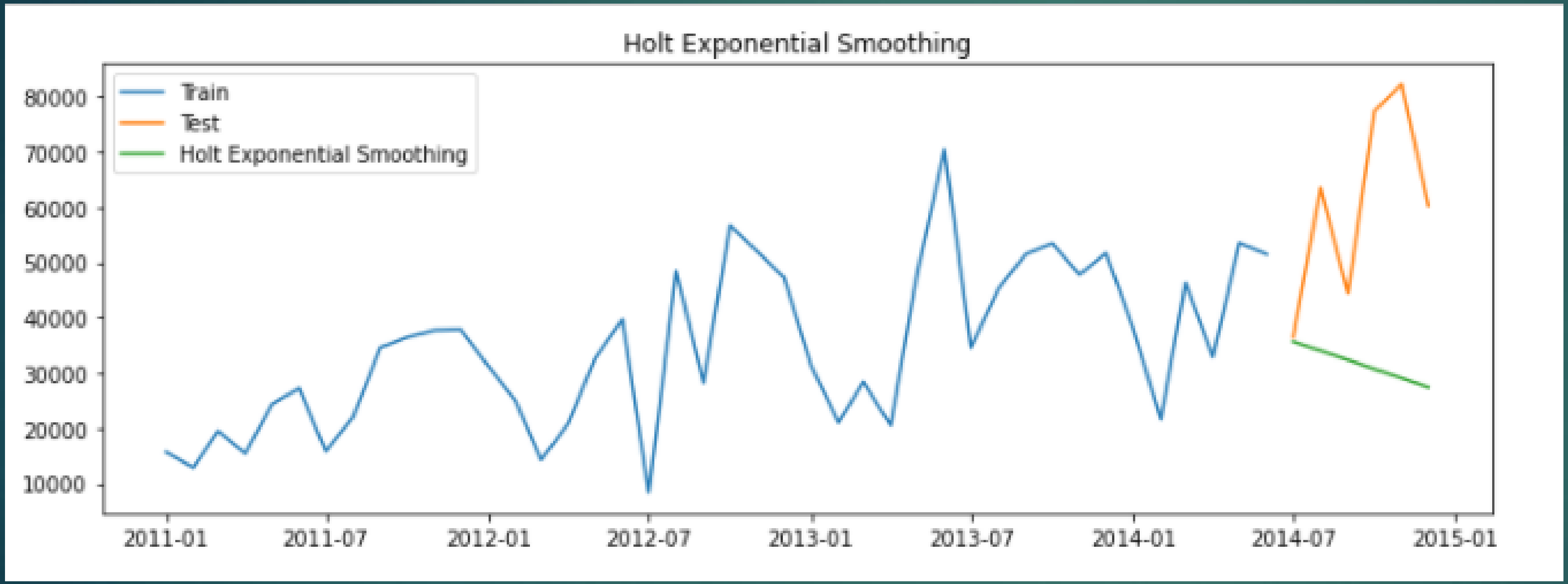
- This is weighted average technique; more weights are given to recent values and lesser weights to past data points. It helps us forecast the level in the time series data



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82

Smoothing Technique : Holt's Winter Exponential Smoothing

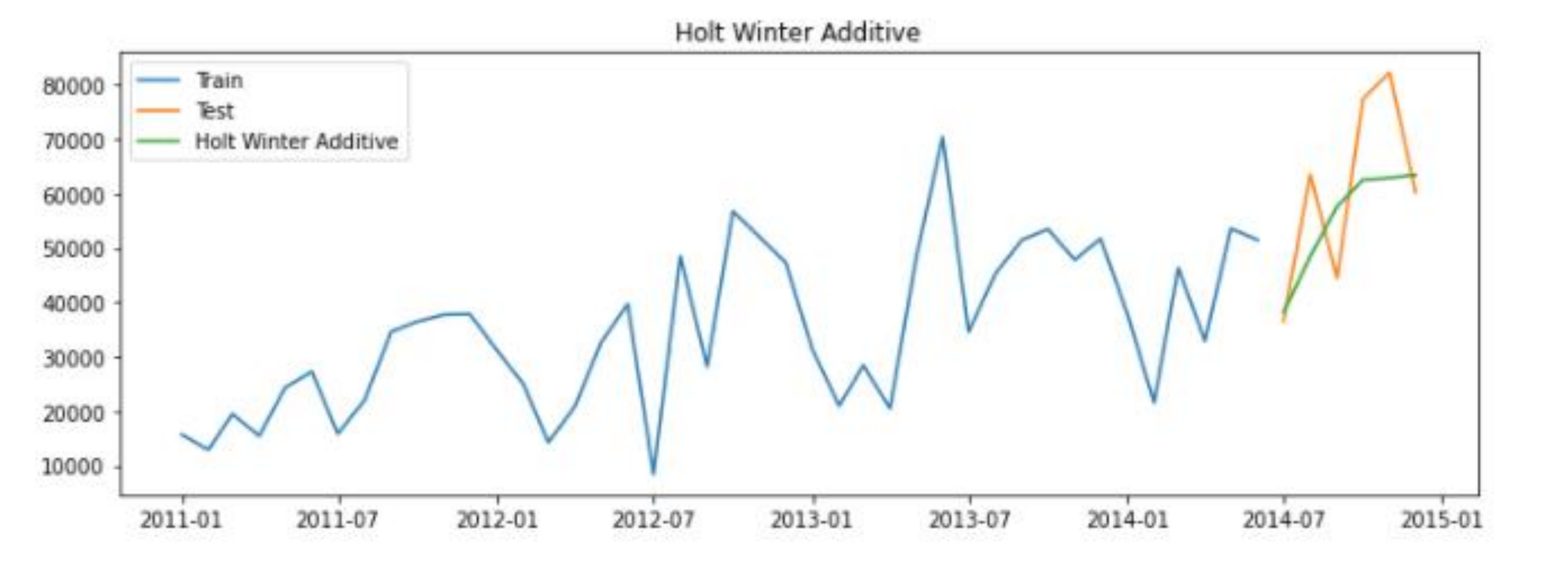
- This forecasts the level, trend as well as the seasonality for a time series data.
- There are two methods of performing the Holt-Winters' smoothing techniques: additive and multiplicative methods.



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57

Smoothing Technique : Holt's Winter Additive Smoothing

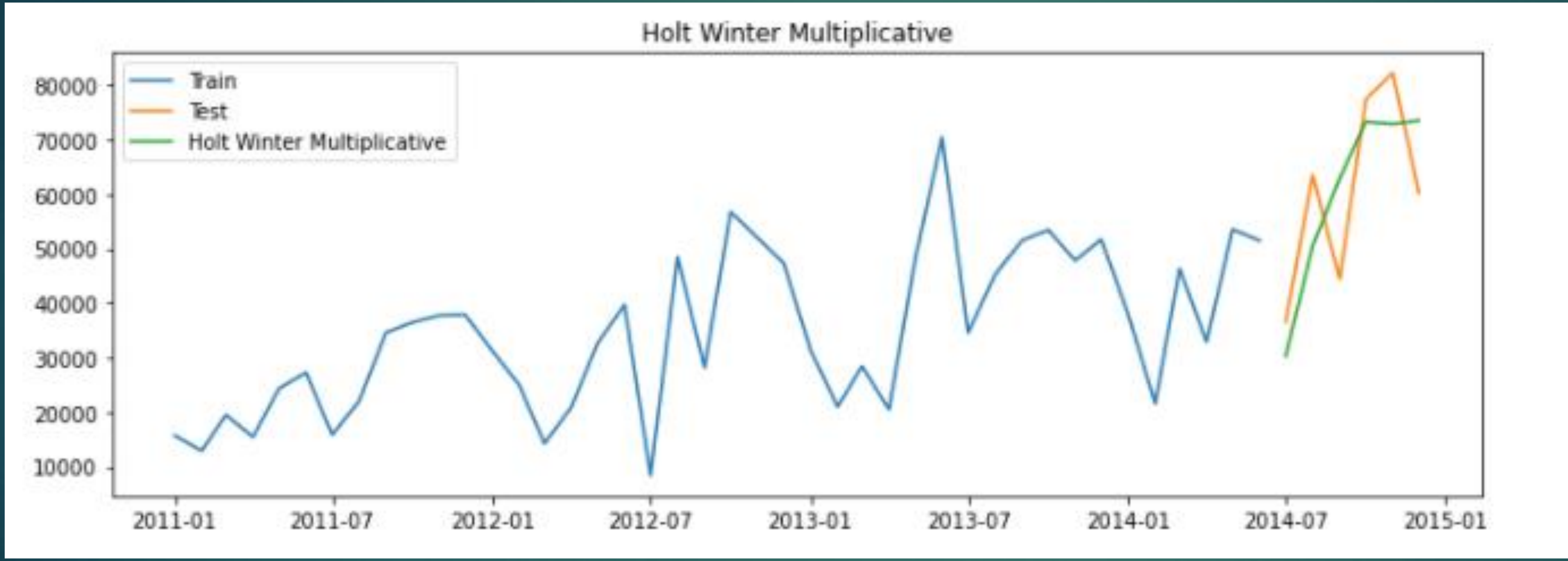
- This forecasts the lev el, trend as well as the seasonality for a time series data.
- There are two methods of performing the Holt-Winters' smoothing techniques: additiv e and multiplicativ e methods.



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57
0	Holt Winter Additive Method	12971.01	17.61

Smoothing Technique : Holt's Winter Multiplicative Smoothing

- This forecasts the lev el, trend as well as the seasonality for a time series data.
- There are two methods of performing the Holt-Winters' smoothing techniques: additiv e and multiplicativ e methods.



	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57
0	Holt Winter Additive Method	12971.01	17.61
0	Holt Winter Multiplicative Method	11753.42	19.62

Auto Regression Methods

- Forecast the future observations using a linear combination of past observations of the same variable. There are two fundamental assumptions to build an autoregressive model.
  - Stationarity - statistical properties like mean, variance, and covariance will be the same throughout the series
  - Autocorrelation - helps us to know how a variable is influenced by its own lagged values
    - Autocorrelation measures
      - 1. Autocorrelation function (ACF) 2. Partial autocorrelation function (PACF)
- Stationarity Test

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

● Null Hypothesis (H0): The series is stationary ○ p−value>0.05

● Alternate Hypothesis (H1): The series is not stationary ○ p−value≤0.05

Augmented Dickey-Fuller (ADF) Test

● Null Hypothesis (H0): The series is not stationary ○ p−value>0.05

● Alternate Hypothesis (H1): The series is stationary ○ p−value≤0.05

Stationarity Test

```
ADF Statistic: -3.967177
Critical Values @ 0.05: -2.94
p-value: 0.001592
```

```
KPSS Statistic: 0.502104
Critical Values @ 0.05: 0.46
p-value: 0.041193
```

Based on both tests, we can conclude ADF test indicates series is Stationary where as KPSS indicates series is Non-Stationary . After Box Cox Transformation & Differencing both test indicates series is Stationary.

```
ADF Statistic: -5.769275
Critical Values @ 0.05: -2.95
p-value: 0.000001
```

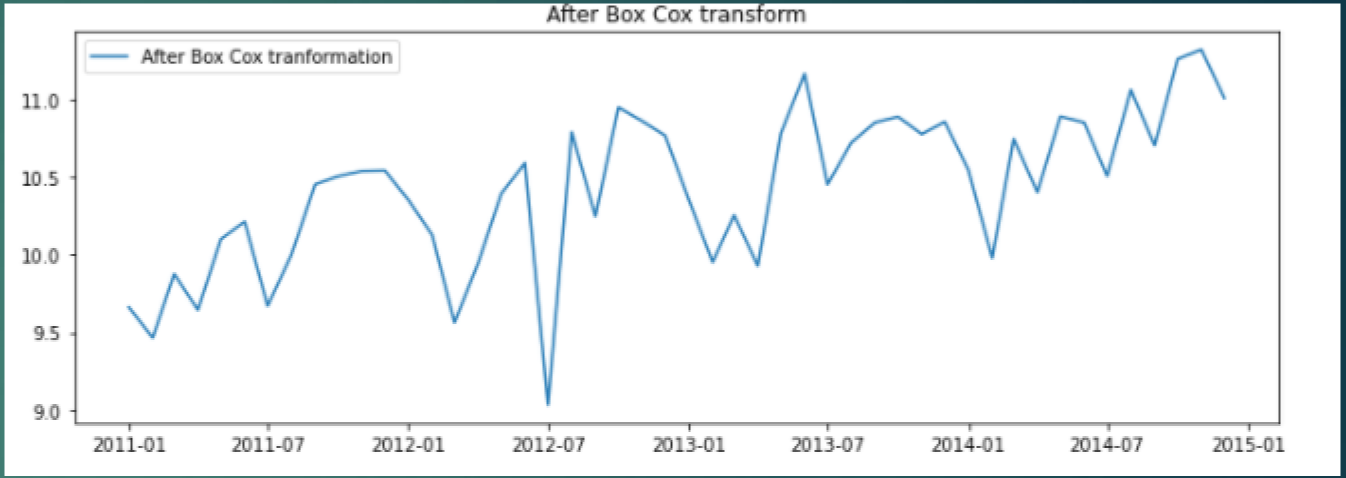
```
KPSS Statistic: 0.135659
Critical Values @ 0.05: 0.46
p-value: 0.100000
```

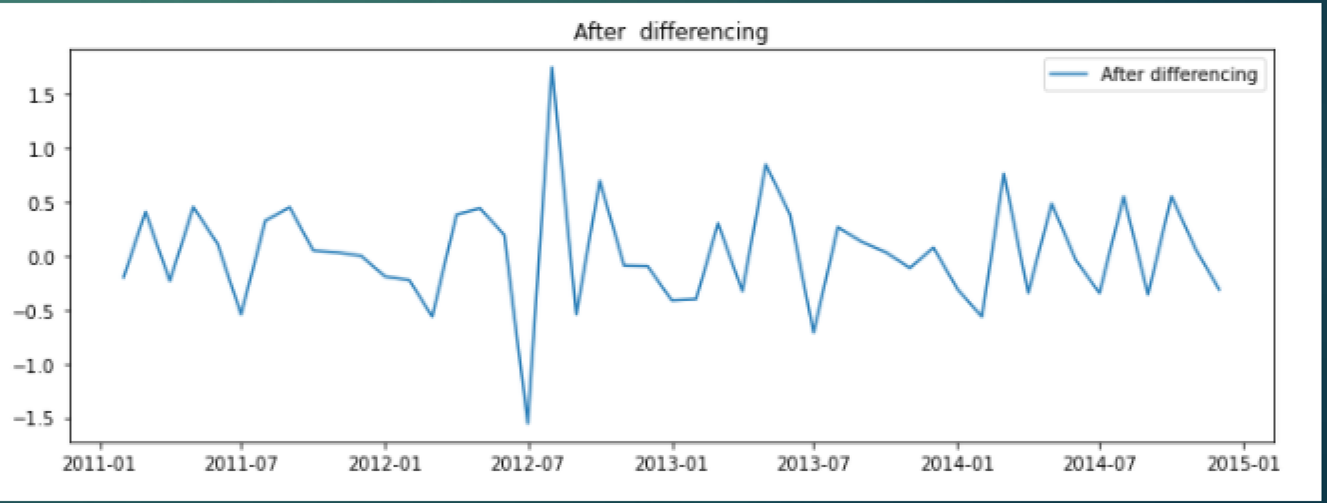
ADF p-value is lesser than .05 and KPSS p-value is greater than .05 , so we conclude series is Stationary

Box Cox Transformation & Differencing

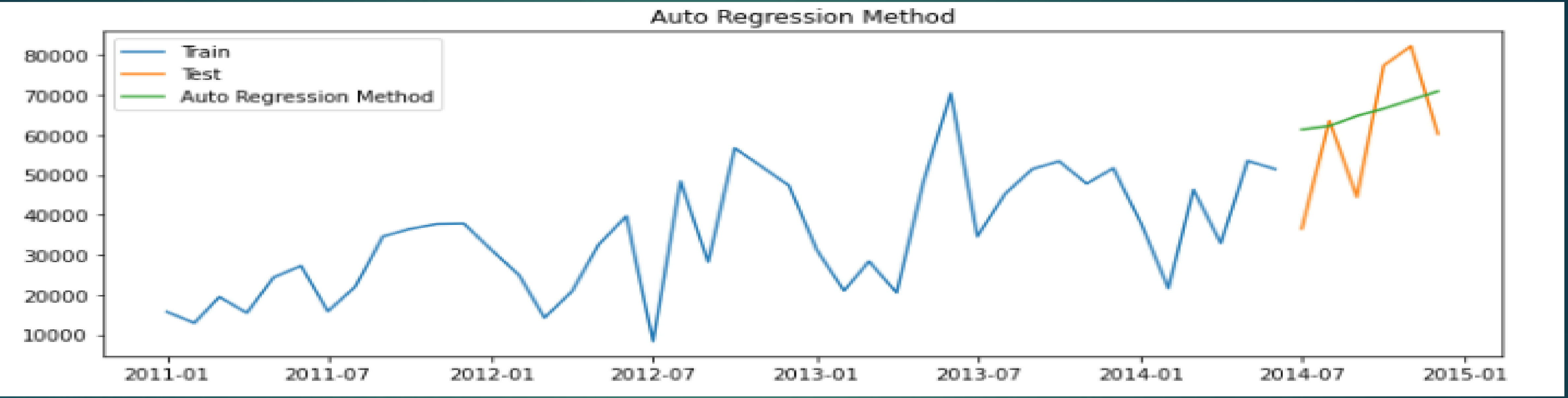
Box Cox Transformation – Makes the variance constant

Differencing – Remove the trend from the time series, differencing we compute the differences between consecutive observations.



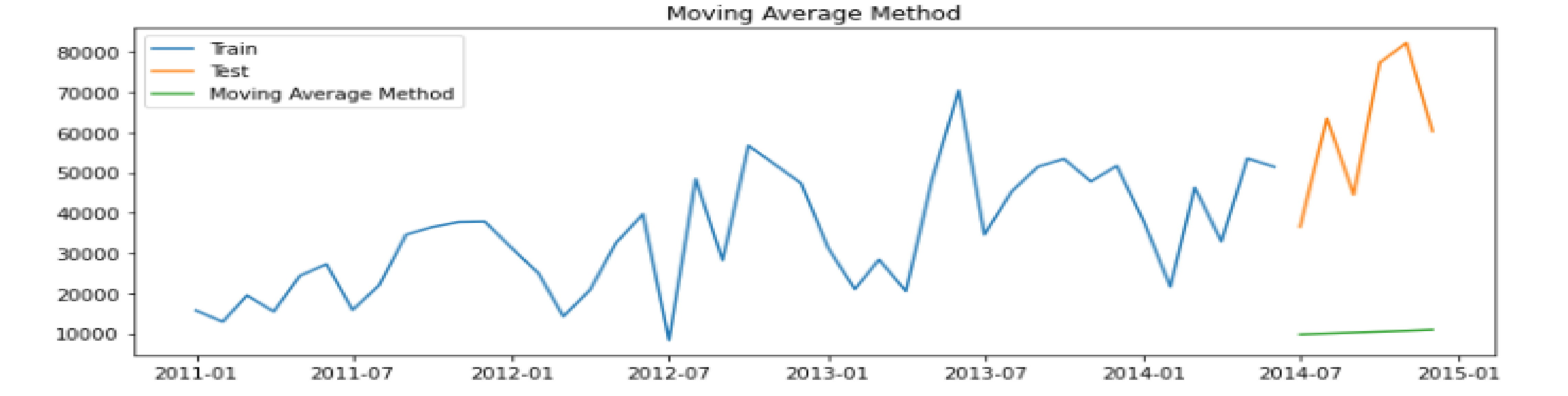


AR – Auto Regression



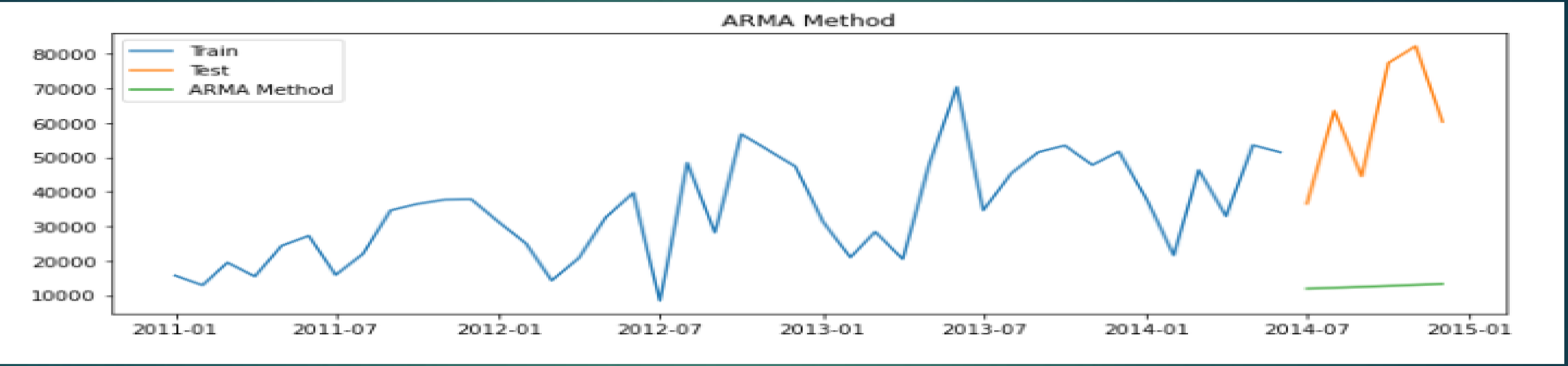
	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57
0	Holt Winter Additive Method	12971.01	17.61
0	Holt Winter Multiplicative Method	11753.42	19.62
0	Auto Regression Method	15505.02	27.27

MA – Moving Average



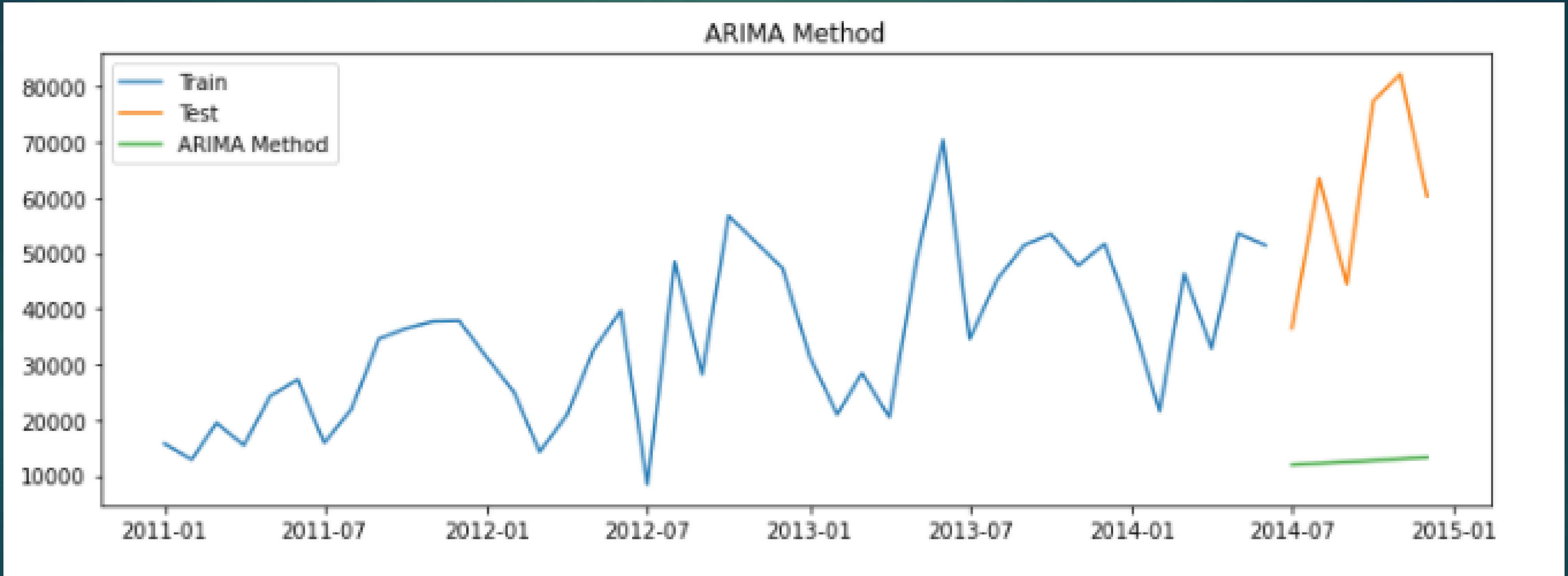
	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57
0	Holt Winter Additive Method	12971.01	17.61
0	Holt Winter Multiplicative Method	11753.42	19.62
0	Auto Regression Method	15505.02	27.27
0	Moving Average Method	52903.35	81.64

ARMA – Auto Regression Moving Average

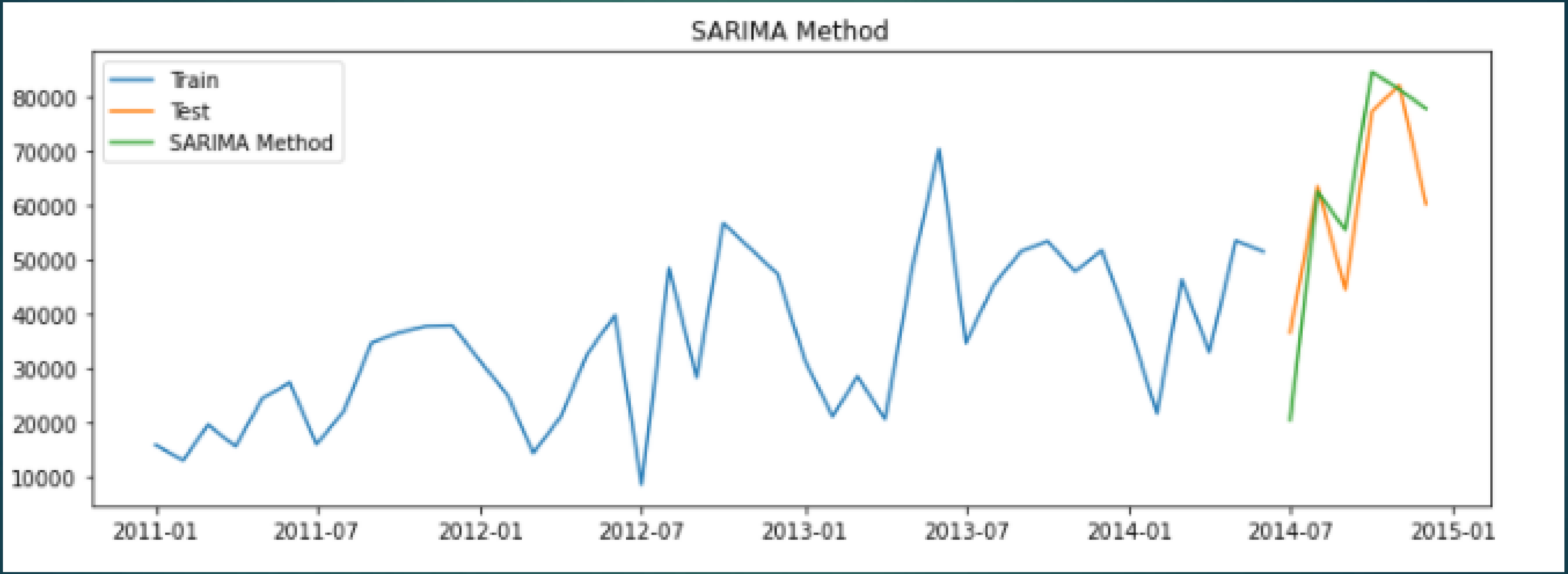


	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57
0	Holt Winter Additive Method	12971.01	17.61
0	Holt Winter Multiplicative Method	11753.42	19.62
0	Auto Regression Method	15505.02	27.27
0	Moving Average Method	52903.35	81.64
0	ARMA	50757.92	77.66

ARIMA -Auto Regression Integrated Moving Average



SARIMA –Seasonal Auto Regression Integrated Moving Average



Conclusion

As per all the below methods performed on Train data and forecast based on test data.

- Holt Winter's Additive method is best performing in smoothing techniques
- SARIMA is best performing in ARIMA techniques

	Method	RMSE	MAPE
0	Naive Method	18774.05	26.86
0	Simple Average Method	30846.00	38.18
0	Simple Moving Average Method	23383.65	28.15
0	Simple Exponential Smoothing Method	23112.44	27.82
0	Holt Exponential Smoothing Method	34412.52	42.57
0	Holt Winter Additive Method	12971.01	17.61
0	Holt Winter Multiplicative Method	11753.42	19.62
0	Auto Regression Method	15505.02	27.27
0	Moving Average Method	52903.35	81.64
0	ARMA	50757.92	77.66
0	ARIMA Method	50757.92	77.66
0	SARIMA Method	11180.27	18.38