Retail Giant Sales Forecasting Assignment

DEEPTI SISODIYA

Business

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.

Problem Statement

The store wants to forecast the sales of the products for the next 6 months, so that they have a proper estimate and can plan inventory and business processes accordingly.

Objective Analysis

- Understand the data
- Find the most profitable segments
- Forecast sales and demand for each of the profitable segments
- Recommendation for inventory management

Analysis Steps

1. Data Preparation

- Find the 21 unique Market Segments
- Convert the order-date into a date-time format
- Perform the train-test split
- Calculate Coefficient of variance(CoV) on the profit
- Find the most profitable market segment by comparing the 21 CoV values

2. Model Building

- Pick the most consistent profitable segment and drop the all other
- Perform the train-test split
- Decompose the time-series
- Calculate hypothesis by applying smoothing technique methods
- Build the different model using ARIMA set of techniques

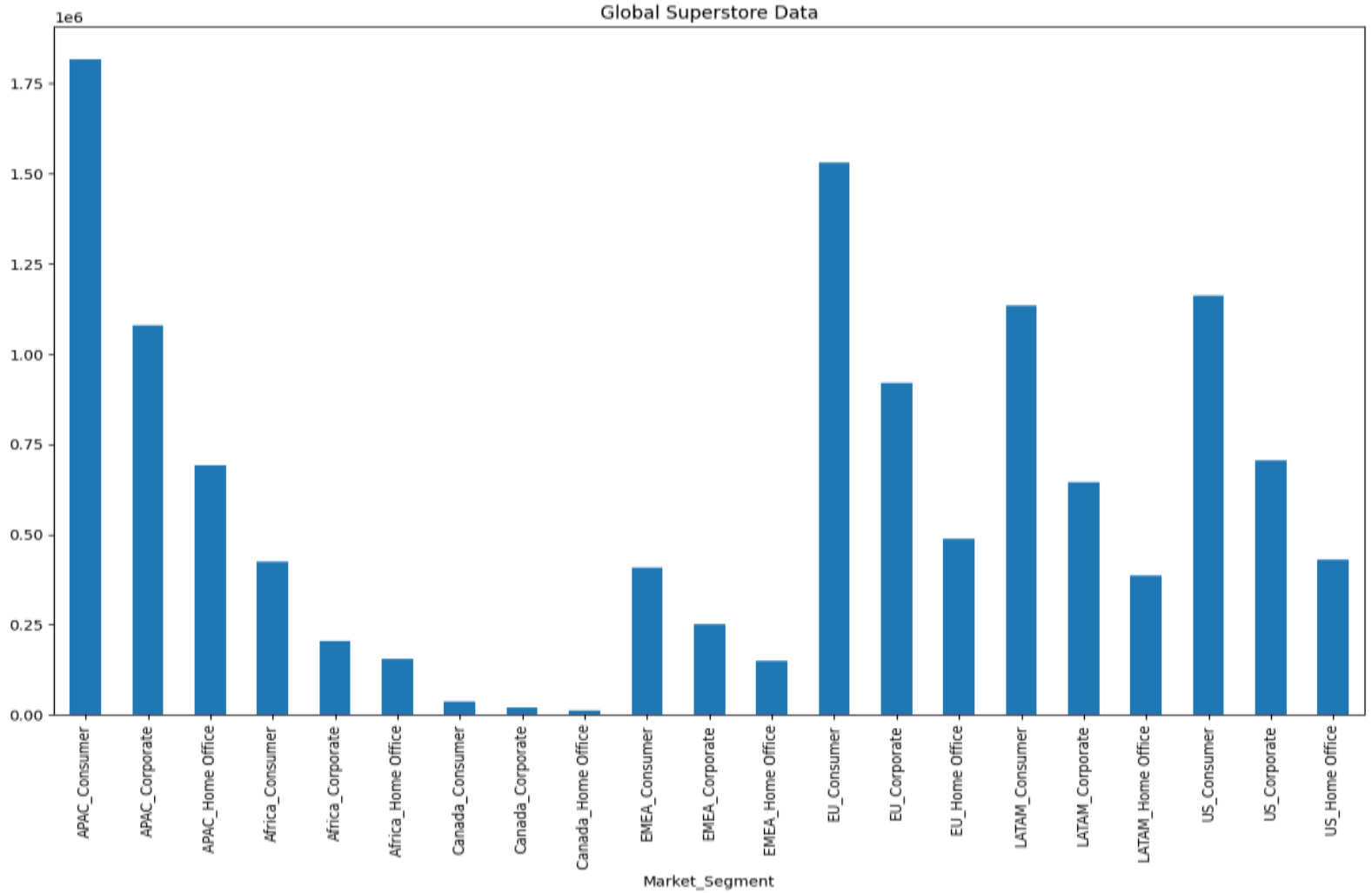
3. Model Evaluation

- Evaluate model on validation set using RMSE and MAPE values
- Choose the best model out of all the Models using MAPE value

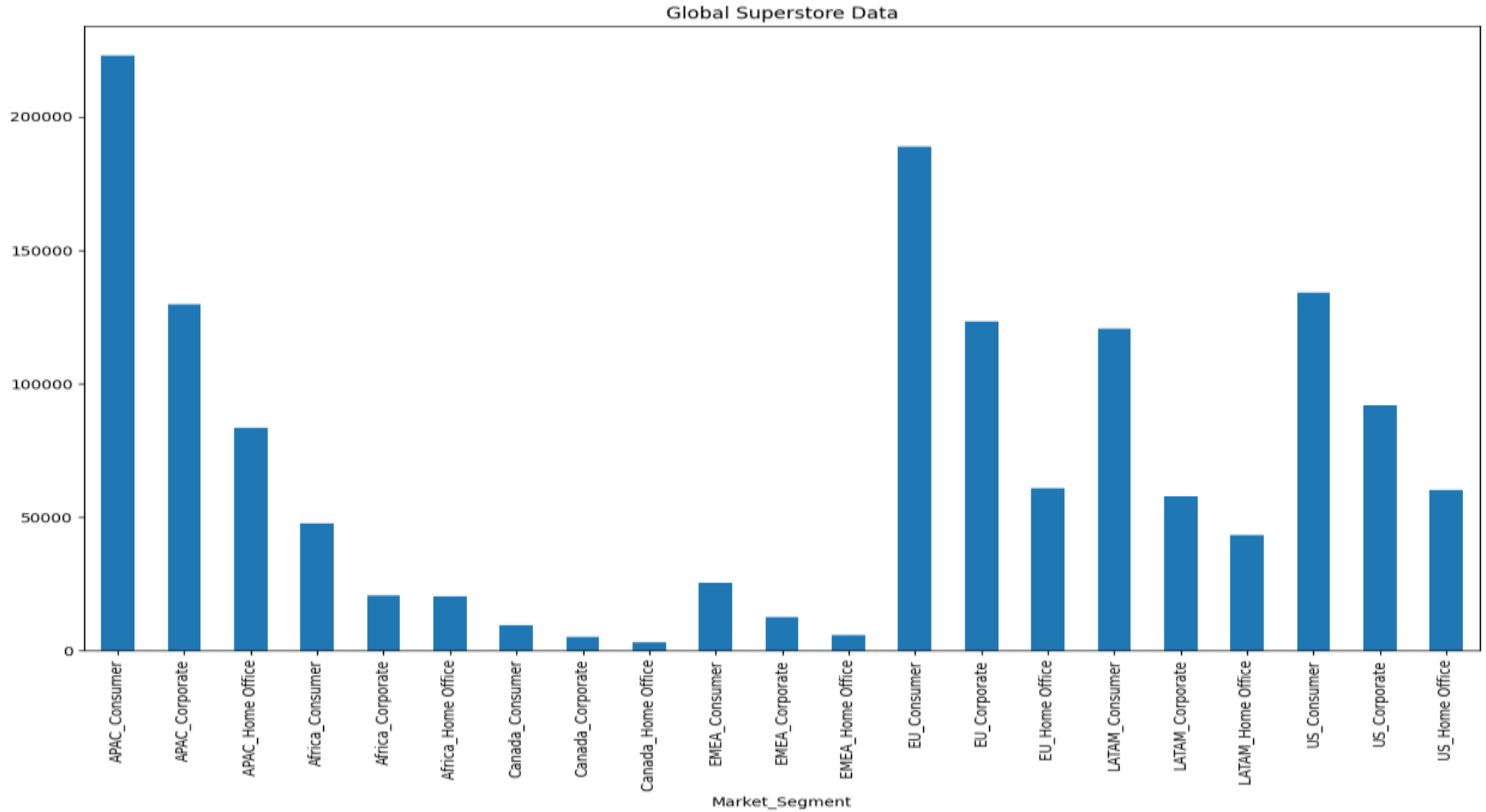
4. Forecasting

- Use the best model to forecast future 6-months sales for the most profitable market segment

Sales Analysis of Market-Segment



Profit Analysis of Market-Segment



Most Profitable Segment

Calculate the Coefficient of Variation using aggregated Profit for each Market_Segment using below formula:

CoV = std(Profit)/mean(Profit)

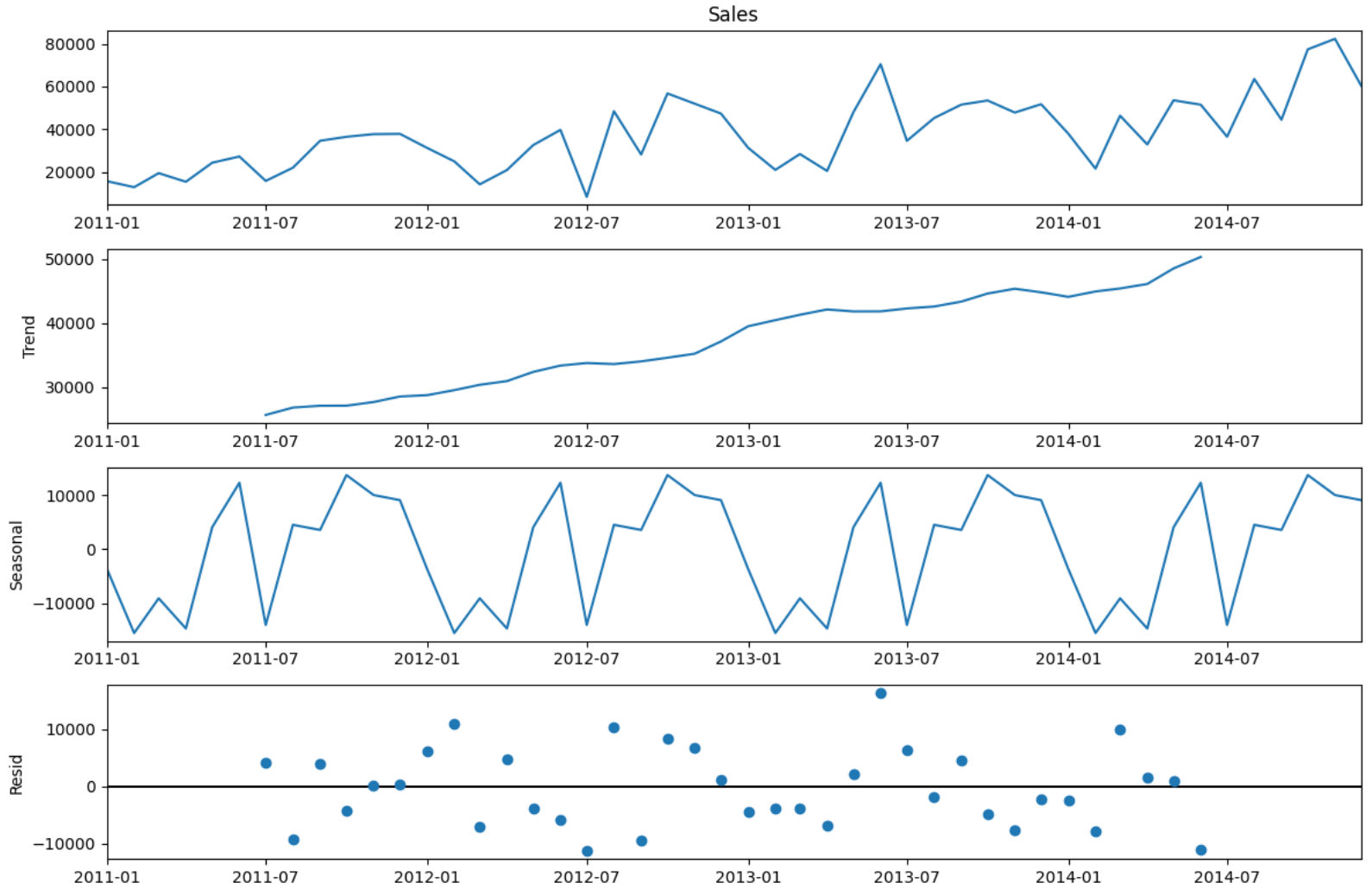
Using the CoV and Profit, found the most profitable Market_Segment as APAC_Consumer with **CoV value – 0.59**

	Market_Segment	CoV
0	APAC_Consumer	0.596404
12	EU_Consumer	0.647485
15	LATAM_Consumer	0.680684
13	EU_Corporate	0.689346
1	APAC_Corporate	0.731926
16	LATAM_Corporate	0.880260
19	US_Corporate	1.027209
2	APAC_Home Office	1.048817
18	US_Consumer	1.095295
14	EU_Home Office	1.114681
7	Canada_Corporate	1.197220
20	US_Home Office	1.217133
17	LATAM_Home Office	1.343696
3	Africa_Consumer	1.429335
6	Canada_Consumer	1.476093
4	Africa_Corporate	1.664827
5	Africa_Home Office	1.989866
8	Canada_Home Office	2.188300
9	EMEA_Consumer	2.716992
11	EMEA_Home Office	6.066684
10	EMEA_Corporate	6.779639

Time Series Decomposition - Additive

Decomposed the data using additive method:

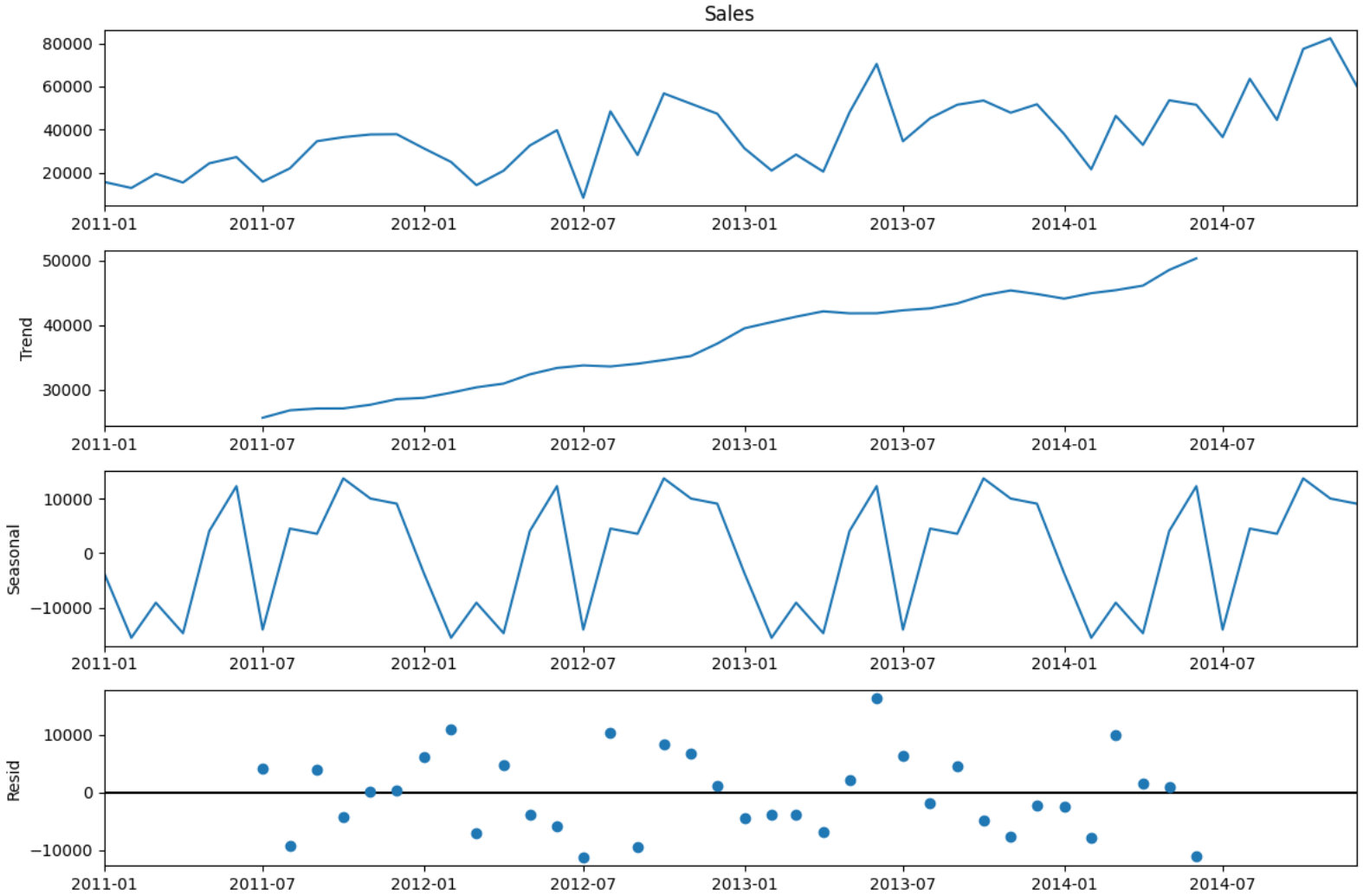
- There is clear upward trends
- There is a yearly seasonality in the data



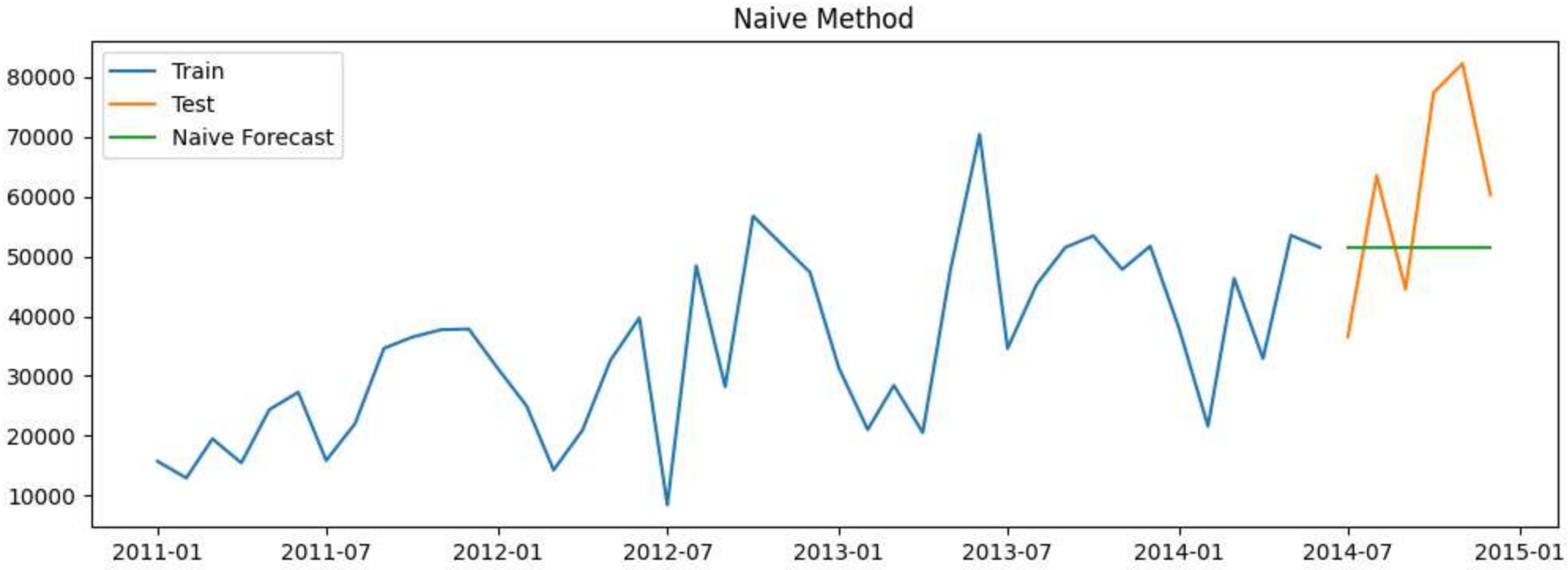
Time Series Decomposition - Multiplicative

Decomposed the data using multiplicative method:

- There is clear upward trends
- There is a yearly seasonality in the data

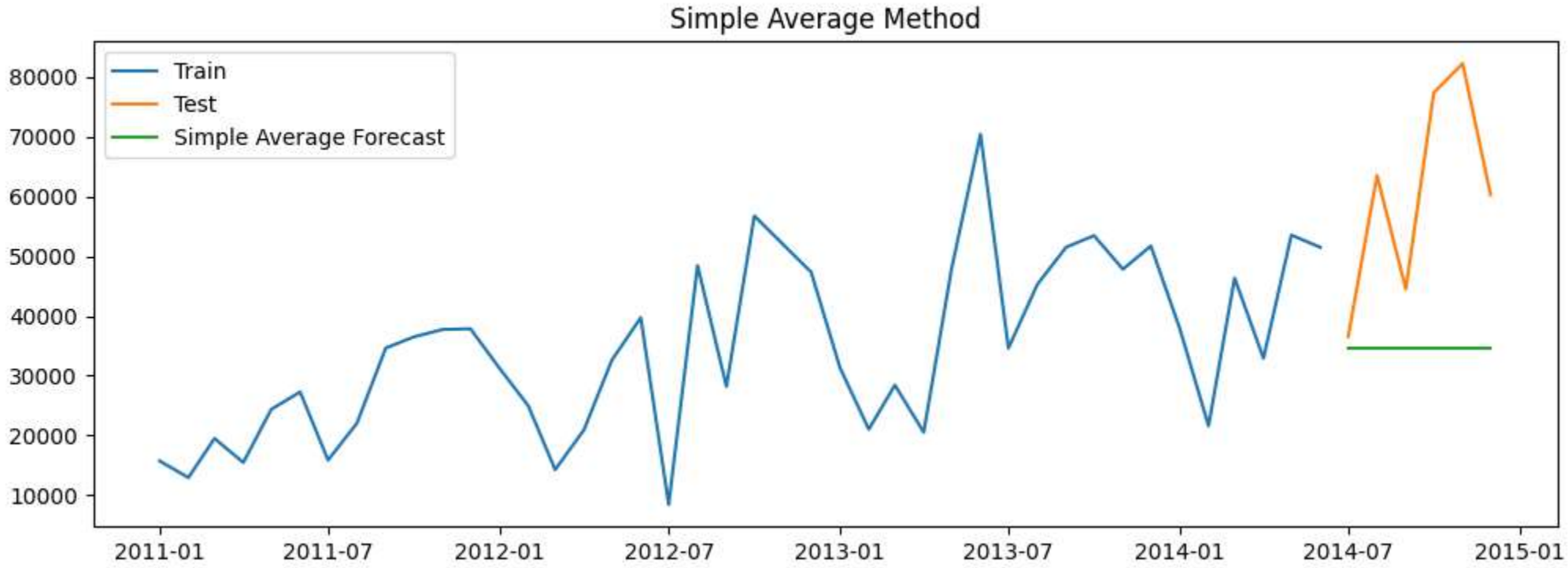


Time Series Model – Naïve Method



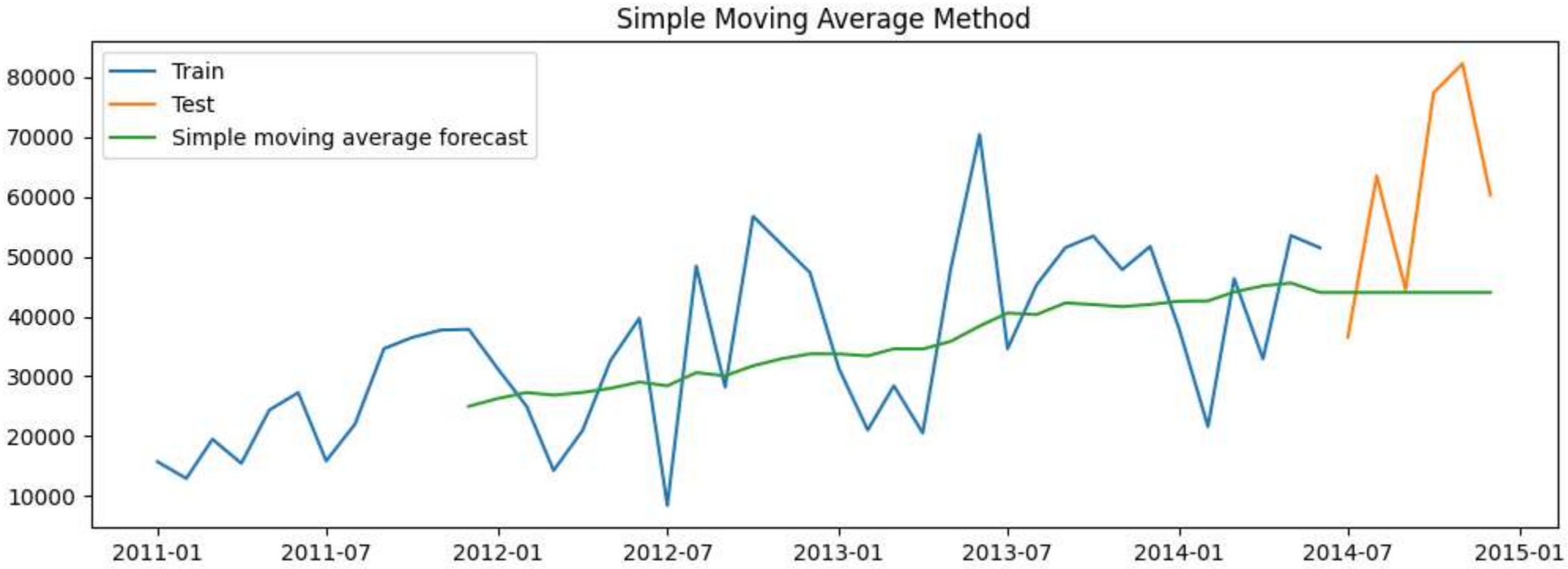
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86

Time Series Model – Simple Advantage Method



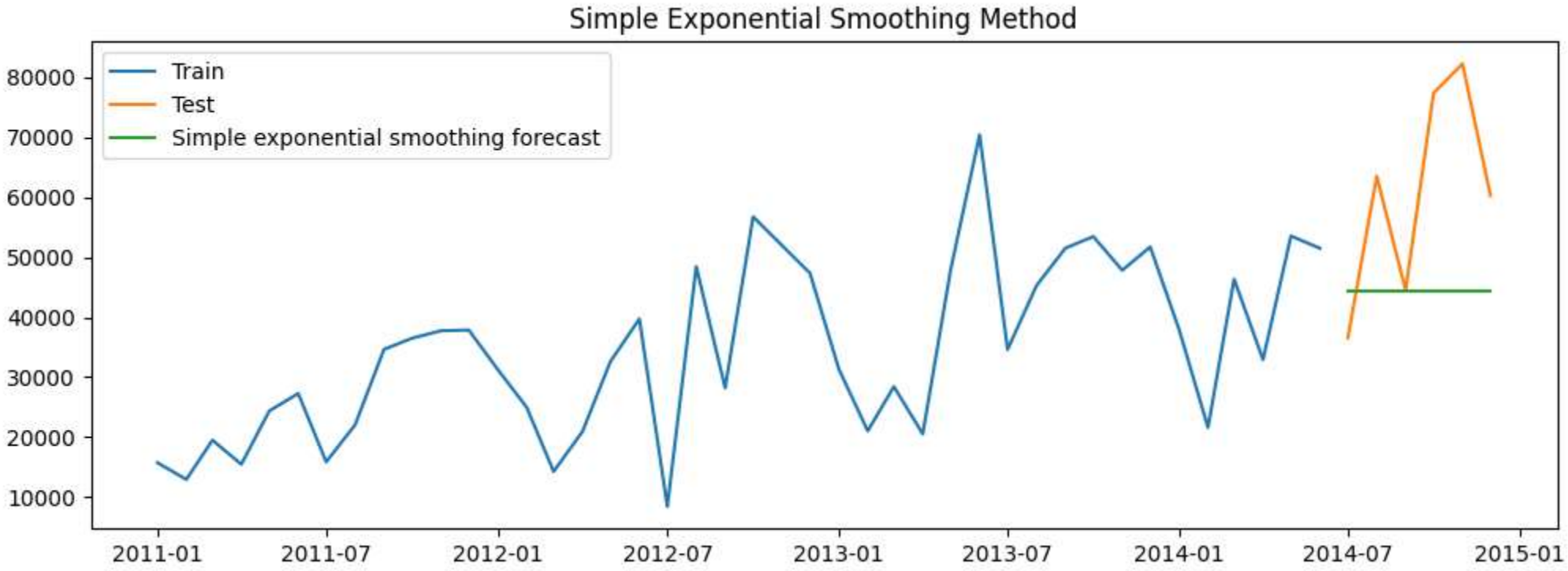
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18

Time Series Model - Simple Moving Average Method



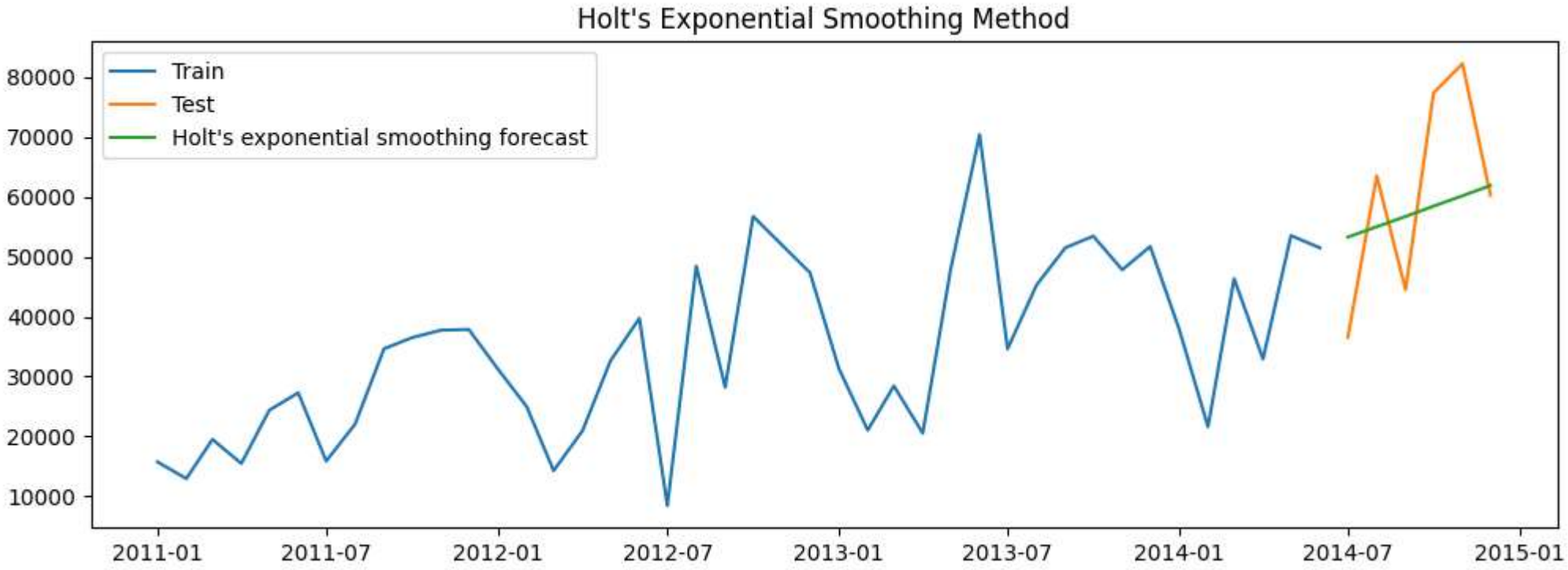
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15

Time Series Model - Simple Exponential Smoothing Method



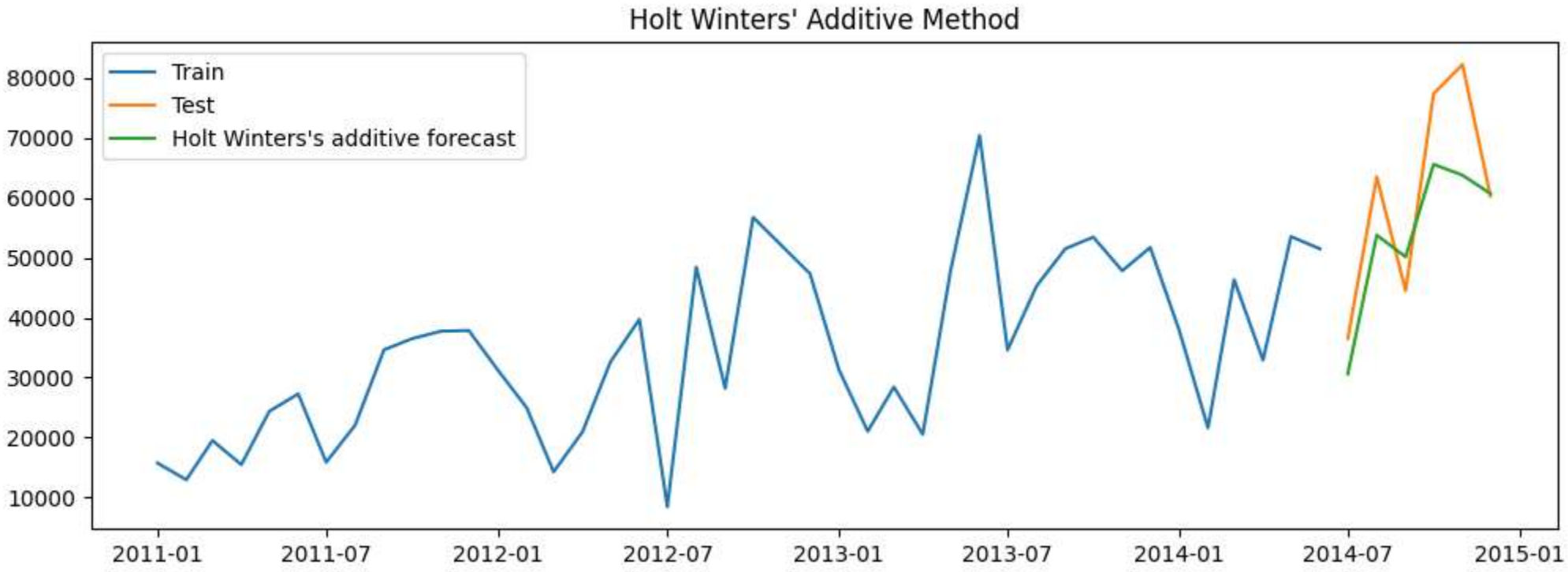
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82

Time Series Model – Holt's Exponential Smoothing Method



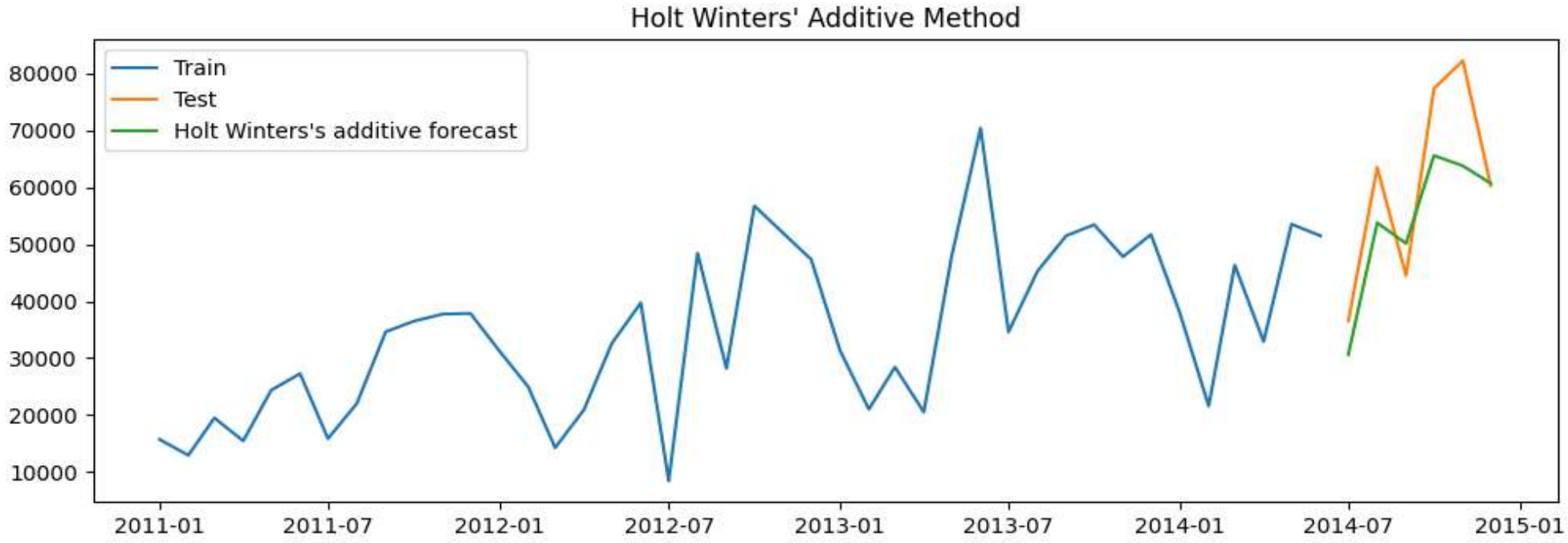
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47

Time Series Model – Holt Winters' Additive Method



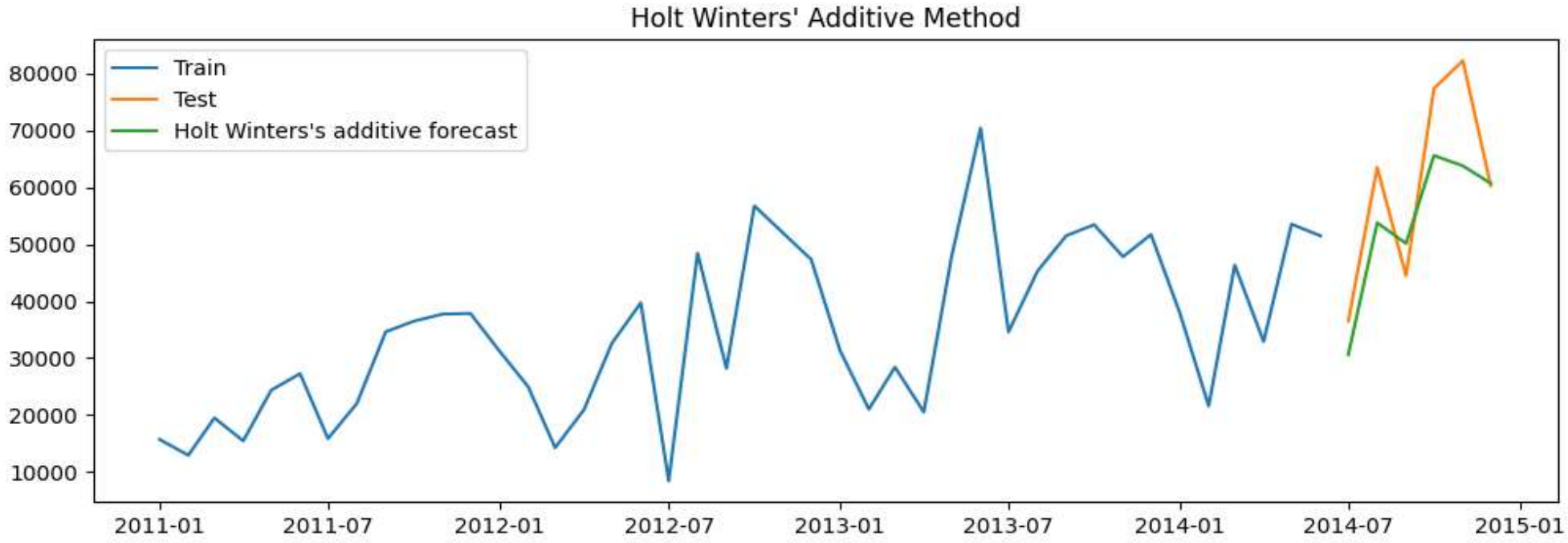
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77

Time Series Model – Holt Winters' Multiplicative Method



	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77

Time Series Model – Holt Winters' Multiplicative Method



	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69

Box Cox Transformation & Differencing

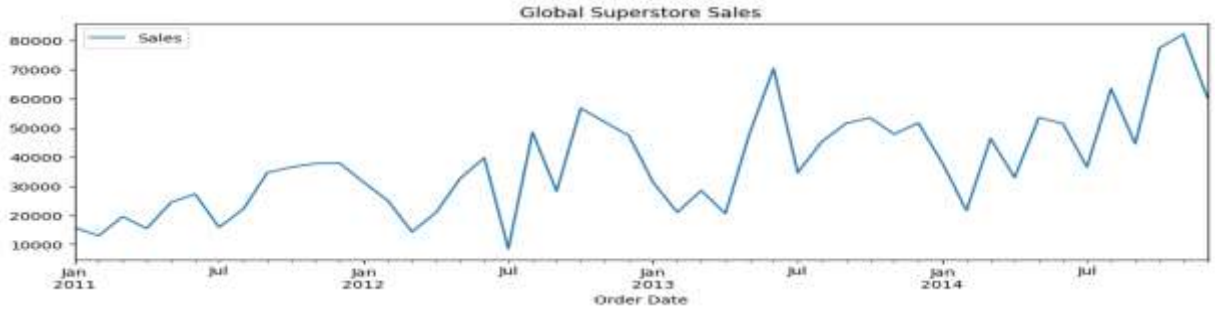
Using ADF and KPSS test we identified that time series data is not Stationary

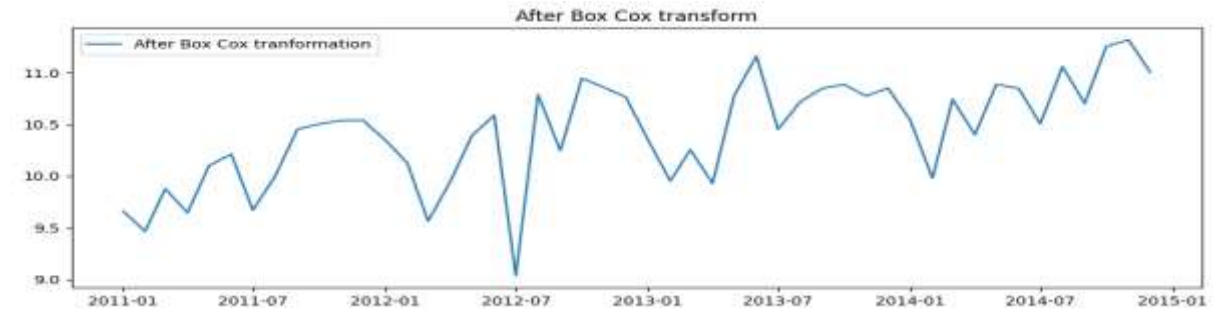
```
adf_test = adfuller(data_boxcox_diff)

print('ADF Statistic: %f' % adf_test[0])
print('Critical Values @ 0.05: %.2f' % adf_test[4]['5%'])
print('p-value: %f' % adf_test[1])
```

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

```
kpss_test = kpss(data_boxcox_diff)

print('KPSS Statistic: %f' % kpss_test[0])
print('Critical Values @ 0.05: %.2f' % kpss_test[3]['5%'])
print('p-value: %f' % kpss_test[1])
```

```
KPSS Statistic: 0.094966
Critical Values @ 0.05: 0.46
p-value: 0.100000
```



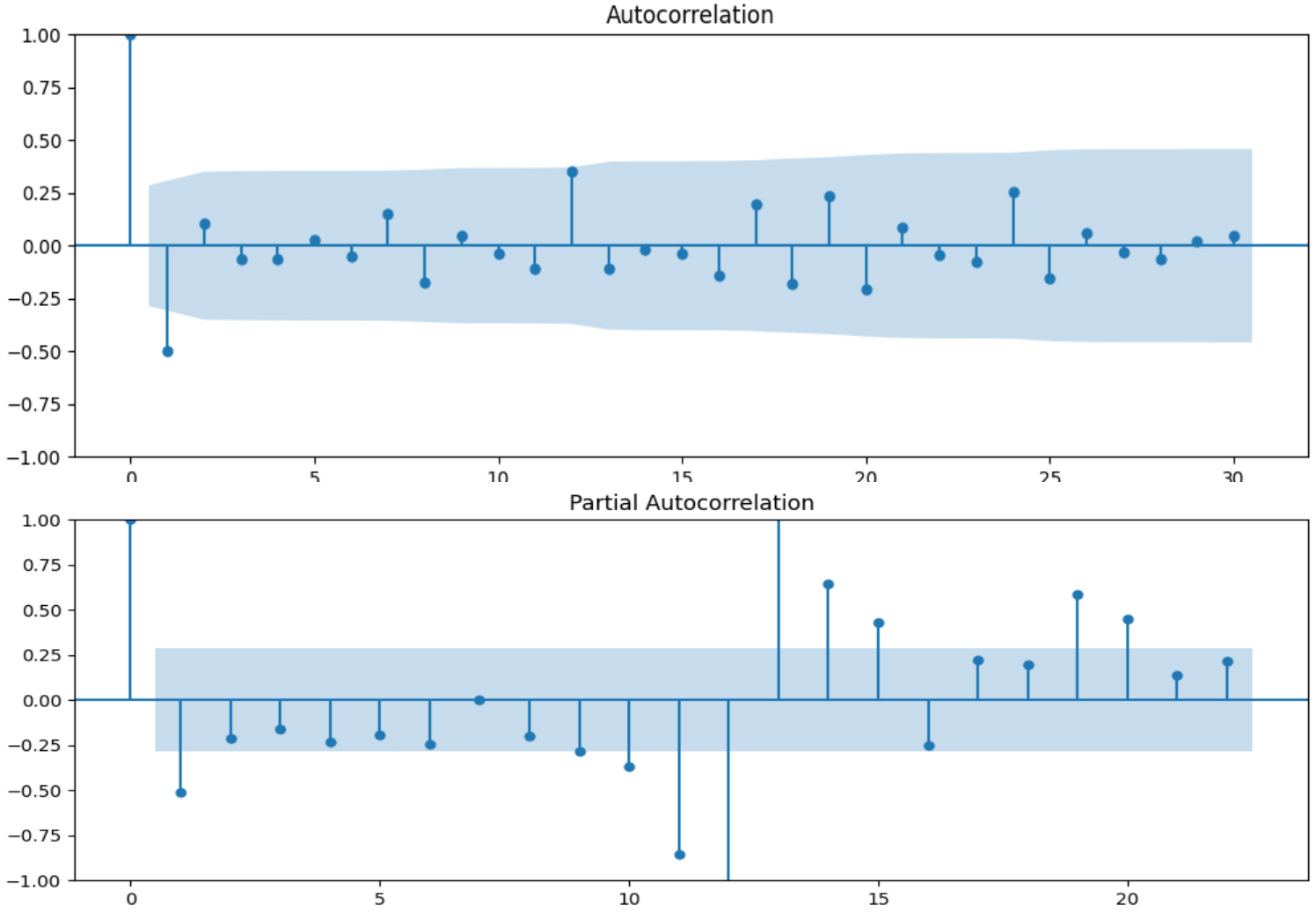


Since data was not stationary we need to do Box Cox and Differentiating to make it stationary

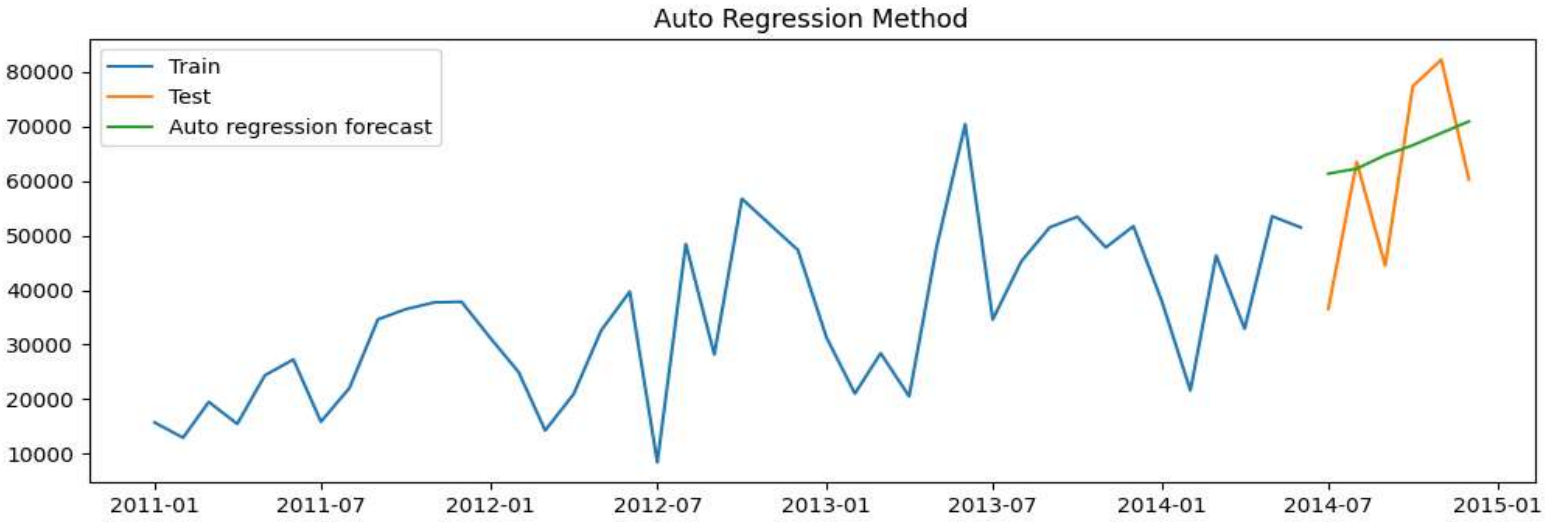
ACF and PACF Plot

From ACF plot we could see that the dependency on the very next node which means MA should be 1

From PACF plot we could see that there is a seasonality in the data

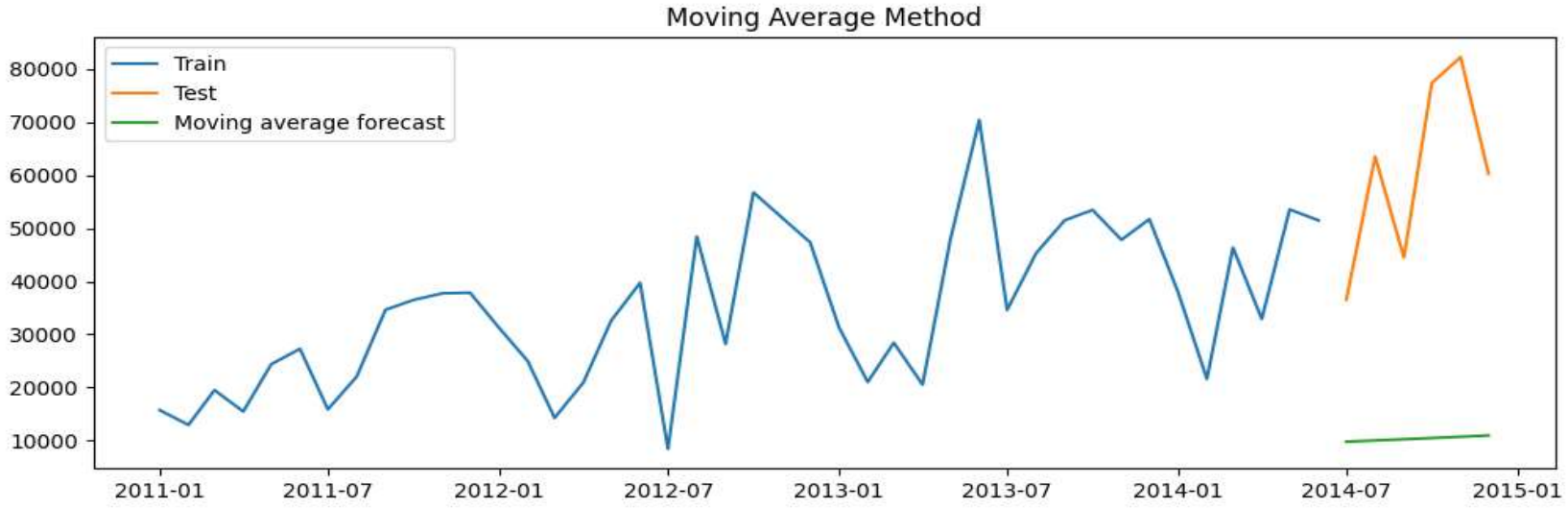


Time Series Model – Auto Regressive Method



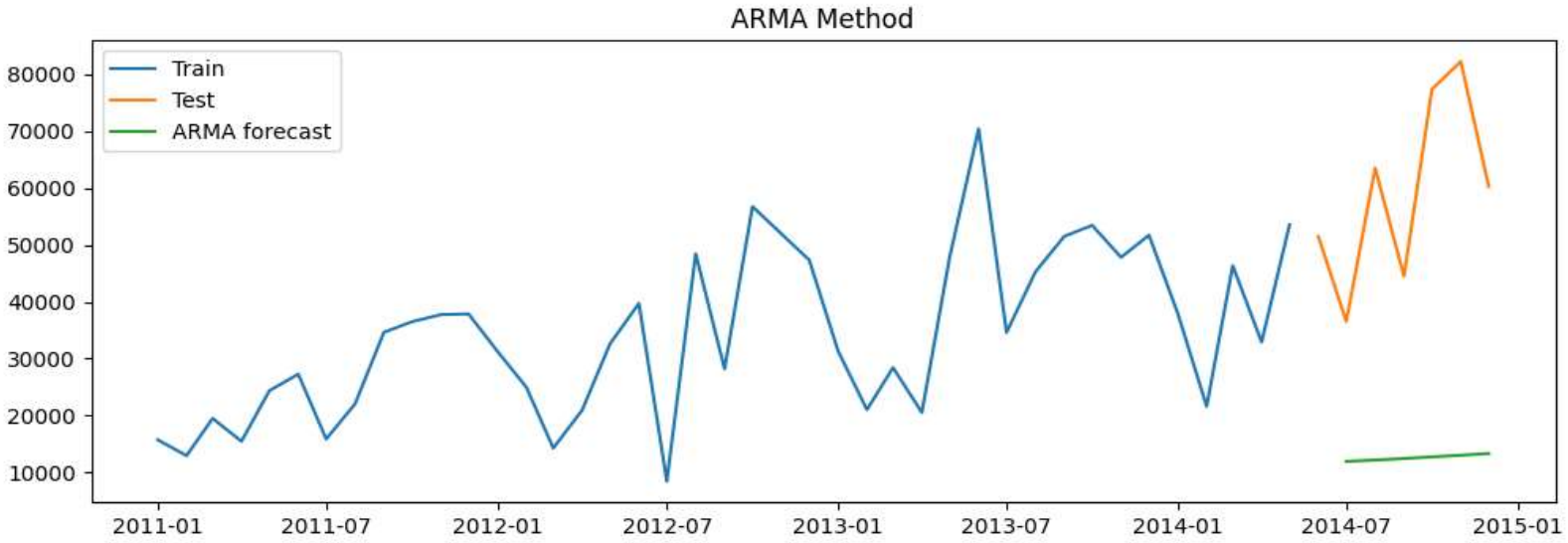
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69
0	Autoregressive (AR) method	15498.94	27.26

Time Series Model – Moving Average Method



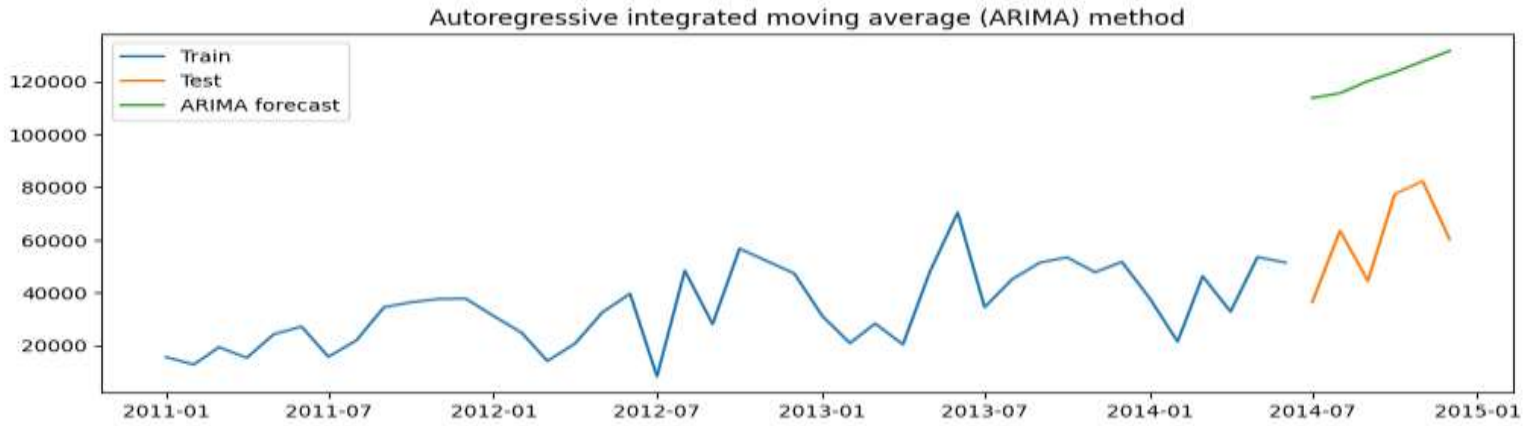
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69
0	Autoregressive (AR) method	15498.94	27.26
0	Moving Average (MA) method	52917.91	81.67

Time Series Model – ARMA Method



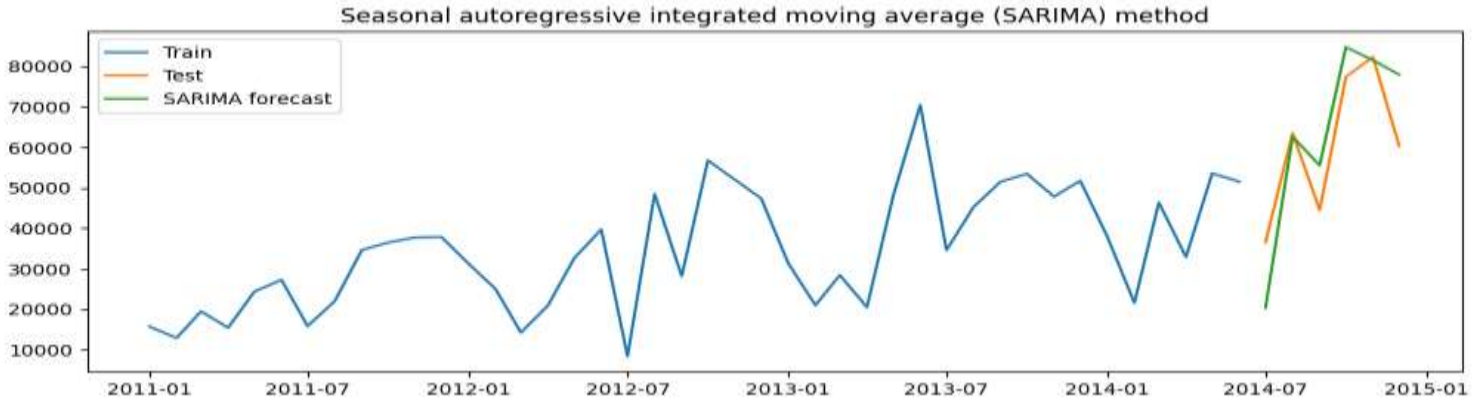
	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69
0	Autoregressive (AR) method	15498.94	27.26
0	Moving Average (MA) method	52917.91	81.67
0	Autoregressive moving average (ARMA) method	50775.82	77.69

Time Series Model – ARIMA Method



	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69
0	Autoregressive (AR) method	15498.94	27.26
0	Moving Average (MA) method	52917.91	81.67
0	Autoregressive moving average (ARMA) method	50775.82	77.69
0	Autoregressive integrated moving average (ARIM...	62769.36	116.02

Time Series Model – SARIMA Method



	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69
0	Autoregressive (AR) method	15498.94	27.26
0	Moving Average (MA) method	52917.91	81.67
0	Autoregressive moving average (ARMA) method	50775.82	77.69
0	Autoregressive integrated moving average (ARIM...	62769.36	116.02
0	Seasonal autoregressive integrated moving aver...	11179.56	18.38

Model Evaluation

The best performing time series model is **Holt Winter' Multiplicative Method** out of 12 time series models

	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	23112.44	27.82
0	Holt's exponential smoothing method	15014.67	23.47
0	Holt Winters' additive method	10350.33	13.77
0	Holt Winters' multiplicative method	9585.23	16.69
0	Autoregressive (AR) method	15498.94	27.26
0	Moving Average (MA) method	52917.91	81.67
0	Autoregressive moving average (ARMA) method	50775.82	77.69
0	Autoregressive integrated moving average (ARIM...	62769.36	116.02
0	Seasonal autoregressive integrated moving aver...	11179.56	18.38

Summary

- Based on data provided we helped Global Superstore in identified most profitable market segment as APAC Consumer
- We created total 12-Forecasting models for APAC Consumer market segment
- Selected the best performing time series model as Hold Winters' Multiplicative Method
- SARIMA – Seasonal Autoagressive Integrated Moving Average is the best method in ARIMA set of techniques
- Generated forecasting for the future 6-months sales
- Below is the summary of key forecasts on the test data:

❖ APAC Consumer Sales is likely to rise in next 6-months with small fluctuation

❖ APAC Consumer is also likely to rise steeply for upcoming 6-months

Thank you