TIME SERIES FORECASTING

BY: Prateek Rana

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

Objective: To Forecast the sales of the profits of the store for the next 6 months.

PROBLEM STATEMENT:

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

The store has 7 geographical markets and 3 segments. On the left, in the table, we can see the 21 market segments We need to find the most consistently profitable market-segment and forecast the sales and demand for that market-segment. We can do that by calculating the "Coefficient of Variation (CoV)" and select the market-segment with the highest CoV. After acquiring the most consistently profitable market-segment, we must forecast the profit for the next 6 months. We must choose the right time series forecasting model according to the possible trends and seasonality.

FORECASTING APPROACH

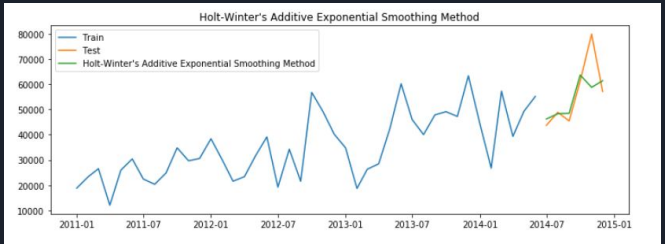
	MS	cov
0	APAC-Consumer	0.522725
1	APAC-Corporate	0.530051
12	EU-Consumer	0.595215
15	LATAM-Consumer	0.683770
13	EU-Corporate	0.722076
16	LATAM-Corporate	0.882177
14	EU-Home Office	0.938072
2	APAC-Home Office	1.008219
18	US-Consumer	1.010530
19	US-Corporate	1.071829
20	US-Home Office	1.124030
17	LATAM-Home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

- The dataset provided is of a Global Supermarket which is clean and has the columns namely: Order Date, Segment, Market, Sales and Profit.
- I need to forecast sales for the most consistently profitable market segment. So first, I convert the Order Date column into a datetime format. Then I create a column named market segment having both market and segment columns' values.
- Aggregate data and find sum of profit for various months and market segments with the help of a pivot table. Split it into train(42) and test(6) data.
- Coefficient of variance helps us to identify which series is more fluctuating. Its calculated as standard deviation/mean. Higher the CoV, more the fluctuations. I then find the CoV value for various market segments.
- Market Segment with the lowest CoV value calculated on the profit is 'APAC-Consumer' with 0.522725 as seen in the table. This means that this market segment has the least fluctuations and is most profitable, thus, making it the most consistently profitable market segment.
- Using the original dataset, I filter the data for the 'APAC-Consumer' market segment.
- Aggregate the data based on Order Date to find sum of sales and split the data into train-test 42:6.
- I then use this train data to train Exponential Smoothing techniques and ARIMA models.

IMPLEMENTATION

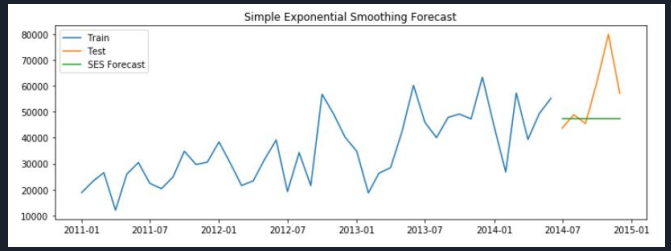
EXPONENTIAL SMOOTHING TECHNIQUES

	Method	RMSE	MAPE
0	Simple Exponential Smoothing method	15011.49	15.99
0	Holt's Exponential Smoothing Method	18976.37	34.57
0	Holt-Winter's Exponential Smoothing Method	8994.00	8.59
0	Holt-Winter's Multiplicative Exponential Smoot...	9976.52	10.12

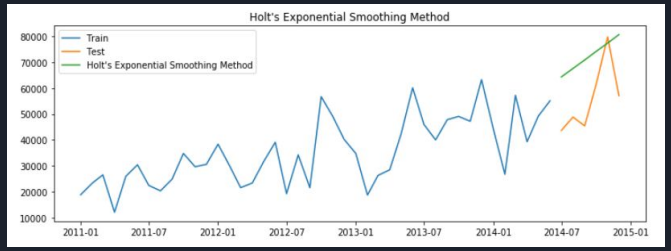


- MAPE Value or mean absolute percentage error is a measure of prediction accuracy of any forecasting method.
- I performed various exponential smoothing techniques namely:
  Simple Exponential Smoothing Method
  Holt's Exponential Smoothing Method
  Holt-Winter's Exponential Smoothing Method (Additive)
  Holt-Winter's Exponential Smoothing Method (Multiplicative)
- As we can see from the table, Holt-Winter's Additive Exponential Smoothing Method has the least MAPE value of 8.59, which means that we must choose this method over others.
- We can see from the train, test and forecast plot as well that the exponential smoothing technique is predicting the sales very closely to the actual sales.
- Holt-Winter's takes into account the trend level and seasonality into the picture, where as the other techniques are able to take either trend or both trend and level. Since there is seasonality present in the time series data, the other techniques could not predict as well as Holt-Winter's.

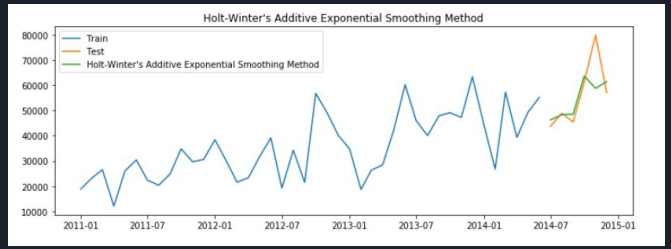
COMPARISON OF EXPONENTIAL SMOOTHING TECHNIQUES



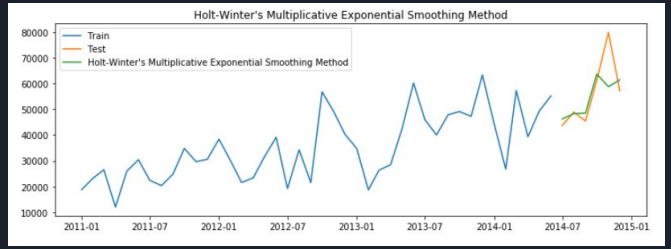
Captures level.



Captures level and trend.

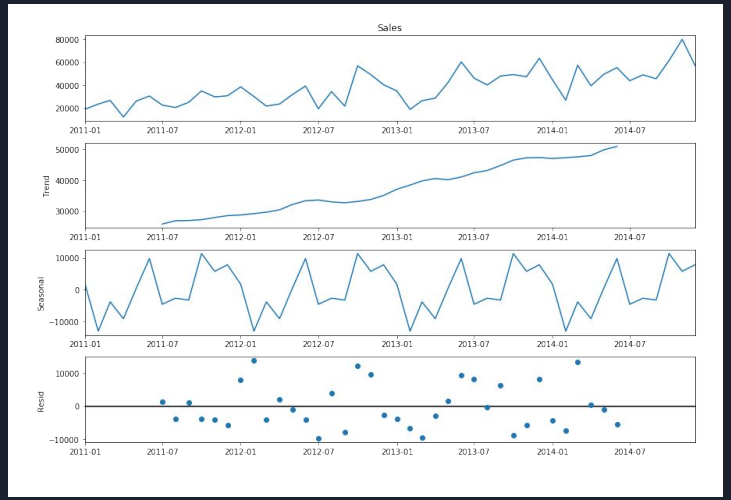


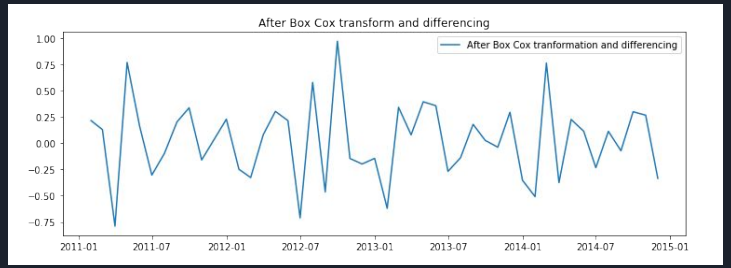
Captures level, trend and seasonality.



Captures level, trend and seasonality.

STATIONARITY CHECKS

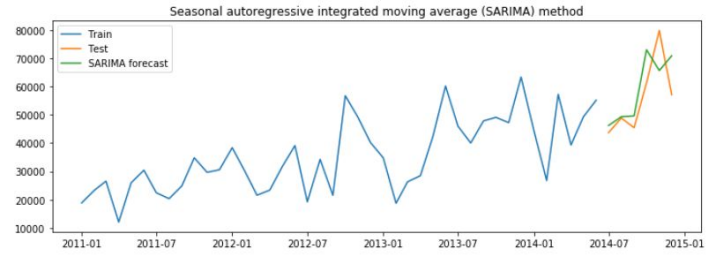




- Before using the data for ARIMA models for forecasting the sales I performed some stationarity checks as the data is required to have constant mean, variance and covariance.
- I performed the Augmented Dickey-Fuller (ADF) test and got a p-value of 0.0118 so I had to reject the null hypothesis of the time series being not stationary.
- However, with the help of time series additive decomposition I could see that there is clearly a trend and a seasonal effect.
- Therefore, I performed box cox transformation with lambda=0 and Differencing methods to get rid of variance and trends.
- After box cox transformation and differencing the time series was stationary as we can see in the figure below.

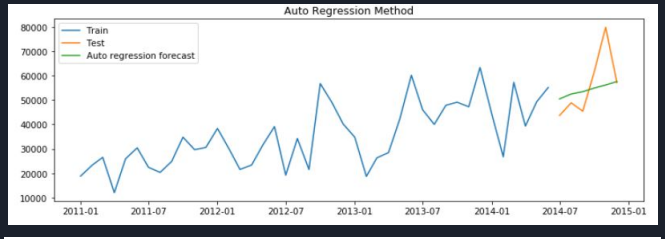
ARIMA METHODS

	Method	RMSE	MAPE
0	Simple Exponential Smoothing method	15011.49	15.99
0	Holt's Exponential Smoothing Method	18976.37	34.57
0	Holt-Winter's Exponential Smoothing Method	8994.00	8.59
0	Holt-Winter's Multiplicative Exponential Smoot...	9976.52	10.12
0	Autoregressive (AR) method	10985.28	13.56
0	Moving Average (MA) method	23360.02	33.93
0	Autoregressive moving average (ARMA) method	22654.32	32.40
0	Autoregressive integrated moving average (ARIM...	22654.32	32.40
0	Seasonal autoregressive integrated moving aver...	9609.20	12.85

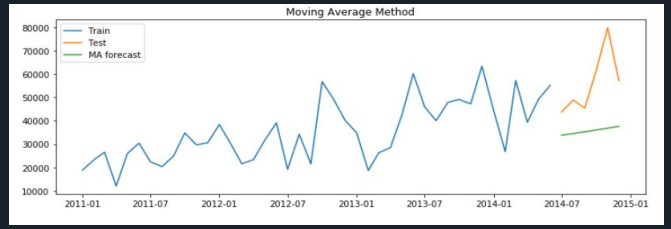


- I performed ARIMA methods and created models to predict the sales.
- From the table we can see that Seasonal Autoregressive Integrated Moving Average method has the least MAPE value of 12.85. Thus, out of all the other ARIMA models we choose SARIMA model.
- From the train, test and forecast plot of the SARIMA model below the table we can see that the model predicts the sale quite closely to the actual sales.

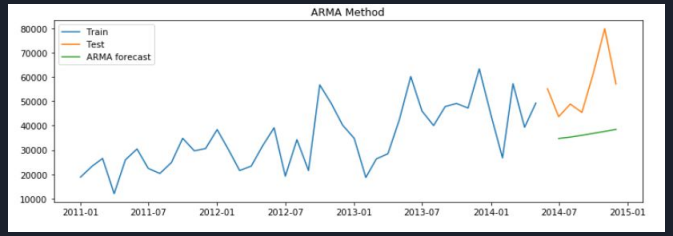
COMPARISON OF ARIMA MODELS



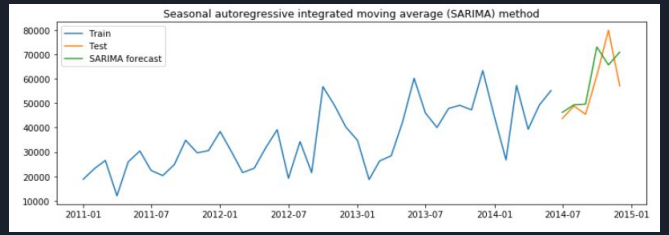
Captures level and trend the better than ARMA and ARIMA.



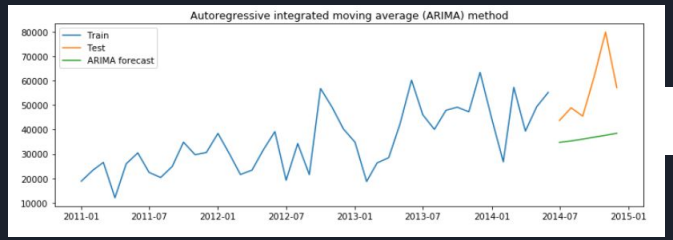
Captures trend.



Captures level and trend



Captures level, trend and seasonality.



Captures level and trend similar to ARMA.

CONCLUSION

We choose the Holt-Winter's Additive Exponential Smoothing Technique with a MAPE value of 8.59 amongst all the Exponential Smoothing Techniques. This method performs best when there is a trend in the time series data and a seasonal effect as well. With the help of decomposition we could see that there is trend and seasonality present in the time series data in the Exponential Smoothing Techniques. Thus, the mean absolute percentage is less as the predicted values are very close to not only the trend but the peaks and troughs of the seasons.

From the ARIMA Models we choose the SARIMA model with the MAPE value of 12.85 amongst all the other models. SARIMA or Seasonal Auto Regression Integrated Moving Average consists of 6 parameters:

p & seasonal P = Indicate number of autoregressive terms (lags of stationary data).

d & seasonal D = Indicate differencing that must be done to stationarize the series.

q & seasonal Q = Indicate number of moving average terms (lags of forecast errors).

With the help of these parameters SARIMA model is able to take the trend and seasonality into picture and predict the future values accordingly. SARIMA model thus works best when there is trend and seasonality present in the time series data amongst all the ARIMA models. The mean absolute percentage is less because the prediction or the forecast is very close to the actual sales, also capturing the level, trend and peaks and troughs of the season.

THANK YOU