Retail Gaint Sales Forecast Case study

Jaswanth NA

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.

- Based on these, there are 21 unique "Market-Segments" for which the sales forecasts can be made. That is the dataset needs to be prepared such that you get the Order-Date, Sales and Profit for the 21 market segments.

- **Business Objective:**

- Check the forecast plot calculated on the test data and also the MAPE values for each of the above methods. (keep adding the MAPE values in a single table to compare them)

- Which forecast is able to predict the sales closer to the actual values and whose MAPE values is the least among all the methods done in the ARIMA above.

- Overall, we want the best forecasting method in the smoothing technique as well as the ARIMA set of techniques.

Solution Methodology

► Data cleaning and data manipulation

1. Check and handle duplicate data.

2. Check and handle NA values and missing values.

3. Find the 21 unique Market Segments by combining the respective 7 geographical markets for each of 3 segments such as Home office, Consumer and Corporate.

4. Imputation of the values, if necessary.

5. Once you have understood the dataset of the global data store, you need to get the order date in the required month-year format to make it a monthly aggregated transaction data. For this, convert the order-date into a date-time format for getting it into the Month-year format; you will get the data for 48 months now.

► Exploratory Data Analysis

► Perform the train-test split such that you take the 42 months as the train data and the 6 months as the test data.

► Calculate the CoV on the profit for each of the 21 market segments on the train data.

► Find the most profitable market segment by comparing the 21 CoV values.

► Calculated CoV and found the most profitable market segment.

► Perform the train-test split for that market segment that takes the 42 months data as the train data and the 6 months data as the test data.

➢ Decompose the time-series to get an idea of the trend, seasonality and residual parts of the data
➢ Applying Smoothing Techniques.
  ○ Simple exponential smoothing
  ○ Holt's exponential smoothing
  ○ Holt-Winters' exponential smoothing - Additive
  ○ Holt-Winters' exponential smoothing - Multiplicative
  ○ Check the forecast plot calculated on the test data and also the MAPE values for each of the above methods.

Apply the ARIMA set of techniques and check the sales forecast there. But here you will have to use the box-cox transformation and differencing to make the time-series stationery. For the box-cox transformation, you should use the value of lambda = 0. Note, you perform the box-cox on the entire dataset of that market segment and then train-test split it again before proceeding to the ARIMA set of methods

  ● AR model
  ● MA model
  ● ARMA model
  ● ARIMA model
  ● SARIMA model
  ● Conclude the method whose forecast is able to predict the sales closer to the actual values and whose MAPE values is the least among all the methods done in the ARIMA above. Check if this matches what you found from the flow chart.

21 Market Segments
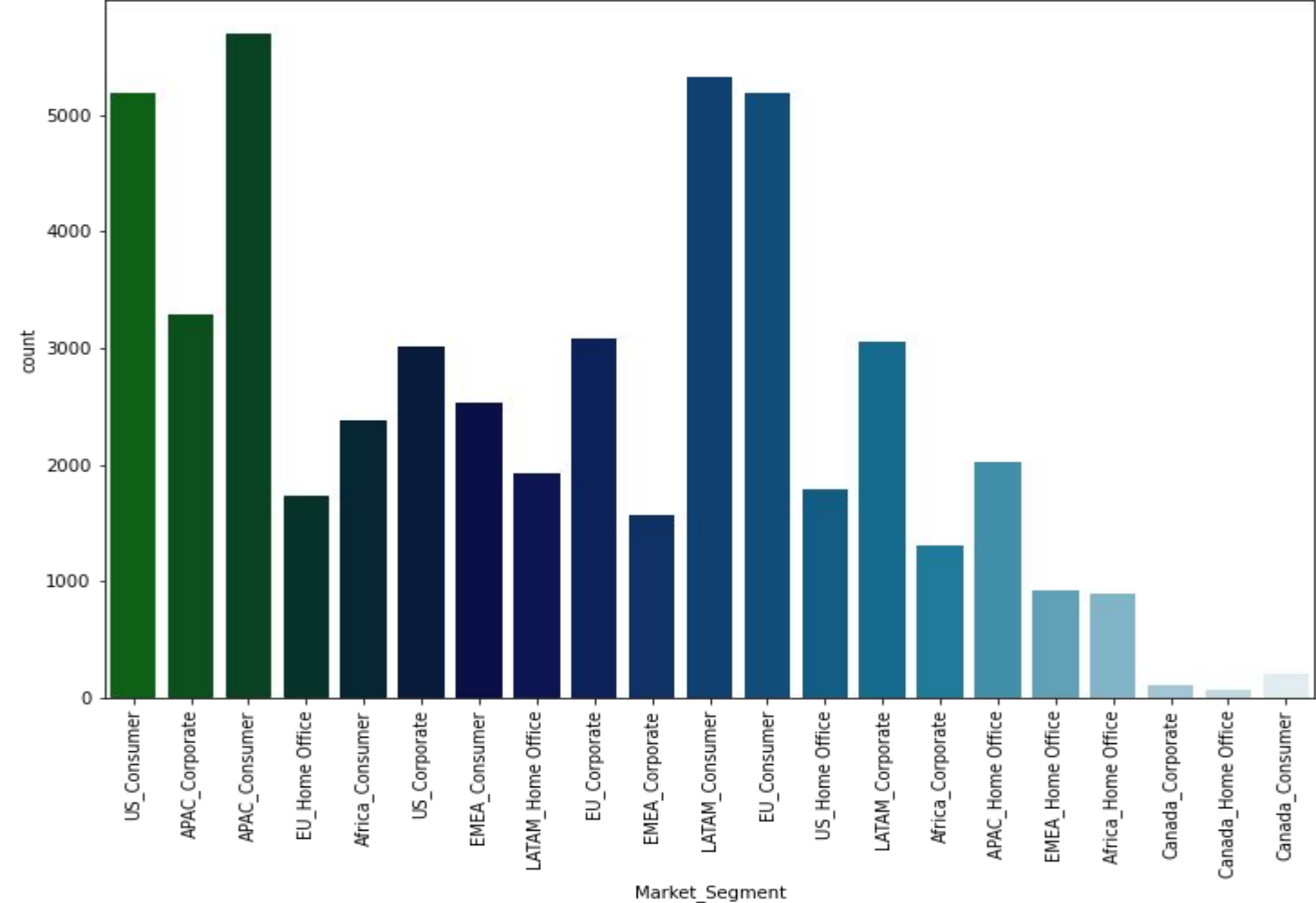
Market Segments

- 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

CoV on the profit for each of the 21 market segments on the train data.

We compare the variance between the segments using the coefficient of variation which will normalise the standard deviation with the mean and give a comparative figure on the basis of which we can identify the most profitable market segment. We want to forecast the sales where the market segment is reliable or in other words, there is less variation in the profits.

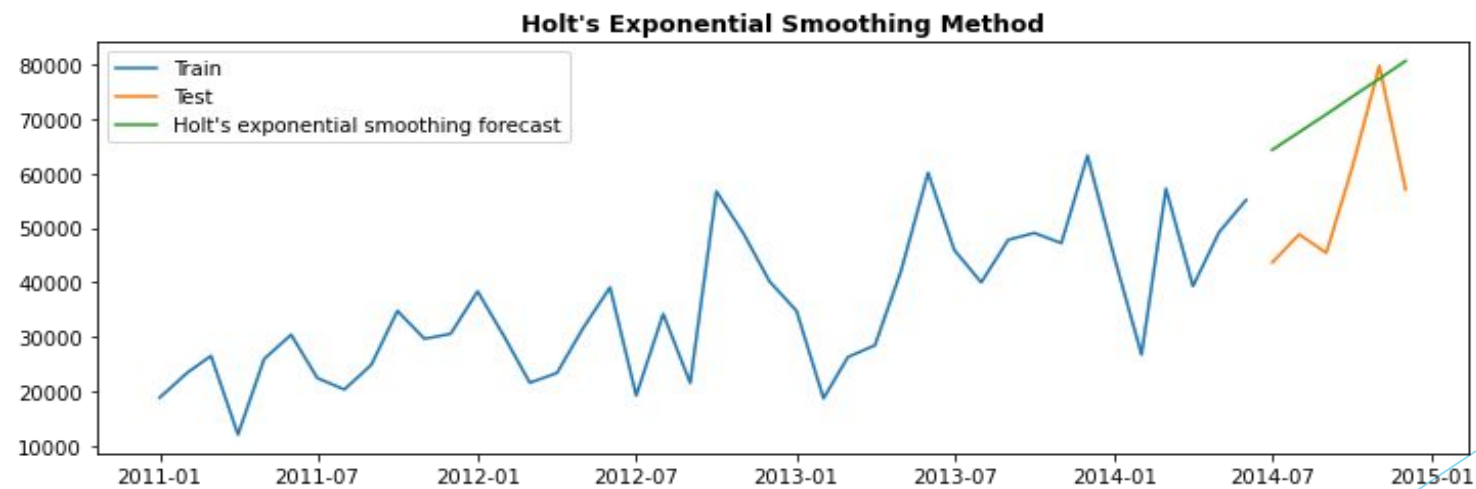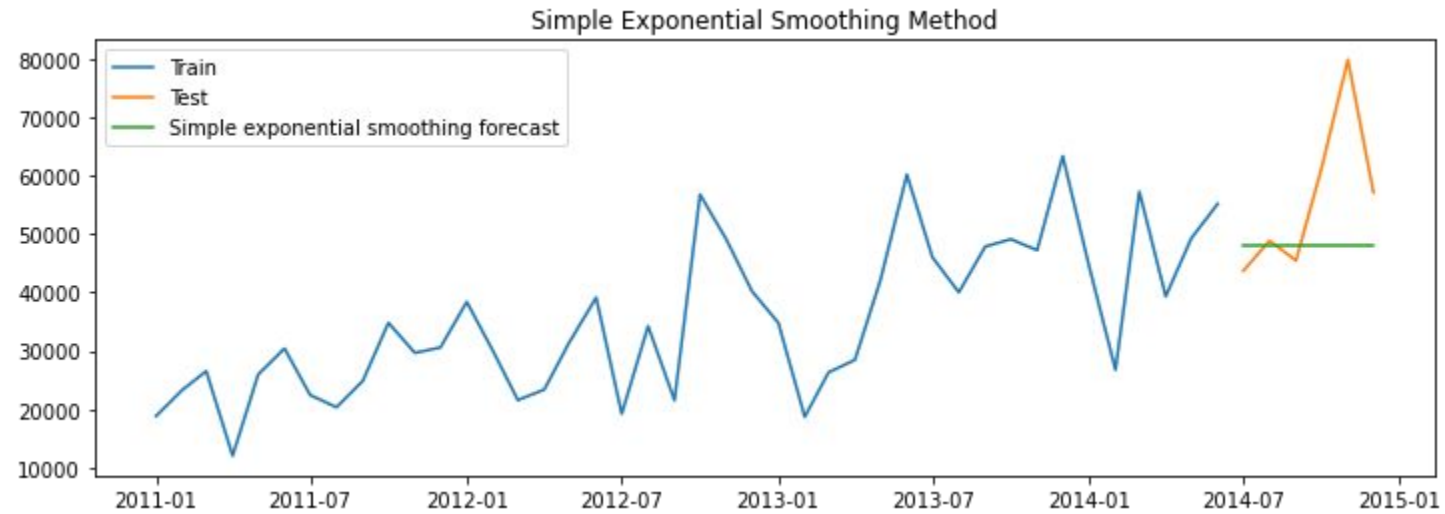|	Market_Segment	Mean	Std	CoV
0	APAC_Consumer	4400.894243	2300.457687	0.522725
1	APAC_Corporate	2574.919807	1364.837734	0.530051
12	EU_Consumer	3699.977143	2202.282289	0.595215
15	LATAM_Consumer	2295.555697	1569.632686	0.68377
13	EU_Corporate	2216.299429	1600.336696	0.722076
16	LATAM_Corporate	1122.633016	990.36088	0.882177
14	EU_Home Office	1224.456536	1148.627937	0.938072
2	APAC_Home Office	1511.088314	1523.508658	1.008219
18	US_Consumer	2686.740912	2715.031412	1.01053
19	US_Corporate	1754.199083	1880.200775	1.071829
20	US_Home Office	1132.065762	1272.476439	1.12403
17	LATAM_Home Offic	818.398941	957.275713	1.169693
6	Canada_Consumer	225.987632	282.555788	1.250315
3	Africa_Consumer	957.707	1254.932072	1.310351
7	Canada_Corporate	90.980294	162.493114	1.786025
4	Africa_Corporate	412.617571	780.56685	1.891744
5	Africa_Home Office	377.221071	759.322203	2.012937
8	Canada_Home Offic	118.00375	279.632866	2.369695
9	EMEA_Consumer	423.960286	1124.552711	2.652495
10	EMEA_Corporate	182.642643	1160.69843	6.355024
11	EMEA_Home Office	84.231366	651.283095	7.732073

As APAC_Consumer has lowest COV Value. which makes it as Best Market Segment
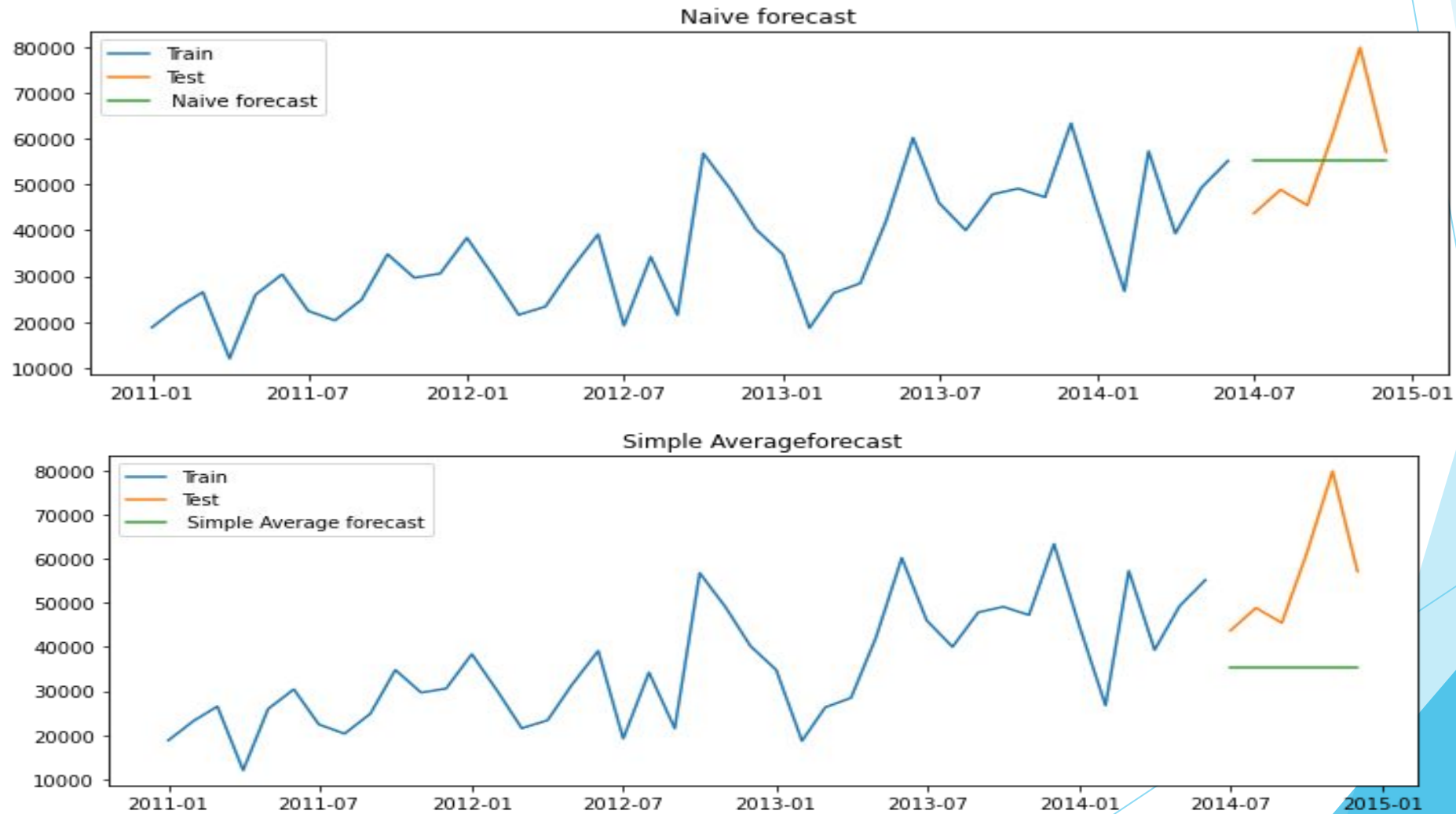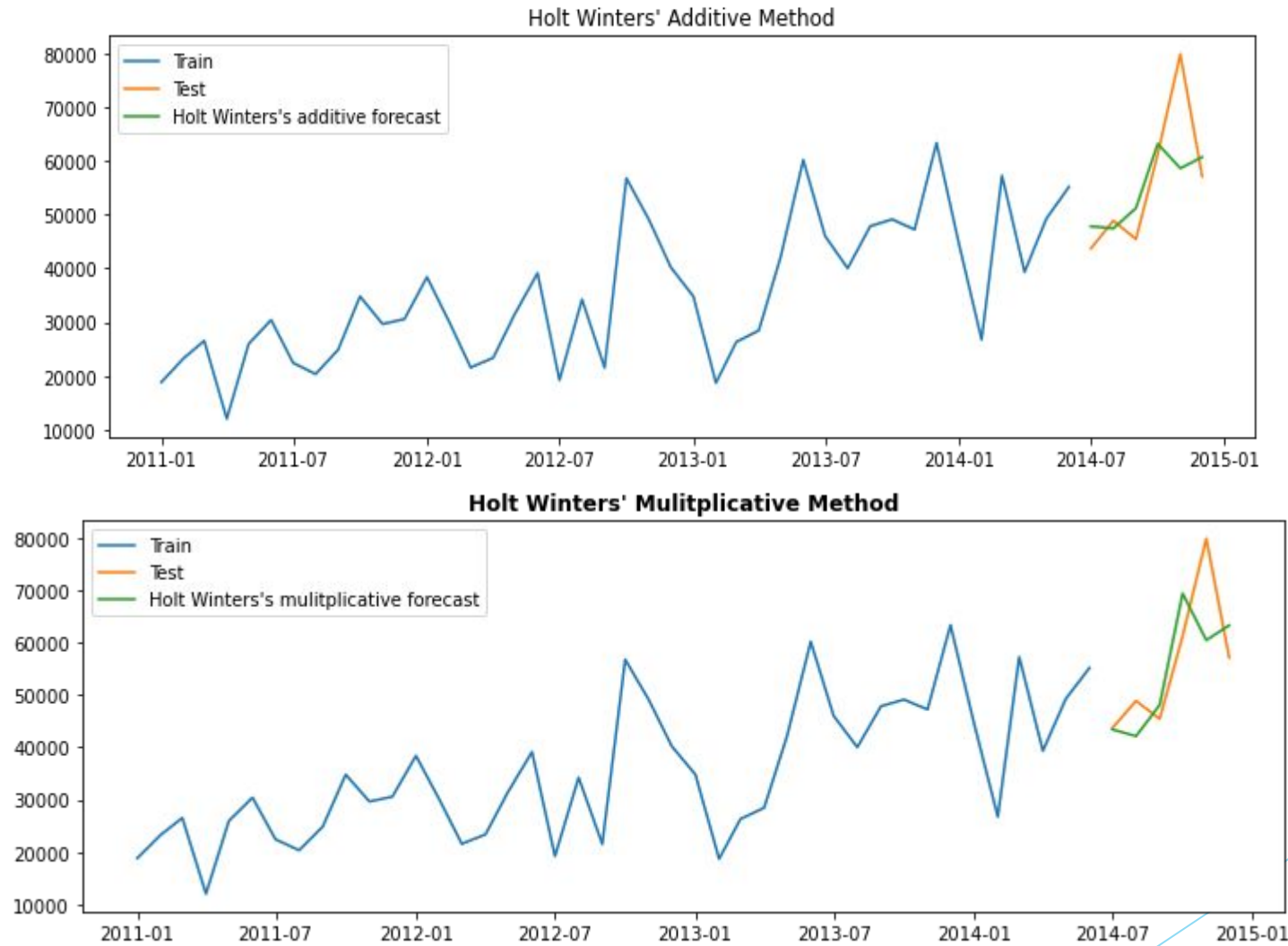
Exponential Smoothing Techniques

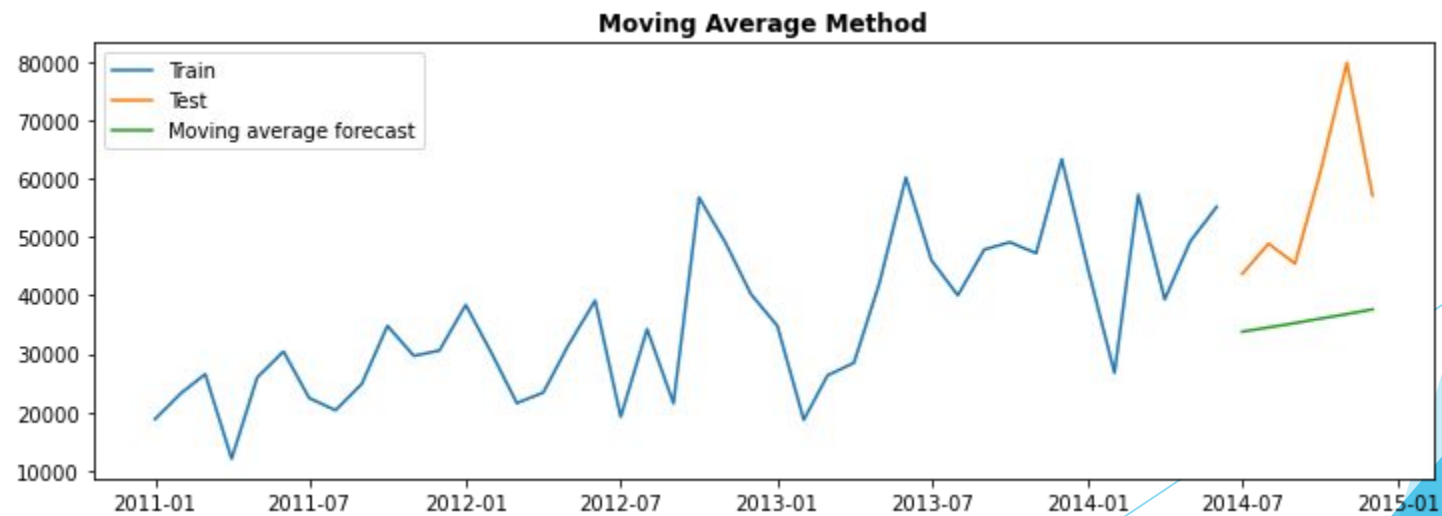# Exponential Smoothing Techniques

Simple Exponential Smoothing Method
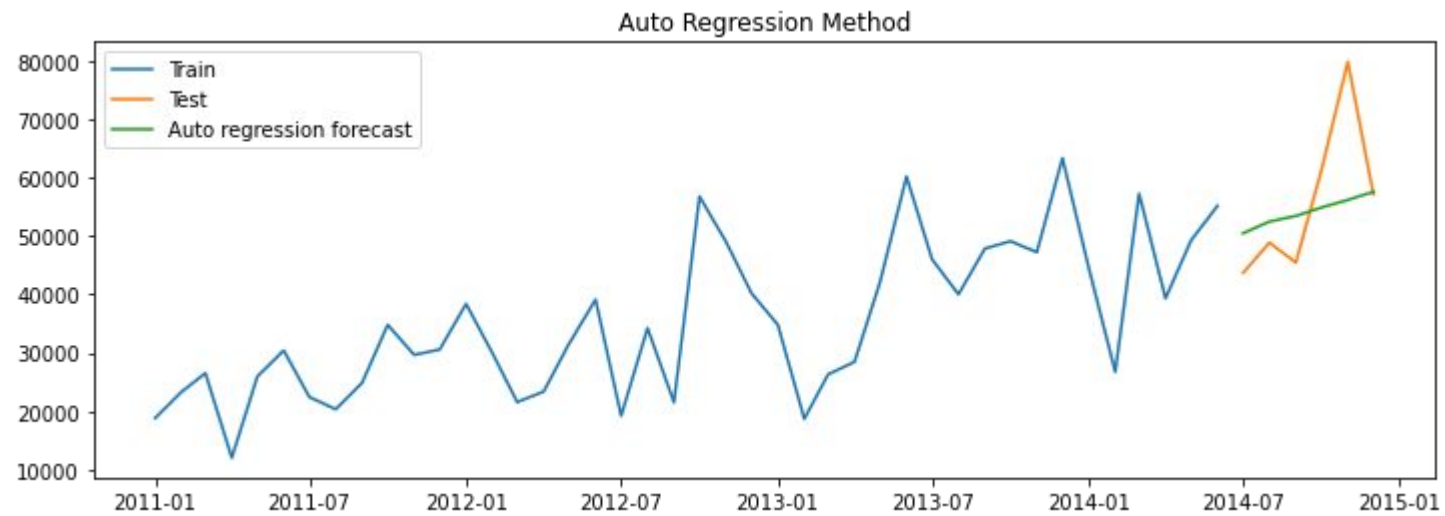

Holt's Exponential Smoothing Method

Exponential Smoothing Techniques

Holt Winters' Additive Method

Holt Winters' Mulitplicative Method

Exponential Smoothing Techniques

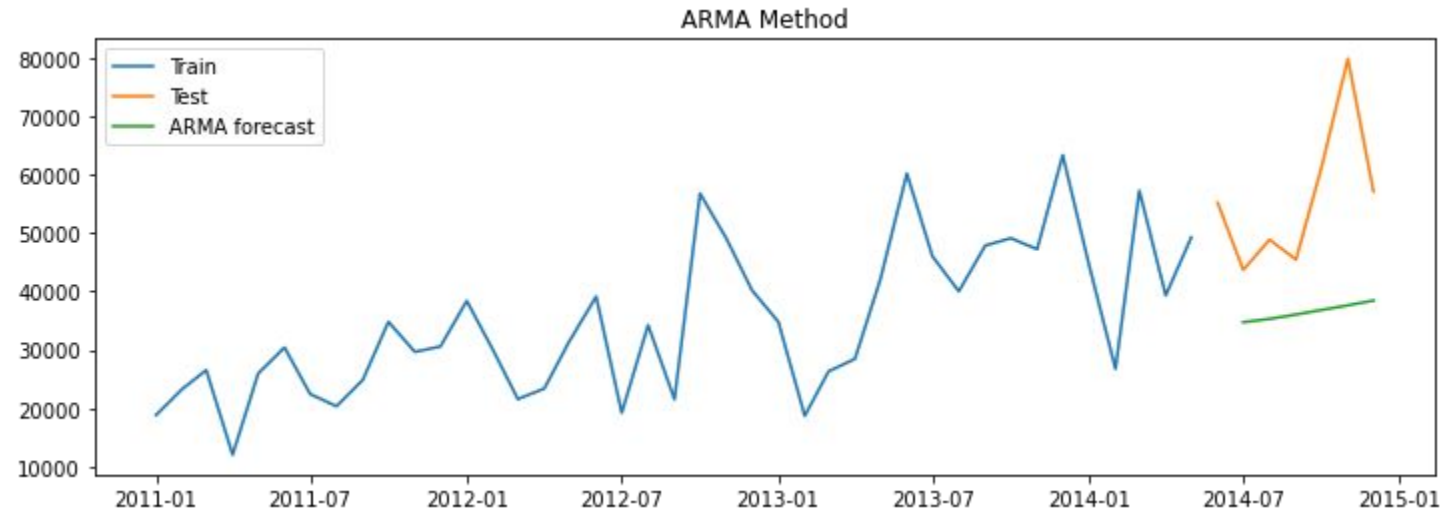| Method	RMSE	MAPE
Naïve Method	12355.97	17.47
Simple Moving Average	24146.06	34.34
Simple Moving Average Forecast	14756.73	15.82
Simple Exponential Smoothing Forecast	14627.34	15.74
Holt's Exponential Smoothing Method	18976.37	34.57
Holt Winters's Additive Method	9306.82	10.17

- From the Smoothing Techniques performed we can conclude that Holt Winter's Additive Method is giving the better forecast to of the sales for the 6 months, since the predicted sales are closer to the actual sales.
- We could also see that the RMSE and MAPE values is the least among all the methods done above.
- **Holt Winter's Additive method has the lowest RMSE and MAPE values. Which means error measures are very less in this model**
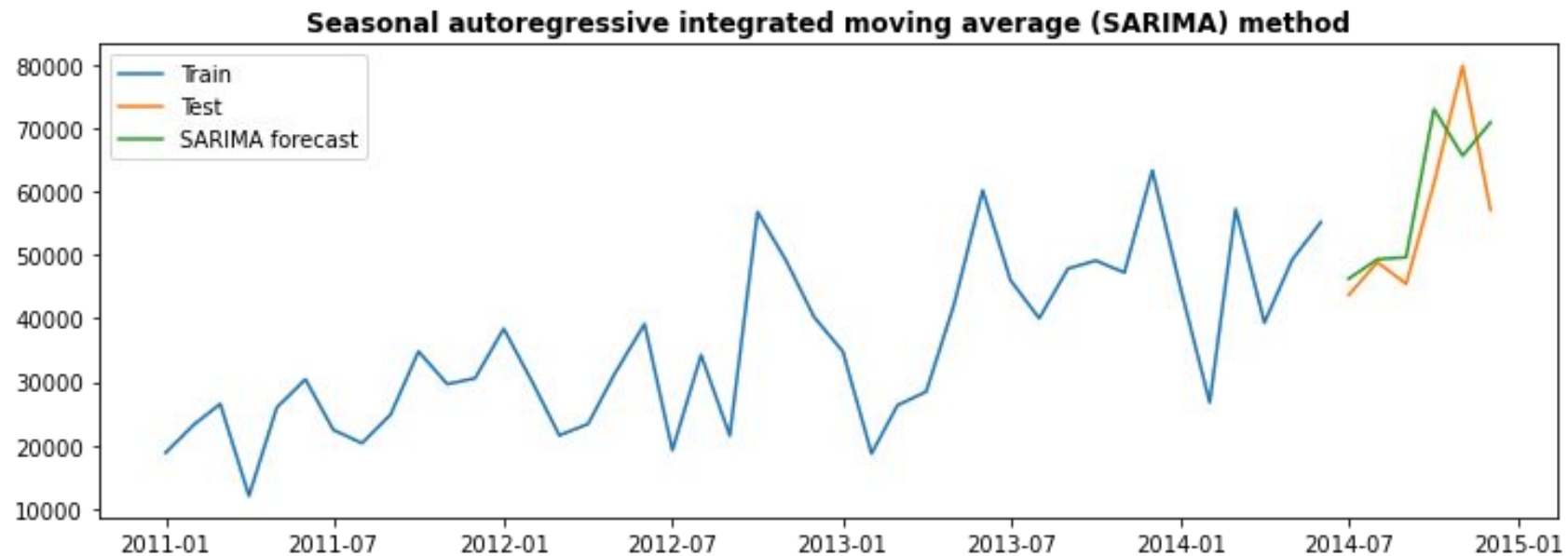
AR Techniques

Auto Regression Method

Moving Average Method

AR Techniques

ARMA Method

Autoregressive integrated moving average (ARIMA) method

AR Techniques

Seasonal autoregressive integrated moving average (SARIMA) method
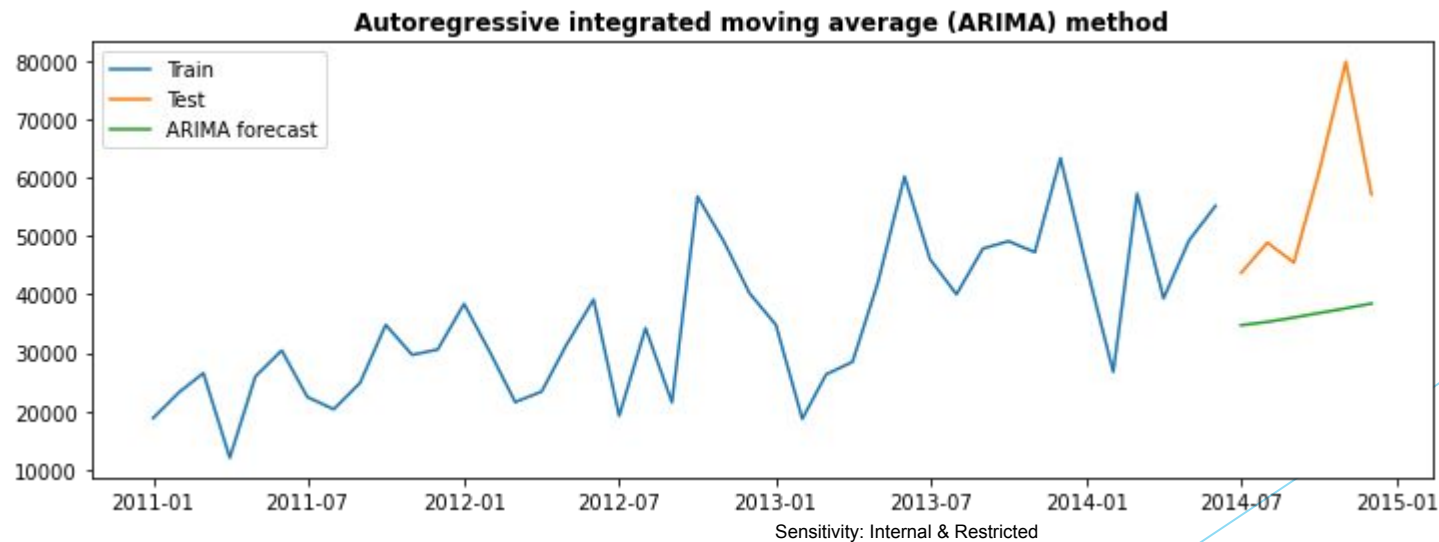
SARIMA forecast captured both trend and seasonality

Smoothing & AR Techniques

Method	RMSE	MAPE
Naïve Method	12355.97	17.47
Simple Moving Average	24146.06	34.34
Simple Moving Average Forecast	14756.73	15.82
Simple Exponential Smoothing Forecast	14627.34	15.74
Holt's Exponential Smoothing Method	18976.37	34.57
Holt Winters's Additive Method	9306.82	10.17
Auto Regression Method	10985.28	13.56
Moving Average Method	23360.02	33.93
Auto Regression Moving Average Method	22654.32	32.4
Auto Regression Integrated Moving Average Method	22654.32	32.4
Seasonal Auto Regression Integrated Moving Average Method	9616.57	12.87

- **Among all the methods done in the ARIMA above, we can conclude that forecast done using SARIMA method is able to predict the sales closer to the actual values**
- **RMSE and MAPE values for this method is the least among all the methods done**
- **Seasonal Auto Regression Integrated Moving Average Method**

Conclusion

➤ Holt Winters' additive method is the best method in Smoothing Technique

➤ SARIMA – Seasonal Autoregressive Integrated moving average is the best method in ARIMA set of techniques