

# Retail Giant Sales Forecast Case Study

By- Zishan Ahmed

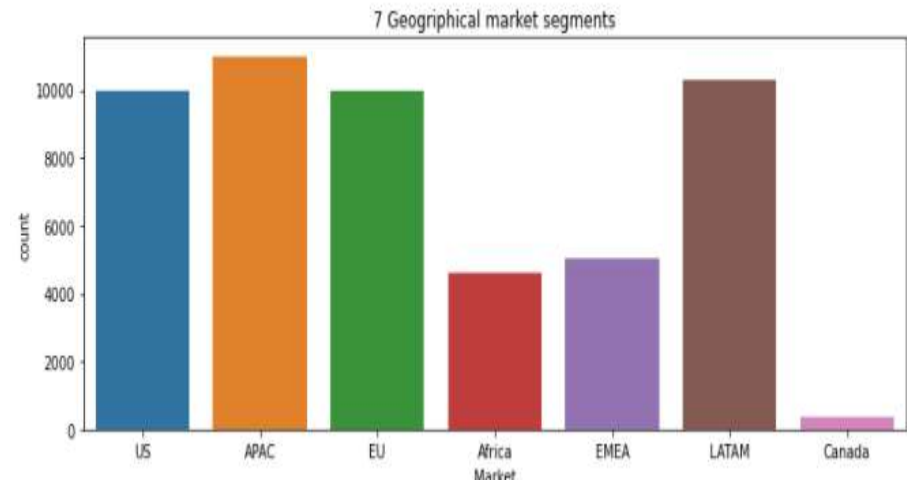
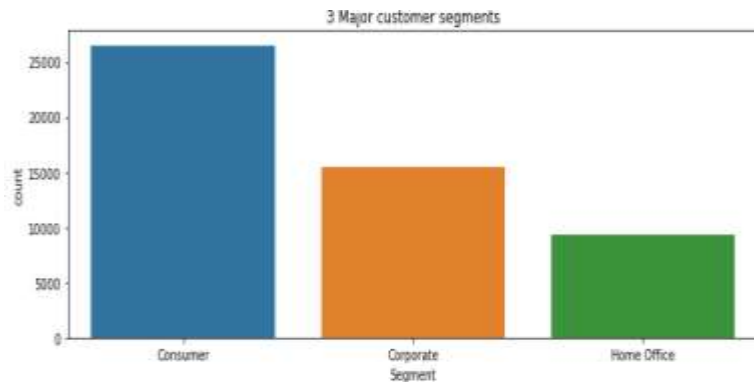
# Problem Statement

Online supergiant store, 'Global Mart' has worldwide operations - taking orders and delivering across the globe. Its major product categories are Consumer, Corporate and Home Office. As a Sales Manager, we are expected to forecast the sales of products for the next 6 months - this will help managing inventory and business processes accordingly - for the combination of Market and Segment which is profitable with least variation in Profits.

# Objective

- Find the most profitable and consistent market segment for the company.
- Forecast the sales and demand for the next 6 months .

# ANALYSIS

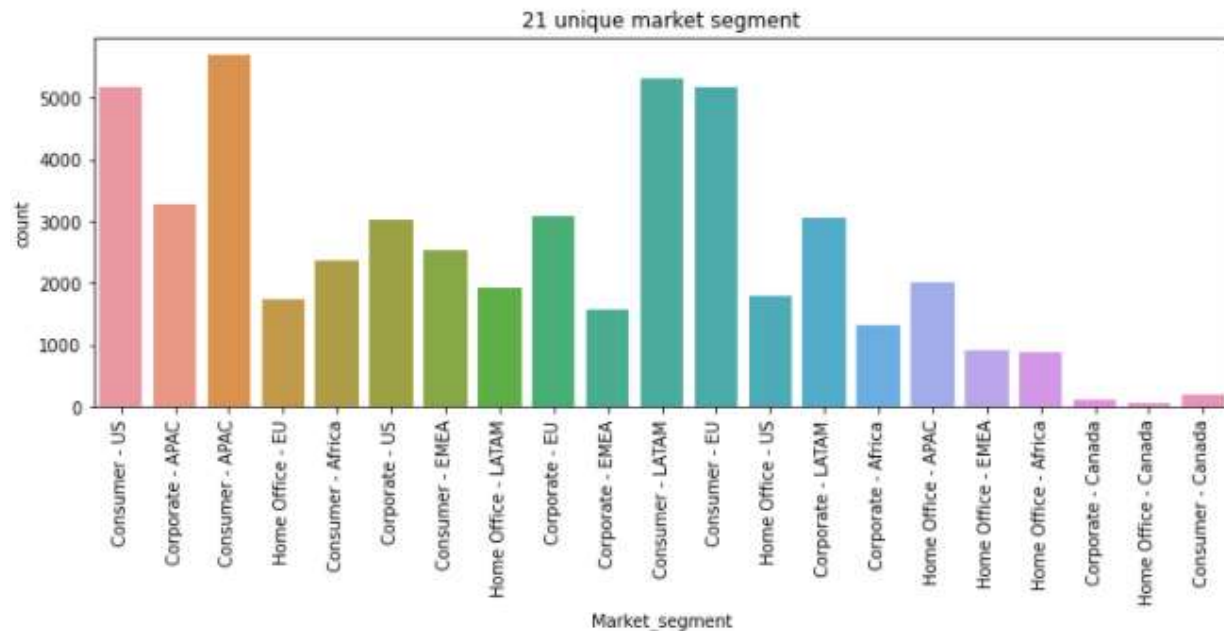


- The data seems to be well spread with 3 Segments and 7 Markets.
- Segment 'Consumer' and Market 'APAC' seem to be the leading categories.

# Market Segment

|                      |      |
|----------------------|------|
| Consumer - APAC      | 5699 |
| Consumer - LATAM     | 5321 |
| Consumer - US        | 5191 |
| Consumer - EU        | 5186 |
| Corporate - APAC     | 3283 |
| Corporate - EU       | 3077 |
| Corporate - LATAM    | 3053 |
| Corporate - US       | 3020 |
| Consumer - EMEA      | 2538 |
| Consumer - Africa    | 2381 |
| Home Office - APAC   | 2020 |
| Home Office - LATAM  | 1920 |
| Home Office - US     | 1783 |
| Home Office - EU     | 1737 |
| Corporate - EMEA     | 1574 |
| Corporate - Africa   | 1312 |
| Home Office - EMEA   | 917  |
| Home Office - Africa | 894  |
| Consumer - Canada    | 202  |
| Corporate - Canada   | 110  |
| Home Office - Canada | 72   |

Name: Market\_segment, dtype: int64



We observe that there are a total of 21 Market-Segments with all the combinations of Market and Segment. The highest number of transactions are from APAC-Consumer, followed by LATAM-Consumer, US-Consumer and EU-Consumer. Canada seems to have the least number of transactions across Segments.

# Comparison using CoV

We compare the variance between the segments using the coefficient of variation which will normalize the standard deviation with the mean and give a comparative figure on the basis of which we can identify the most profitable market segment

|    | Market_segment       | Mean    | Std     | CoV      |
|----|----------------------|---------|---------|----------|
| 0  | Consumer - APAC      | 4400.89 | 2300.46 | 0.522726 |
| 7  | Corporate - APAC     | 2574.92 | 1364.84 | 0.530051 |
| 4  | Consumer - EU        | 3699.98 | 2202.28 | 0.595214 |
| 5  | Consumer - LATAM     | 2295.56 | 1569.63 | 0.683768 |
| 11 | Corporate - EU       | 2216.30 | 1600.34 | 0.722077 |
| 12 | Corporate - LATAM    | 1122.63 | 990.36  | 0.882178 |
| 18 | Home Office - EU     | 1224.46 | 1148.63 | 0.938071 |
| 14 | Home Office - APAC   | 1511.09 | 1523.51 | 1.008219 |
| 6  | Consumer - US        | 2686.74 | 2715.03 | 1.010529 |
| 13 | Corporate - US       | 1754.20 | 1880.20 | 1.071828 |
| 20 | Home Office - US     | 1132.07 | 1272.48 | 1.124029 |
| 19 | Home Office - LATAM  | 818.40  | 957.28  | 1.169697 |
| 2  | Consumer - Canada    | 225.99  | 282.56  | 1.250321 |
| 1  | Consumer - Africa    | 957.71  | 1254.93 | 1.310344 |
| 9  | Corporate - Canada   | 90.98   | 162.49  | 1.785997 |
| 8  | Corporate - Africa   | 412.62  | 780.57  | 1.891741 |
| 15 | Home Office - Africa | 377.22  | 759.32  | 2.012937 |
| 16 | Home Office - Canada | 118.00  | 279.63  | 2.369746 |
| 3  | Consumer - EMEA      | 423.96  | 1124.55 | 2.652491 |
| 10 | Corporate - EMEA     | 182.64  | 1160.70 | 6.355125 |
| 17 | Home Office - EMEA   | 84.23   | 651.28  | 7.732162 |

# Market Selection

Considering we are focussing on only one Market-Segment which is profitable with least variation in Profits, we chose **APAC-Consumer**. The reason we prefer lowest CoV is because we want to estimate sales for the most consistently profitable Market-Segment, so that our sales forecasts are reliable.

```
: Lowest_CoV_value =min(CoV_data["CoV"])\nLowest_CoV_value
```

```
: 0.5227260849509985
```

- Lowest Cov value is of Consumer - APAC

```
: highest_CoV_value =max(CoV_data["CoV"])\nhighest_CoV_value
```

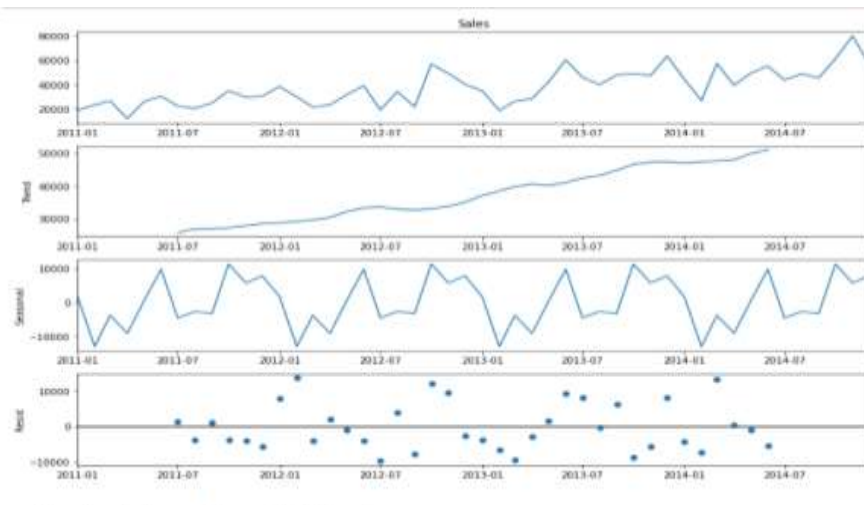
```
: 7.732161937551941
```

- highest CoV values is of Home Office -EMEA

# Decomposition

- Number of data points is more than 10, hence we will utilize Exponential Smoothing or ARIMA and we should NOT utilize Simple Moving Average or Naive Methods.
- We observe that there is an upward trend in the dataset
- We also observe that there is Seasonality (at 12 months) in the dataset.
- **Exponential Smoothing:** We should use Simple Exponential Smoothing, Holt's Exponential Smoothing, Holt Winter's Smoothing techniques among others.
- We should also look at **ARIMA:** There is an Upward Trend in the dataset, we can use ARIMA techniques. As we can observe there is seasonality as well, we can use SARIMA. However, we don't have any information on Exogenous variables, hence we would NOT use ARIMAX or SARIMAX.

## Additive Method



## Multiplicative Method

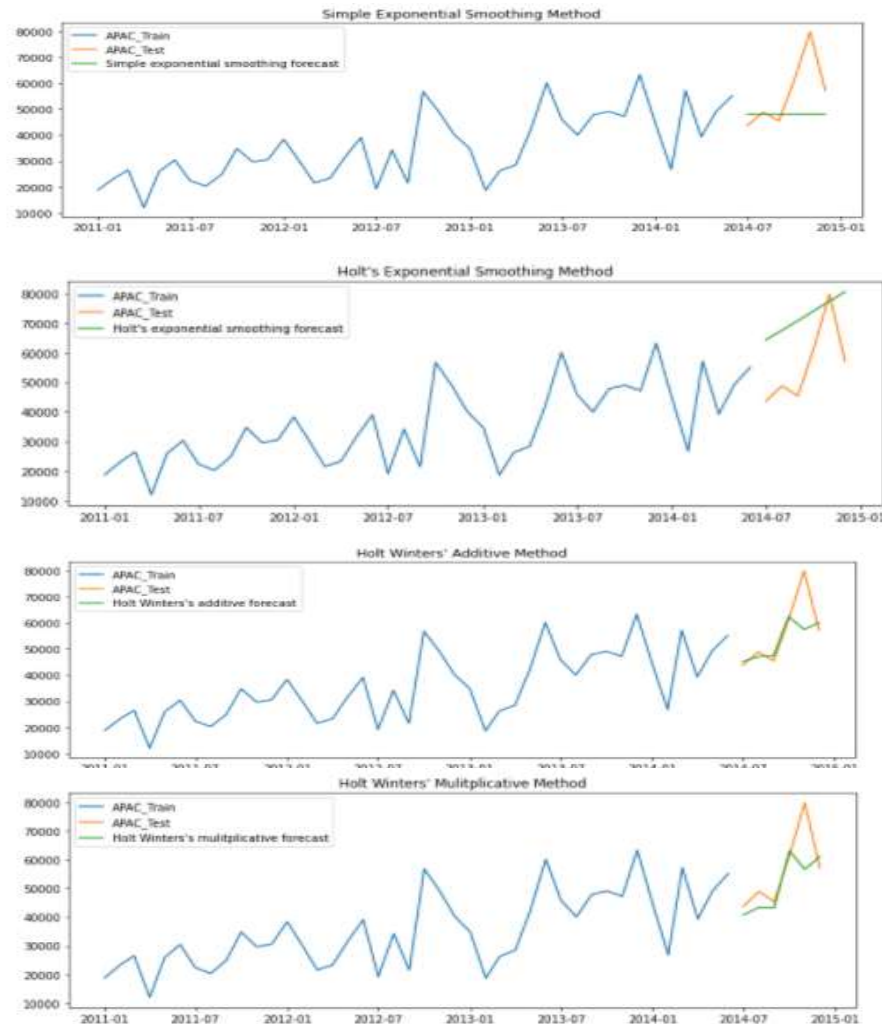




# Exponential Smoothing

|   | Method                                | RMSE     | MAPE  |
|---|---------------------------------------|----------|-------|
| 0 | Simple exponential smoothing forecast | 14784.66 | 15.83 |
| 0 | Holt's exponential smoothing method   | 18976.37 | 34.57 |
| 0 | Holt Winters' additive method         | 9304.48  | 7.76  |
| 0 | Holt Winters' multiplicative method   | 10021.79 | 10.30 |

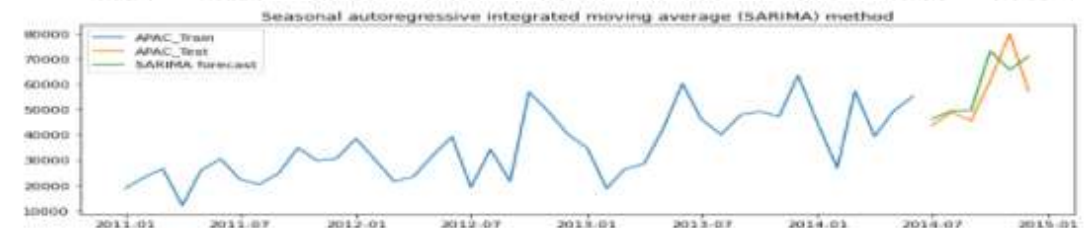
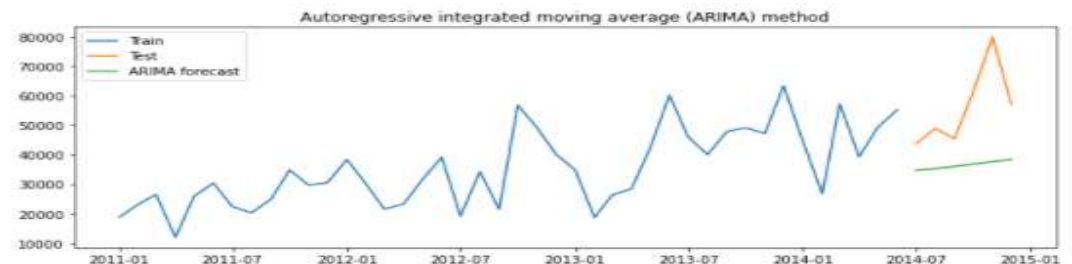
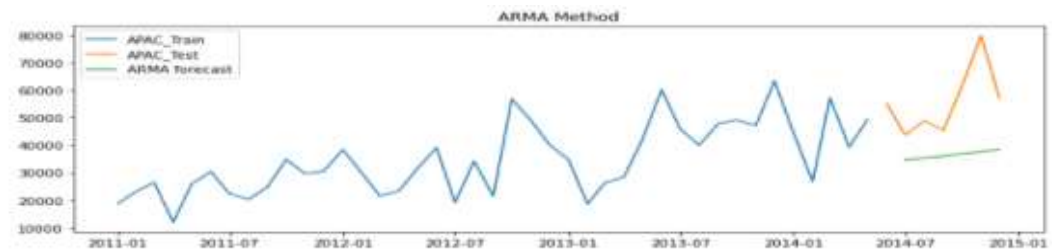
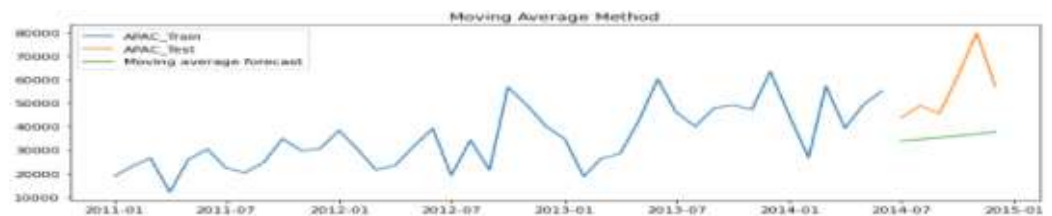
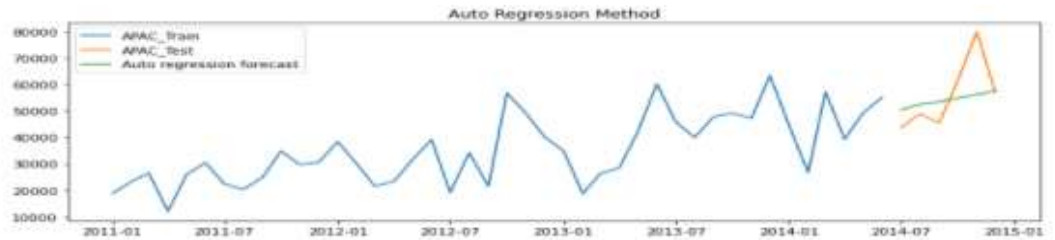
Overall, among the 4 Smoothing techniques, Holt Winter's Additive Method has the least MAPE and hence, the best Smoothing technique for the given Time-Series data.



# ARIMA

|   | Method  | RMSE     | MAPE  |
|---|---|----------|-------|
| 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... | 9612.36  | 12.86 |

Overall, among the 5 ARIMA techniques, SARIMA Forecasting has the least MAPE and hence, the best Forecasting technique for the give Time-Series data



# Conclusion

Overall, the best forecasting methods for the given case study are as below (based on their MAPE values) :-

- For Smoothing method, the best option is Holt Winter's Additive Method
- For Auto-Regressive methods, the best option is SARIMA

Thank  
you