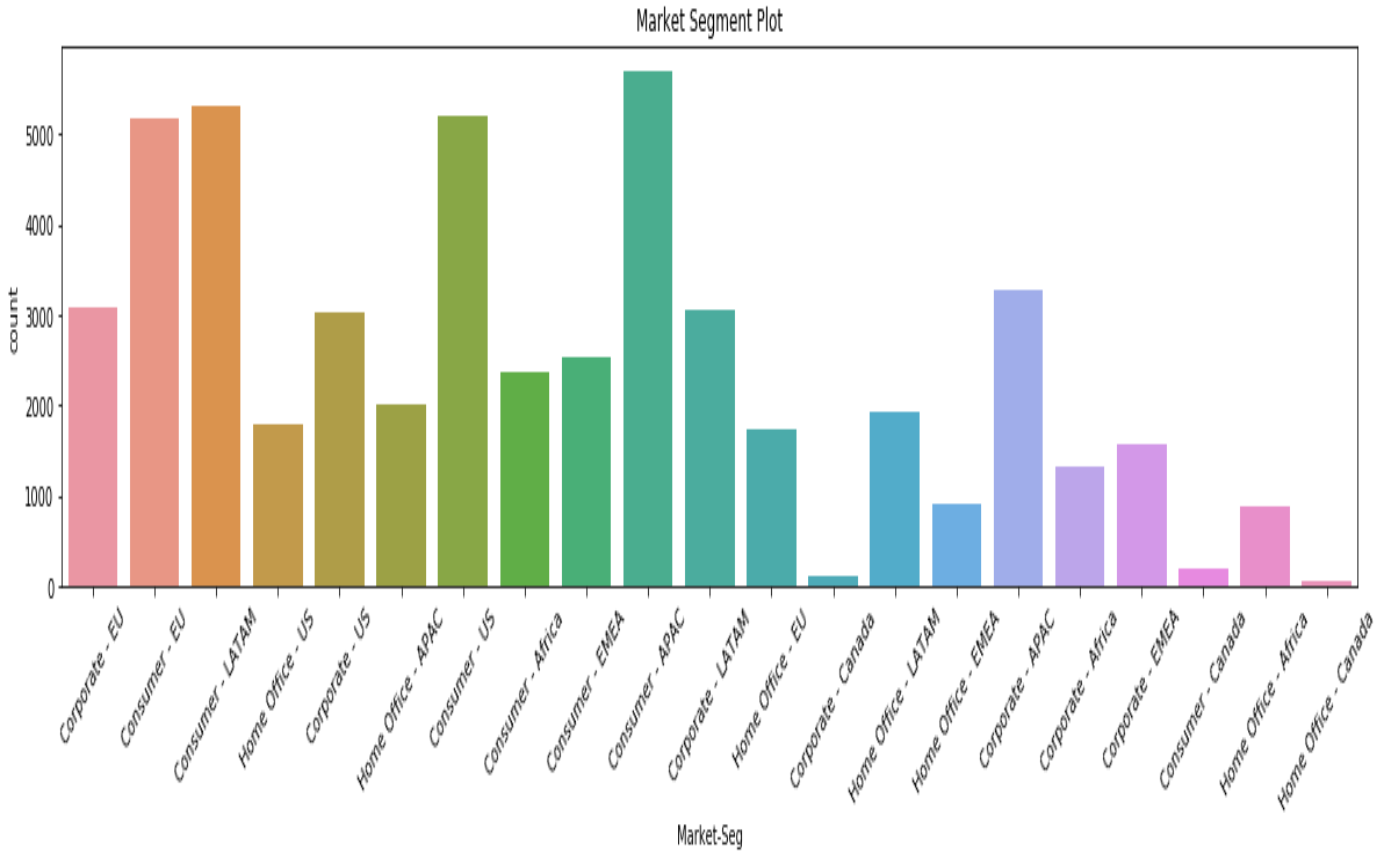
Retail Giant Sales Forecasting Assignment

Shyamala Rajasekar

Overview

- Global Mart is an online supergiant that caters to 7 Markets and 3 segments.
- Markets are namely
  - US (United States)
  - APAC (Asia Pacific)
  - EU (European Union)
  - Africa
  - Canada
  - EMEA(Middle East)
  - LATAM (Latin America)
- Segments are Consumer, Home-office and Corporate.
- Data is provided for sales in the period of 2011 to 2014 comprising of 48 months. It has 5 attributes: Order Date, Sales, Profit, Market and segment.



- Global Mart caters to 21 Market-Segments-Formed by combining the Markets representing the geography and Segments representing the Customer type.
- **Consumer-APAC** has the highest number of entries in the dataframe indicating that it has the maximum no. of Sales(not Sales in Total)

	Market-Seg
Consumer - APAC	11.111328
Consumer - LATAM	10.374342
Consumer - US	10.120881
Consumer - EU	10.111133
Corporate - APAC	6.400858
Corporate - EU	5.999220
Corporate - LATAM	5.952427
Corporate - US	5.888087
Consumer - EMEA	4.948333
Consumer - Africa	4.642230
Home Office - APAC	3.938390
Home Office - LATAM	3.743420
Home Office - US	3.476311
Home Office - EU	3.386625
Corporate - EMEA	3.068824
Corporate - Africa	2.558004
Home Office - EMEA	1.787873
Home Office - Africa	1.743030
Consumer - Canada	0.393839
Corporate - Canada	0.214467
Home Office - Canada	0.140378

Objective

- The objective is to forecast the sales of the products for the next 6 months to have a proper estimate and to plan inventory and business processes accordingly.
- However, due to certain unpredictable circumstances in the market, the company is prioritizing only the best and most consistent market segment in terms of profitability.
- The sales for the most consistently profitable market-segment is to be forecasted. Investing in this market segment will be beneficial for the company as the forecasts will be reliable.

- **Process Flow:**
    1. Finding the most consistently profitable market-Segment
    2. Forecasting the sales for that segment for the next 6 months.

Process Flow

1. **Finding the Most profitably consistent market-segment**
   - Upon combining the market and segment columns, a dataset with order-date, sales and profit for each market-segment.
   - The order date comprises of date, month and year of order. Only the month and year of order date was extracted resulting in 48 months of data
   - **Coefficient of Variation** for all the 21 market segments in train set(42 Months) was calculated to find the most consistently profitable market-segment.
     - CoV calculates the variation in data.

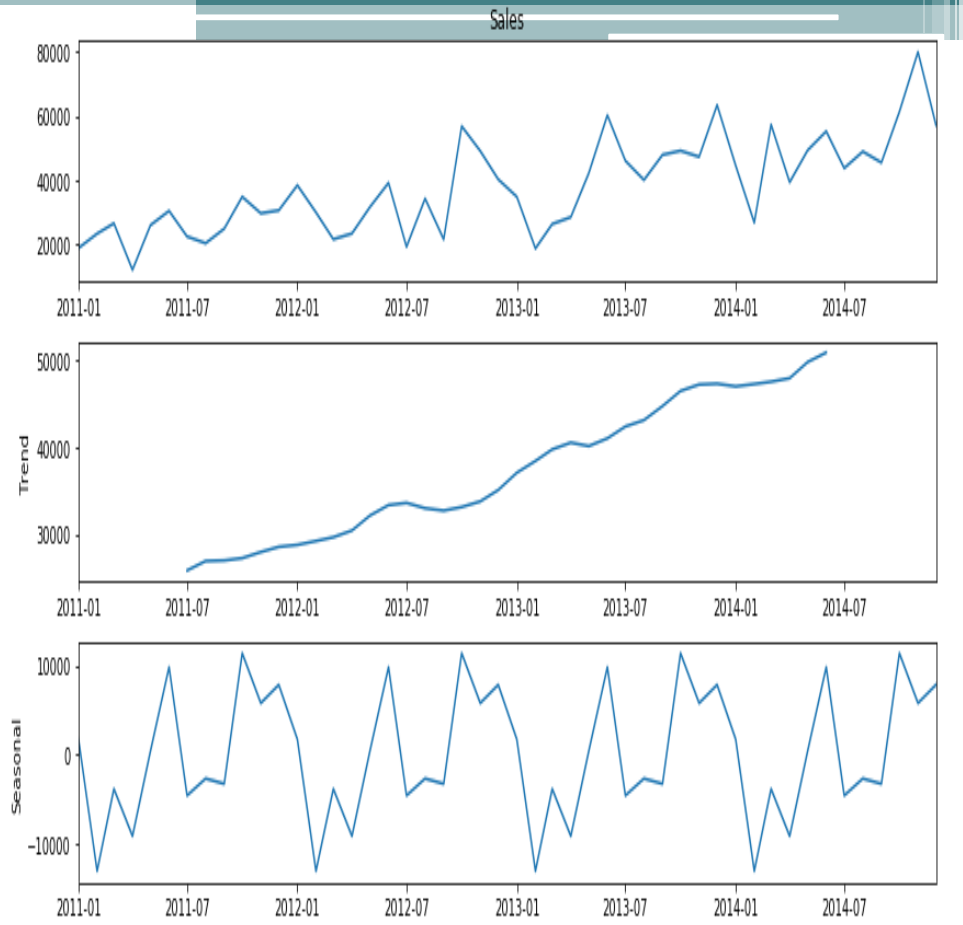
       CoV = Standard Deviation/Mean
     - Data with low CoV is more consistent and has a narrower spread than the data with high CoV.
     - By this principle, the most consistent market-segment is the dataset with the least CoV value.

	Market segment	Most Consistent
0	Consumer - APAC	0.522725
7	Corporate - APAC	0.530051
4	Consumer - EU	0.595215
5	Consumer - LATAM	0.683770
11	Corporate - EU	0.722076
12	Corporate - LATAM	0.882177
18	Home Office - EU	0.938072
14	Home Office - APAC	1.008219
6	Consumer - US	1.010530
13	Corporate - US	1.071829
20	Home Office - US	1.124030
19	Home Office - LATAM	1.169693
2	Consumer - Canada	1.250315
1	Consumer - Africa	1.310351
9	Corporate - Canada	1.786025
8	Corporate - Africa	1.891744
15	Home Office - Africa	2.012937
16	Home Office - Canada	2.369695
3	Consumer - EMEA	2.652495
10	Corporate - EMEA	6.355024
17	Home Office - EMEA	7.732073

The Most Consistently profitable Market-Segment is APAC(Asia Pacific)-Consumer.

2. Forecasting Sales

- Dataset with 21 market-segments is filtered for APAC-Consumer segment. As only sales is to be forecasted, only sales and order date is extracted from the dataset.
- Now the data contains 48 months of sales data for the APAC-Consumer segment.
- The sales of the 6 months is forecasted for this market-segment using smoothing and Auto Regression techniques.
- Sales of the most profitable month is plotted to check for trend, seasons and residuals. Same is done using decomposition process. Seasonality is observed upon decomposition along with an upward trend.



Decomposition of sales plot

Forecasting techniques used:

- **Smoothing techniques used are** :
  - Naive method
  - Simple average method
  - Simple moving average method
  - Simple exponential method
  - Holt's exponential method
  - Holt Winter's additive exponential method
  - Holt Winter's multiplicative exponential method

- **Auto Regression Techniques used are:**
  - Simple auto regressive method (AR)
  - Moving average method (MA)
  - Auto regressive and Moving average method (ARMA)
  - Auto regressive integrated moving average method (ARIMA)
  - Seasonal auto regressive integrated and moving average method (SARIMA)

Train , test and forecasted values from each technique was plotted and MAPE values were calculated to find the most suitable forecasting technique for this dataset.

Sales Data

Data is not noisy.

Time series has 48 Observations

Series has a trend and Seasonal nature

***The best forecast methods are indicated in purple*

Auto Regression

Auto-regression & Moving Average

This method doesn't predict trend and seasonality

ARMA and ARIMA

This method predicts Trend but not seasonality

Not the right fit

SARIMA

SARIMA method predicts trend and seasonality.

Hence the right fit.

Smoothing Techniques

Naïve Method

Series has more than 10 observations

Series has seasonality and trend nature

This method is not effective

Simple Average Method

Data is not noisy

It has a trend and seasonal nature

This method is not effective

Simple Exp Method

Data has a trend and seasonality

This method captures the level but not trend and seasonality.

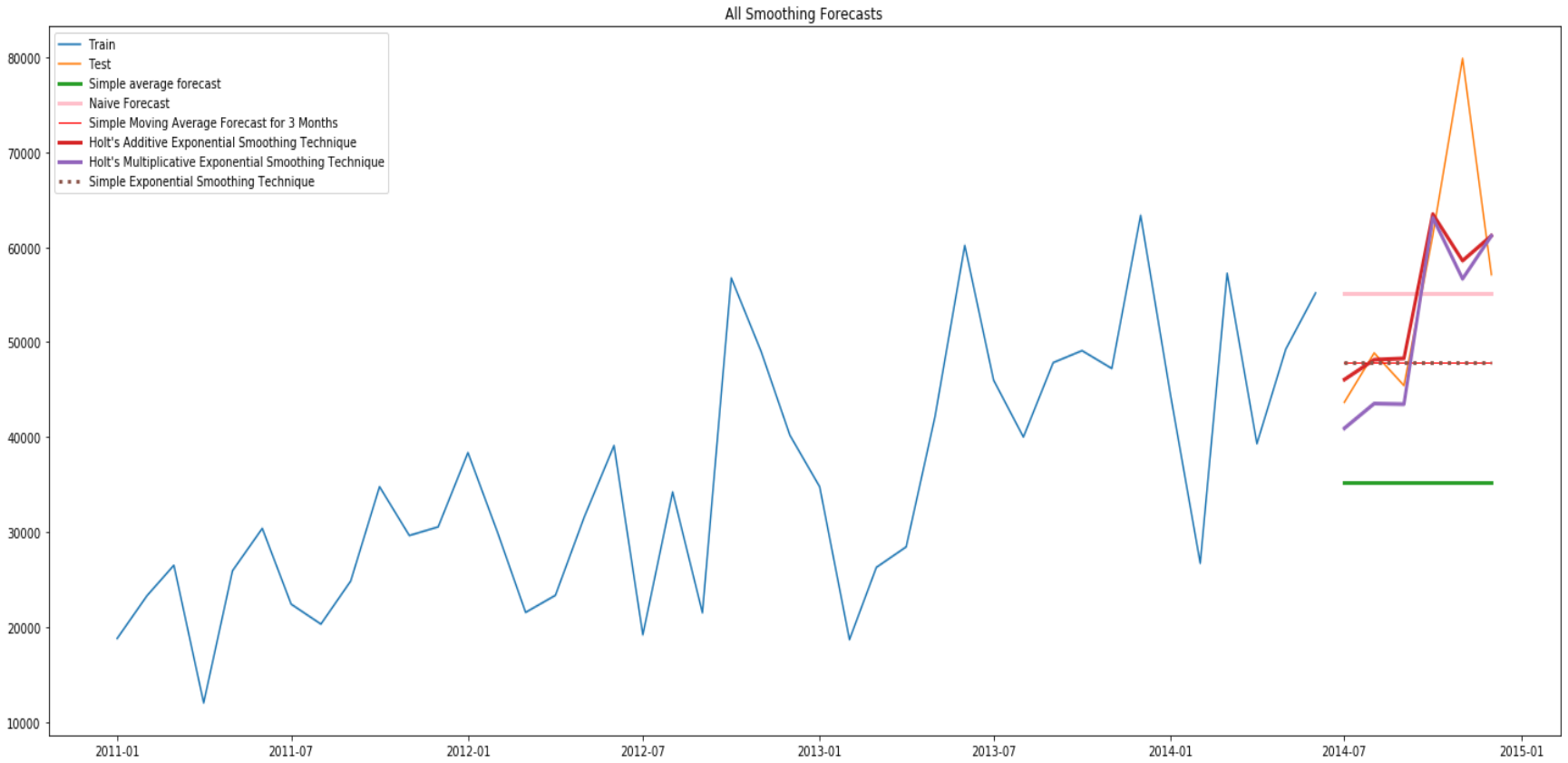
Holt's Exp Smoothing

This method predicts level and trend but not seasonality. It is not effective.

Holt's Winters Exp Smoothing

Predicts both trend and seasonality. This is the best method to forecast values.

FORECAST PLOTS OF ALL SMOOTHING TECHNIQUES



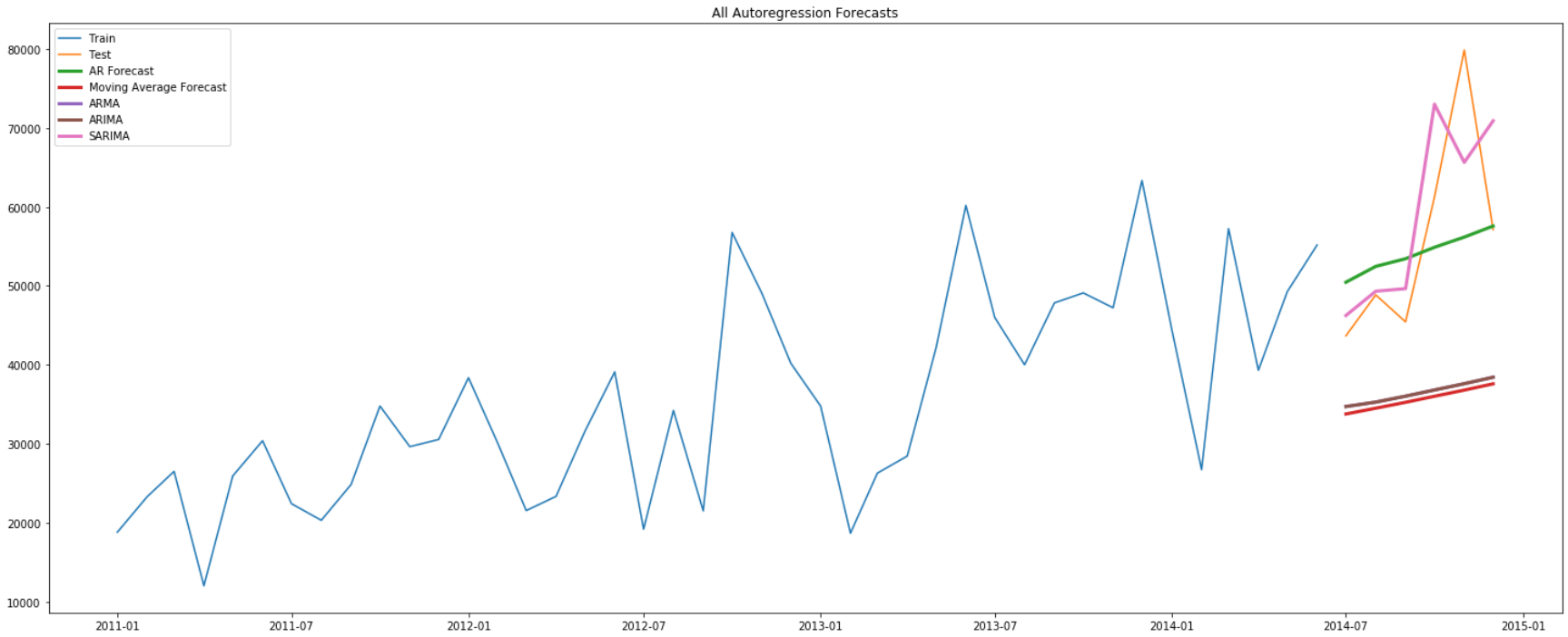
As is evident from the plot, Holt Winters Exponential addition method captures seasonality and trend better than other techniques.
Moving average technique performs the poorest among the smoothing techniques.

Observations

- **Holt Winter's Exponential Additive method** performs the best as it predicts seasonality and trend accurately
- **Holt Winters Exponential Multiplicative method** performance is poor as compared to Additive method because the seasonality is not a function of level component.
- **Simple Moving average method** model forecasts are the least accurate followed by Naïve Method.
- **Holt's Exponential method** captures trend but fails to capture seasonality.

Method	RMSE	MAPE
Naive method	12355.97	17.47
Simple Average method	24146.06	34.34
Simple Moving Average Forecast (12 Months)	15192.01	16.10
Simple Moving Average Forecast (6 Months)	16294.34	16.80
Simple Moving Average Forecast (3 Months)	14756.73	15.82
Simple Exponential method	14764.97	15.83
Holt Exp method	11315.31	15.68
Holt Exponential Additive method	9026.50	8.44
Holt Exponential method Multiplicative	9976.49	10.12

FORECAST PLOTS OF ALL AUTO-REGRESSIVE TECHNIQUES



SARIMA performs the best among the Auto-regressive methods in capturing the forecasts accurately.
Results of ARMA and ARIMA super impose each other.
Moving average underforecasts and performs the poorest among the AR methods.

MAPE Comparisons for Auto-regression and Smoothing Techniques

Observations

p=1,q=1,d=1
P=1,Q=1,D=1,m-12

- **SARIMA method** performs the best as it predicts seasonality and trend accurately
- **Moving average method** model forecasts are the least accurate followed by ARMA and ARIMA.
- **ARMA and ARMA** forecasts are almost similar.

Method	RMSE	MAPE
Naive method	12355.97	17.47
Simple Average method	24146.06	34.34
Simple Moving Average Forecast (12 Months)	15192.01	16.10
Simple Moving Average Forecast (6 Months)	16294.34	16.80
Simple Moving Average Forecast (3 Months)	14756.73	15.82
Simple Exponential method	14764.97	15.83
Holt Exp method	11315.31	15.68
Holt Exponential Additive method	9026.50	8.44
Holt Exponential method Multiplicative	9976.49	10.12
Autoregressive (AR) method	10985.28	13.56
Moving Average (MA) method	23360.02	33.93
Autoregressive moving average (ARMA) method	22654.33	32.40
Autoregressive integrated moving average (ARIMA) method	22654.33	32.40
Seasonal autoregressive integrated moving average (SARIMA) method	9616.86	12.88

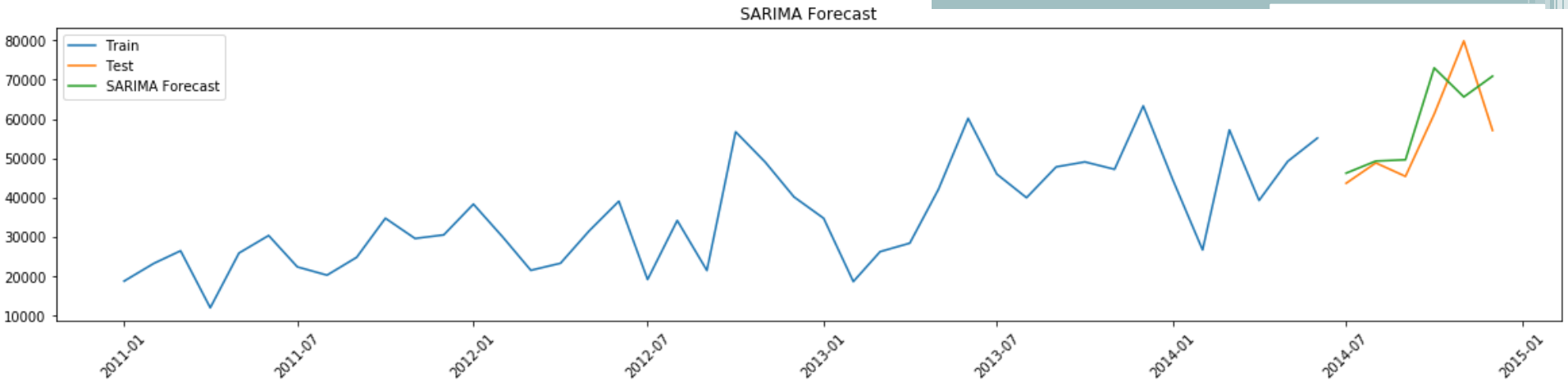
Conclusions:

- The series as noted from the plot is
  - Not noisy
  - There are more than 10 (48) observations
  - It shows an evident trend and seasonality.

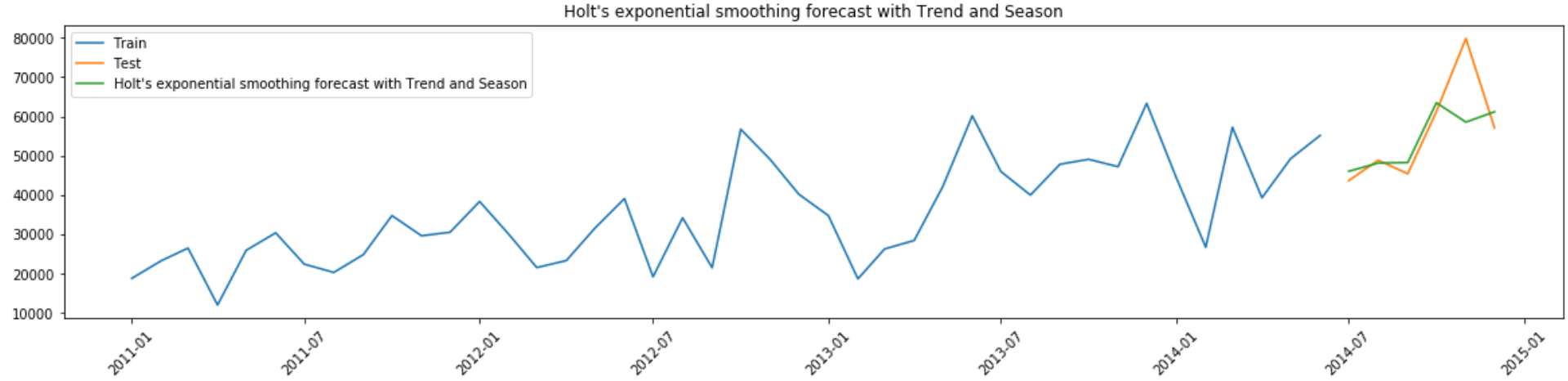
Hence, techniques which forecast trend and seasonality are the best fit here.
The MAPE values from the two best techniques are listed below.

Lower the MAPE values, better is the model performance

TYPE	Smoothing Technique	Auto-regressive technique
NAME	Holt Winters Additive Exponential Method	SARIMA(Seasonal autoregressive integrated and moving average)
MAPE Values	8.44	12.88



Plot of SARIMA and HOLT Winter's Method capture the level, trend and seasonality of the series



Recommendations to Global Mart

- Consumer-Asia Pacific is the most consistently profitably market-segment followed by Consumer-Latin America and consumer-US. It is advisable to invest in APAC-Consumer market-segment for now.

- Sales generally take a dip in July and rise up in the second half of the year. It is advised to stock up accordingly.