1. What are the 21 Market Segments?

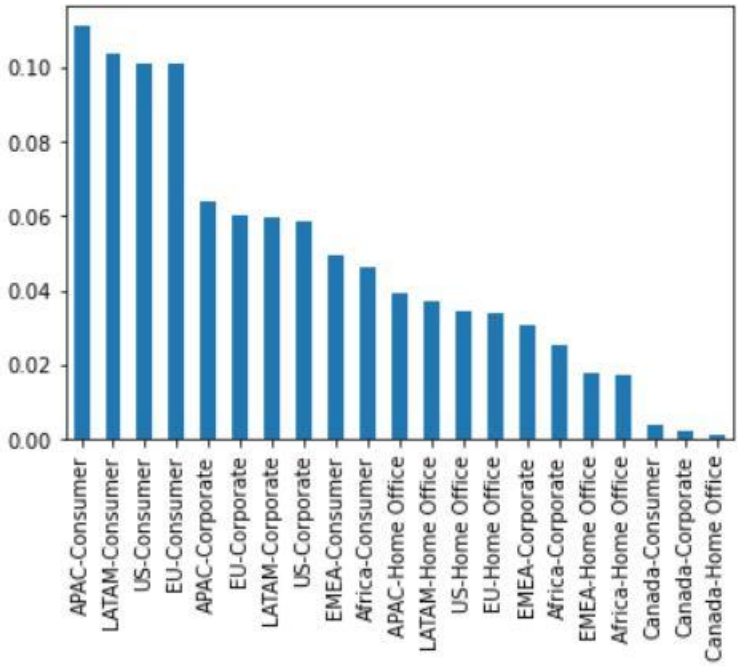
```
In [9]: #We observe that there are now 21 market segments:
        root_df.Market_Segment.value_counts().shape
```

```
Out[9]: (21,)
```

```
In [10]: root_df.Market_Segment.value_counts()
```

```
Out[10]: APAC-Consumer          5699
         LATAM-Consumer         5321
         US-Consumer            5191
         EU-Consumer            5186
         APAC-Corporate         3283
         EU-Corporate           3077
         LATAM-Corporate        3053
         US-Corporate           3020
         EMEA-Consumer          2538
         Africa-Consumer        2381
         APAC-Home Office       2020
         LATAM-Home Office      1920
         US-Home Office         1783
         EU-Home Office         1737
         EMEA-Corporate         1574
         Africa-Corporate       1312
         EMEA-Home Office        917
         Africa-Home Office      894
         Canada-Consumer         202
         Canada-Corporate        110
         Canada-Home Office       72
         Name: Market_Segment, dtype: int64
```

```
In [11]: root_df.Market_Segment.value_counts(normalize=True).plot.bar()
Out[11]: <AxesSubplot:>
```



2. Comparison showing the Table of Values for the Coefficient of Variation(COV) calculated on the profit for the 21 Market Segments

```
In [109]: cov.head(21)
```

Out[109]:

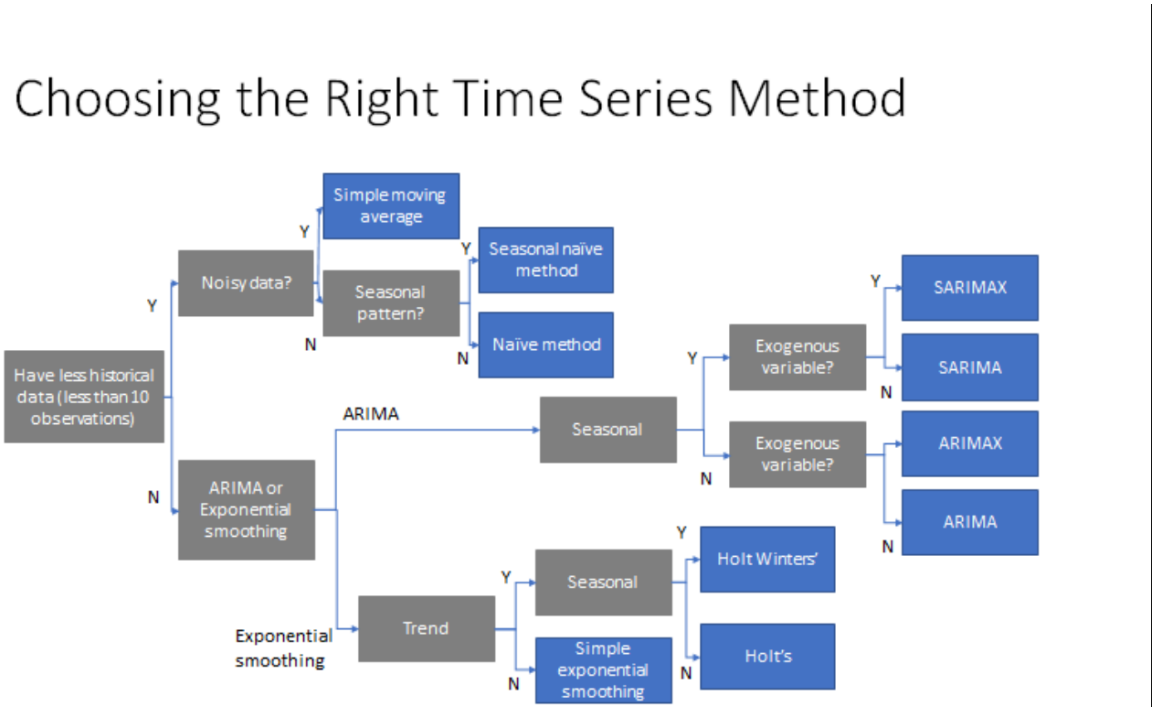
	Market_Segment	COV	mean_profit	stddev_profit
0	APAC-Consumer	0.522725	4400.894243	2300.457687
1	APAC-Corporate	0.530051	2574.919807	1364.837734
12	EU-Consumer	0.595215	3699.977143	2202.282289
15	LATAM-Consumer	0.683770	2295.555697	1569.632686
13	EU-Corporate	0.722076	2216.299429	1600.336696
16	LATAM-Corporate	0.882177	1122.633016	990.360880
14	EU-Home Office	0.938072	1224.456536	1148.627937
2	APAC-Home Office	1.008219	1511.088314	1523.508658
18	US-Consumer	1.010530	2686.740912	2715.031412
19	US-Corporate	1.071829	1754.199083	1880.200775
20	US-Home Office	1.124030	1132.065762	1272.476439
17	LATAM-Home Office	1.169693	818.398941	957.275713
6	Canada-Consumer	1.250315	225.987632	282.555788
3	Africa-Consumer	1.310351	957.707000	1254.932072
7	Canada-Corporate	1.786025	90.980294	162.493114
4	Africa-Corporate	1.891744	412.617571	780.566850
5	Africa-Home Office	2.012937	377.221071	759.322203
8	Canada-Home Office	2.369695	118.003750	279.632866
9	EMEA-Consumer	2.652495	423.960286	1124.552711
10	EMEA-Corporate	6.355024	182.642643	1160.698430
11	EMEA-Home Office	7.732073	84.231366	651.283095

---

3. The Reason Why Market Segment “ABC” is the most profitable segment.

Answer: COV of a series is the ratio of Standard deviation to it's mean. So if the mean is high and standard deviation is low, it indicates that the series is much stable and not prone to variations or fluctuations.
From the Above COV(Coefficient of Variation) table we see that ****APAC-Consumer**** is the market segment with least COV(of 0.522) indicating the most profitable market segment.

4. Concluding the Optimum Technique from the Flow Chart that might work best for the Sales Forecast.



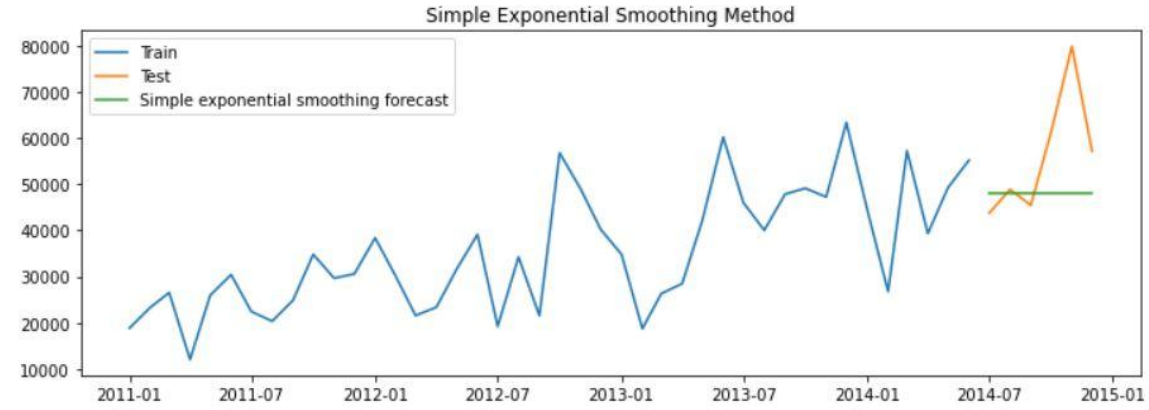
Inference from above flow chart and our plot of Time Series Retail giant sales data:

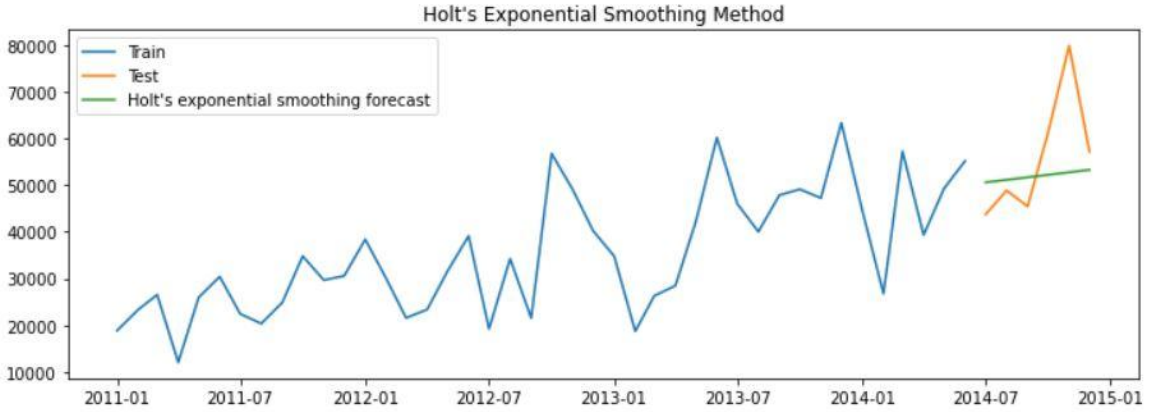
- Our Time Series data has more than 10 observations. Here in our case we have observations for 48 months with seasonality of 12 months.
- Our Time Series Data has all three components of Level, Trend and Seasonality.
- Our Time Series Data doesn't have any other exogenous variable.

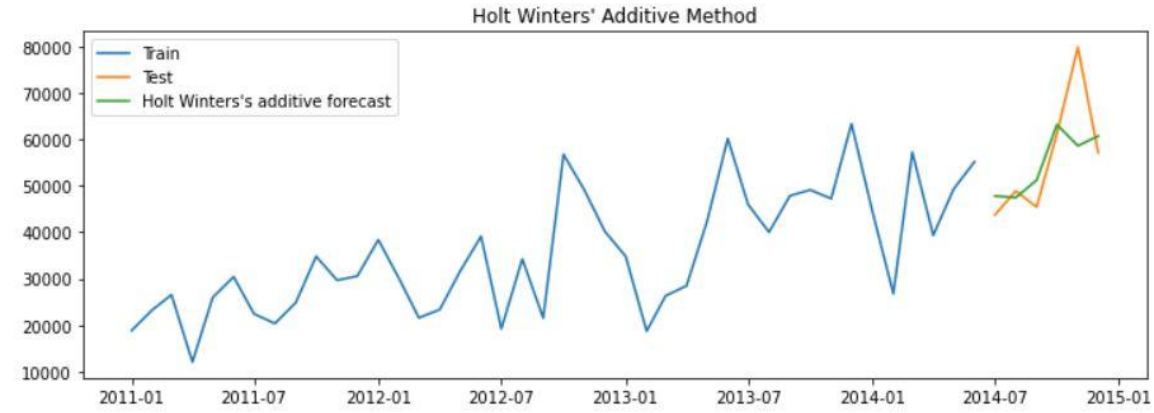
Which model to use?

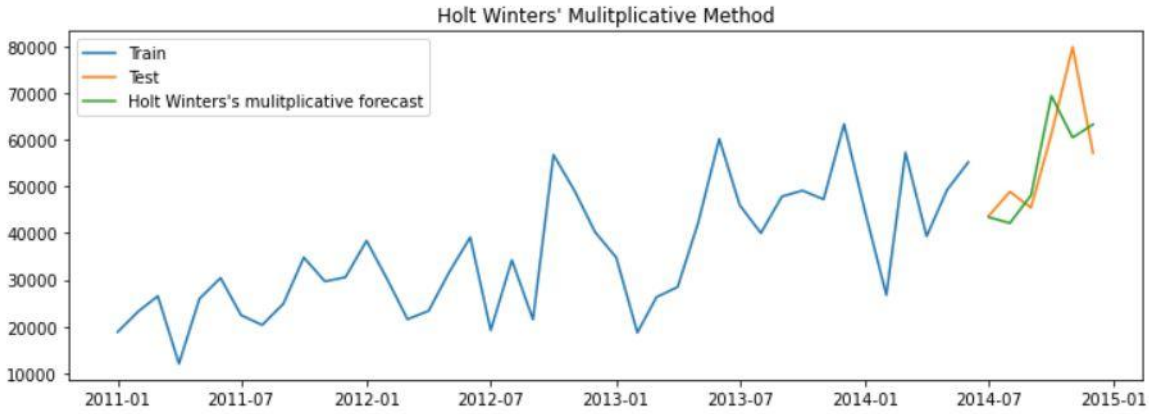
- Thus we can use both ARIMA or Exponential Smoothing Methods.
- Here we conclude that in exponential smoothing we use **"Holt Winters Method"** and in ARIMA method we use the "**SARIMA**" method as these two methods capture all three components of Level, Trend and Seasonality in our Time Series Data.
- Lets however go step by step with all models and check if our prediction is true.

5. Comparing the Sales Forecast plots for all the smoothing techniques and their MAPE values:


Simple Exponential Smoothing Method
Train
Test
Simple exponential smoothing forecast
80000
70000
60000
50000
40000
30000
20000
10000
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01


Holt's Exponential Smoothing Method
Train
Test
Holt's exponential smoothing forecast
80000
70000
60000
50000
40000
30000
20000
10000
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01

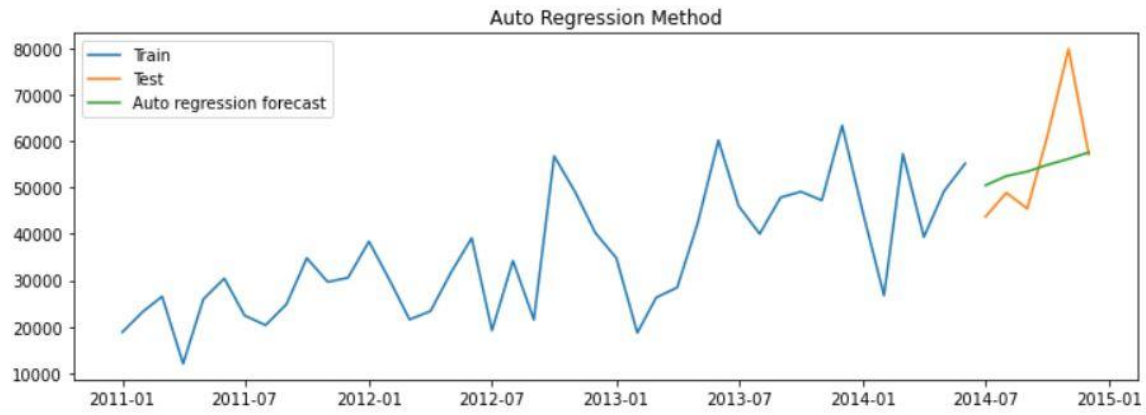

Holt Winters' Additive Method
Train
Test
Holt Winters's additive forecast
80000
70000
60000
50000
40000
30000
20000
10000
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01

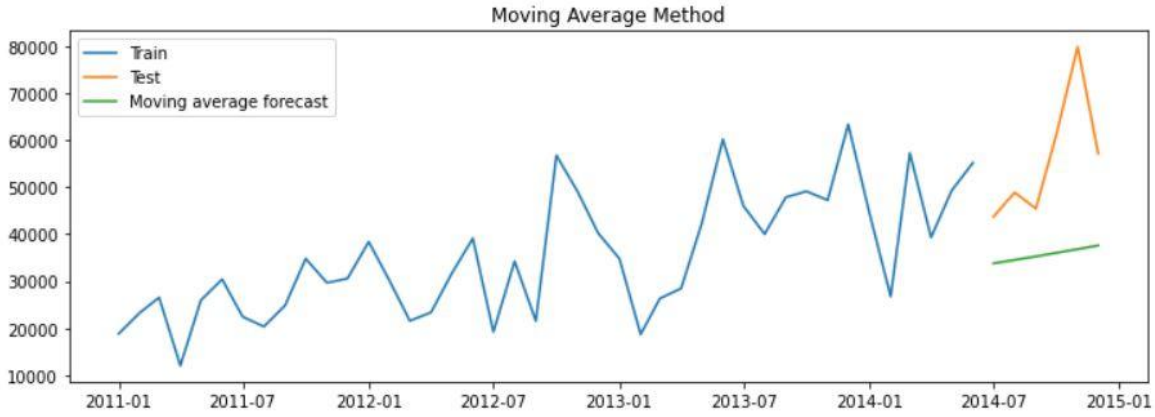


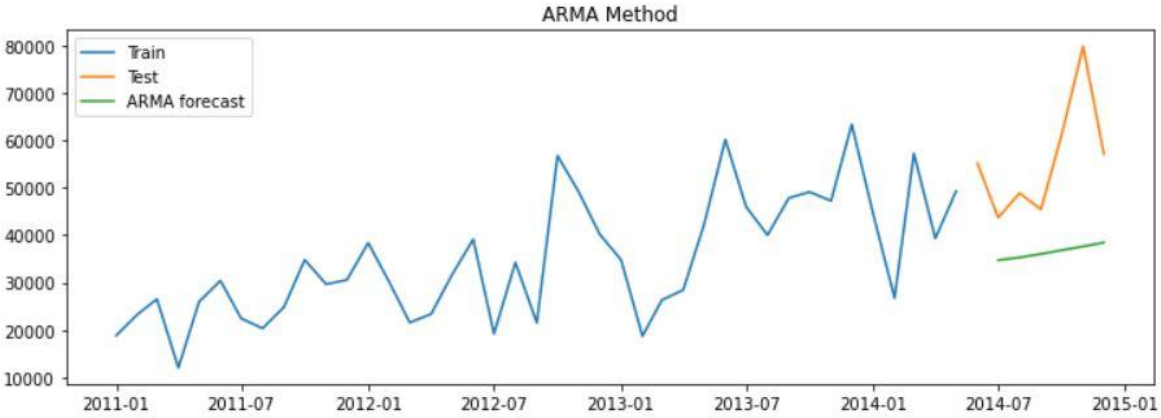
Out[54]:

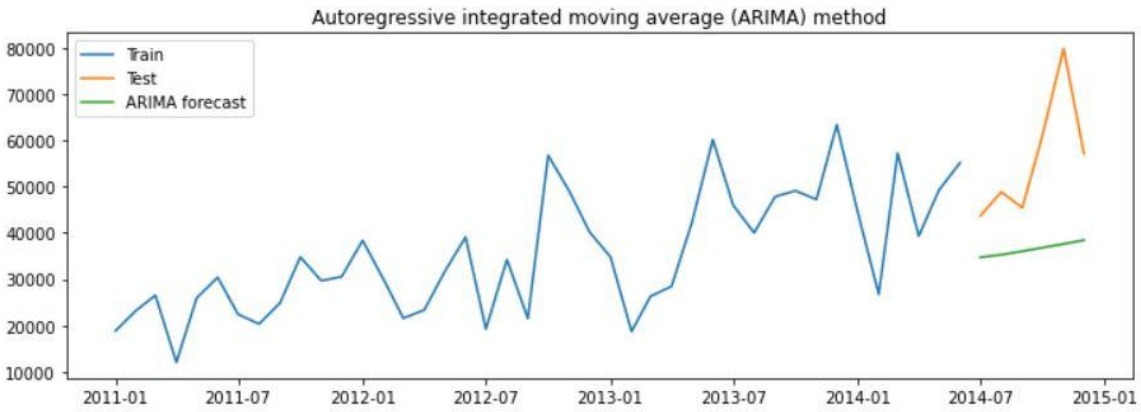
	Method	RMSE	MAPE
0	Simple exponential smoothing forecast	14627.34	15.74
0	Holt's exponential smoothing method	12403.84	14.93
0	Holt Winters' additive method	9306.82	10.17
0	Holt Winters' multiplicative method	9423.23	11.43

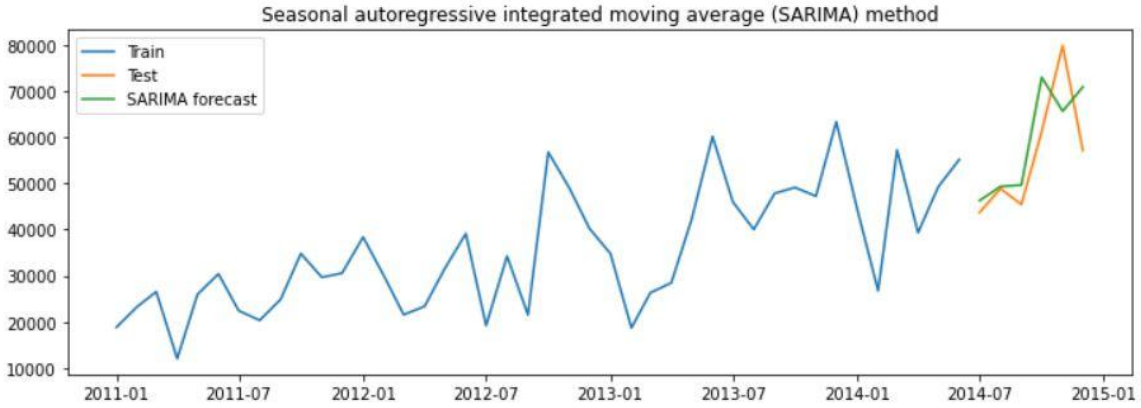
6. Comparing the Sales Forecast plots for all the ARIMA techniques and their MAPE values:


Auto Regression Method
Train
Test
Auto regression forecast
80000
70000
60000
50000
40000
30000
20000
10000
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01


Moving Average Method
Train
Test
Moving average forecast
80000
70000
60000
50000
40000
30000
20000
10000
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01


ARMA Method
Train
Test
ARMA forecast
80000
70000
60000
50000
40000
30000
20000
10000
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01





7. Conclusion on which technique works best for the Sales Forecast and why? Then reason this using the forecast plot and the MAPE values both.

Answer: Final Inferences:

1. We see that in the Exponential Smoothing Methods "Holt Winters Additive Method" with RMSE(Root Mean Squared Error) of `9306.82` and MAPE(Mean Absolute Percentage Error) of `10.17` is the lowest compared to any other smoothing method. Thus we choose this method for our forecasting. This is also intuitively obvious as "Holt Winters Additive Method" captured all three components of Level, Trend and Seasonality. This is also evident in the forecast plot that we saw above, as the Green Line(the forecast line) is quite close to the Yellow line(Actual Line).

2. In ARIMA methods we see that SARIMA method with RMSE of `9618.73` and MAPE of `12.88` is the lowest compared to other ARIMA methods. Thus we choose this method for our forecasting. This is also intuitively obvious as "SARIMA" captured all three components of Level, Trend and Seasonality. This is also evident in the forecast

plot that we saw above, as the Green Line(the forecast line) is quite close to the Yellow line(Actual Line).

**** Thank You IIITB and Upgrad For the Learning Experience****

**** This PPT is submitted by Achal Kagwad: Dated 5th July 2021****