Retail Giant Sales Forecasting Assignment

Submitted by:

Abhinav Choudhary (PGDDS C-24)

Business Objective & 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 of this store, we have to identify the most consistently profitable market-segment & forecast the sales of the products for the next 6 months, so that we have a proper estimate and can plan inventory and business processes accordingly.

Analysis Approach

Understanding the Data

- Import & understand the data at a high level

Data Preparation & CoV Analysis

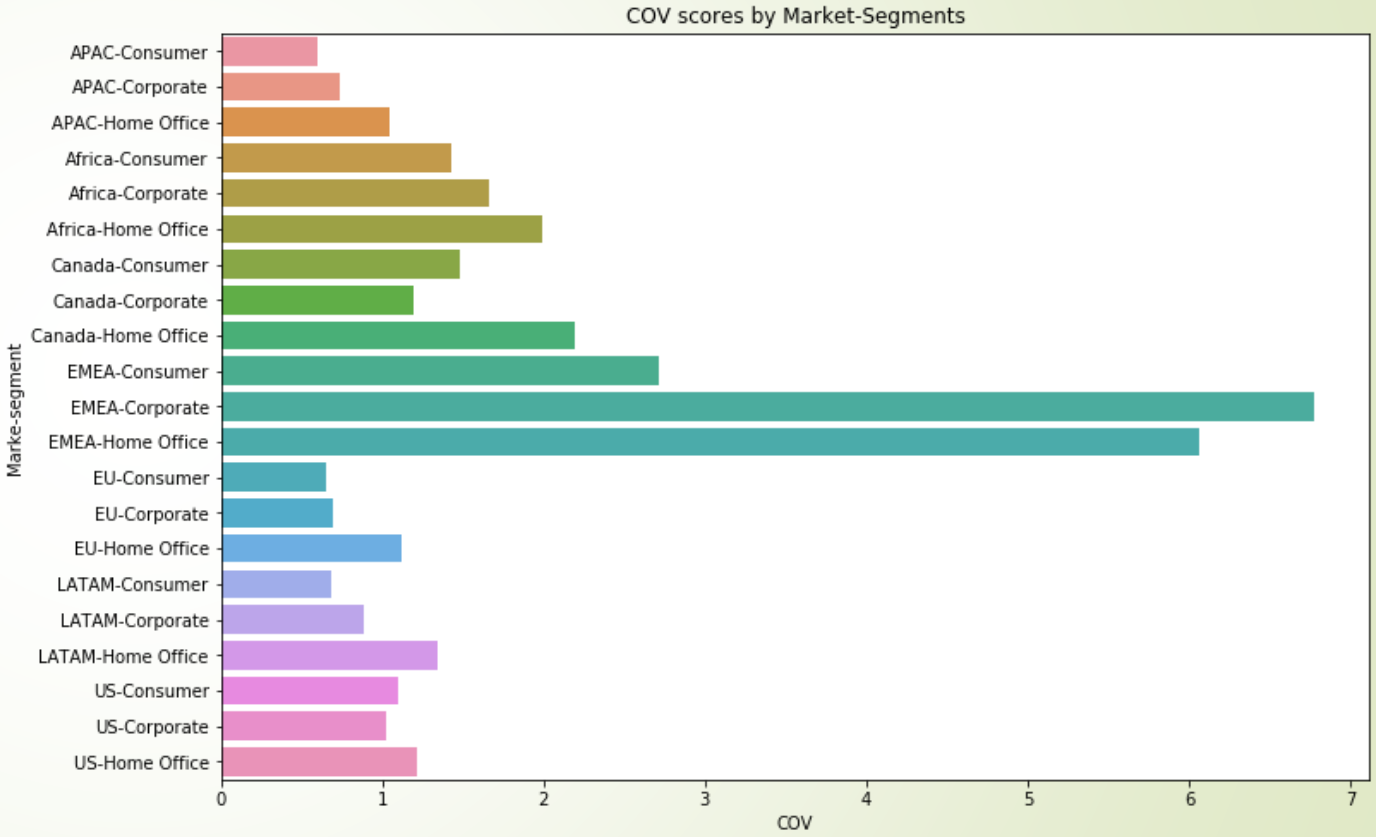
- Deriving required features for CoV analysis
- Aggregating the data & Train-Test Split
- Calculating CoV for each of the Market segment.
- Finding the Market Segment with least CoV.

Forecasting Model Building and Evaluation

- Aggregation & Train-Test split
- Additive & Multiplicative Seasonal decomposition of the dataset
- Understanding the appropriate method from the flow chart
- Building & plotting various forecasting models from Smoothing techniques
- Stationary Test, Box Cox transformation & Differencing
- Building & plotting various Auto Regressive forecasting models
- Model Evaluation by computing & comparing the RMSE & MAPE scores of all the forecasting models.

CoV Analysis
(to find the most consistently profitable market-segment)

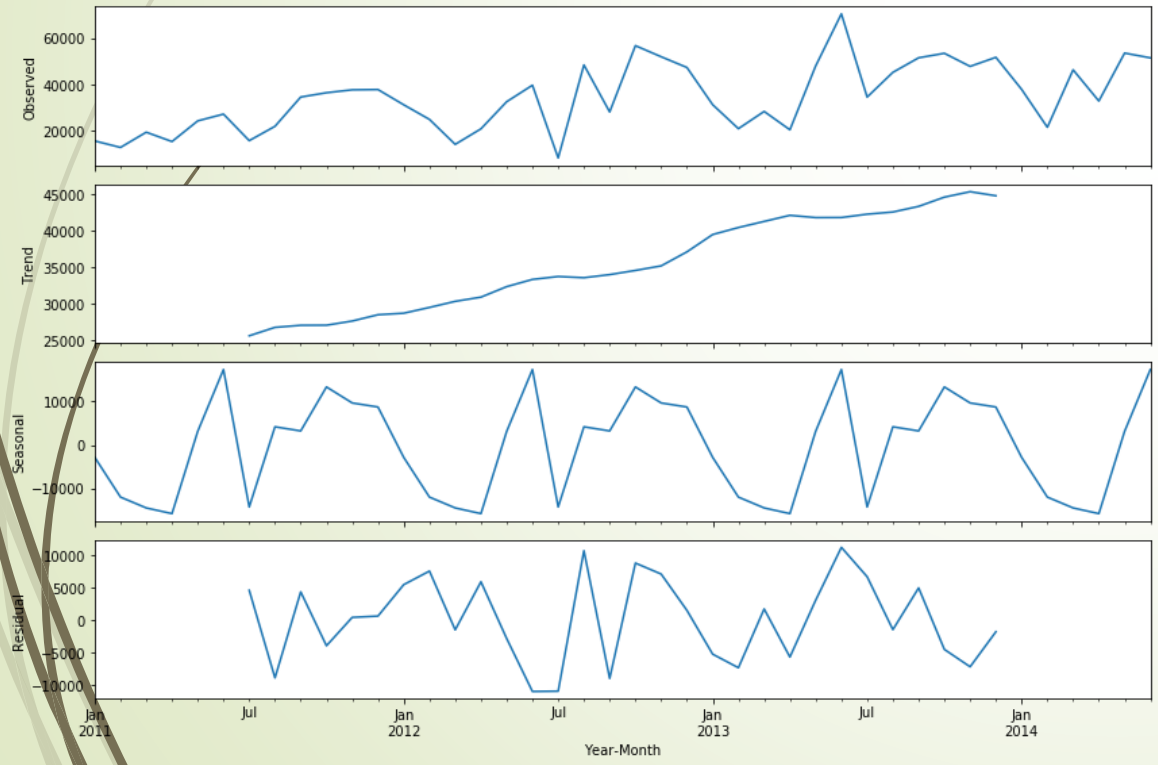
- To find the most consistently profitable market-segment we have leveraged a measure called "Coefficient of Variation (CoV)". The coefficient of variation or CoV is nothing but the ratio of the standard deviation to mean for the data that it is being calculated for.
- Lower the CoV value, lower is the fluctuations & variation in the dataset.
- Computed the CoV for all the 21 Market Segments(refer to the barplot in right) and found that 'APAC-Consumer' has the least CoV value. **Therefore, we can conclude that 'APAC-Consumer' is the most consistently profitable market-segment.**



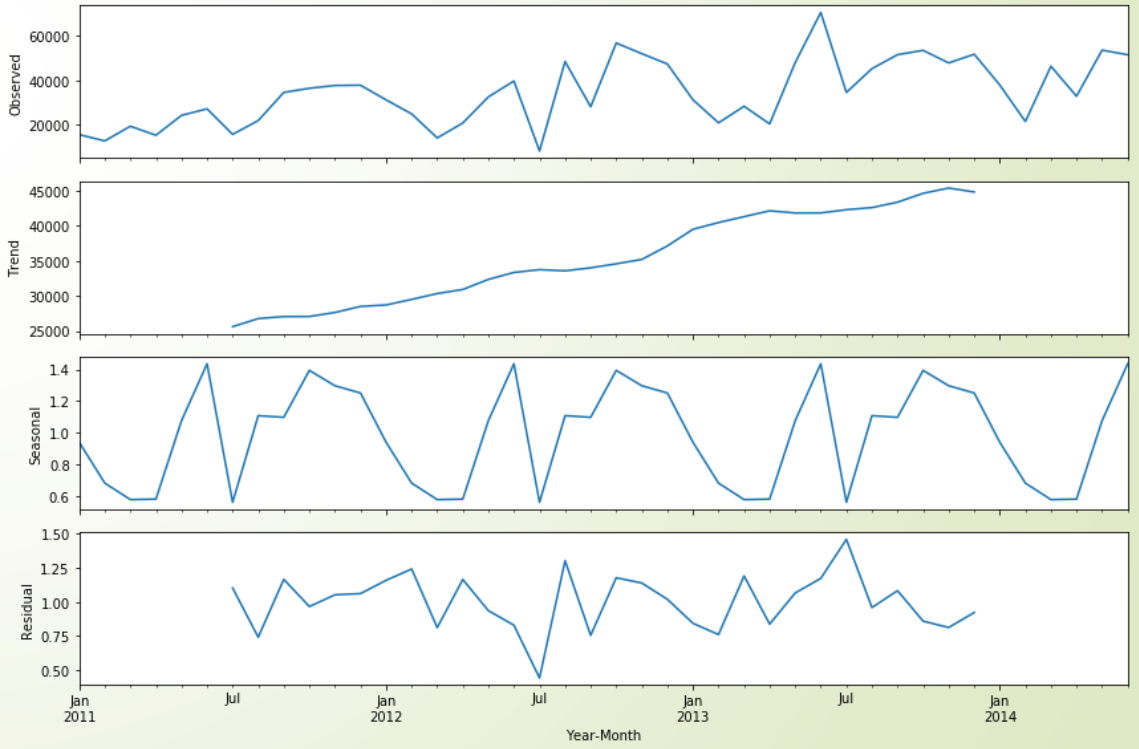
Seasonal Decomposition – Additive & Multiplicative

- After selecting most consistently profitable market-segment, performed the additive & multiplicative decomposition of the dataset.
- From the decomposition plots, we understood that the monthly sales has an upward trend and an annual seasonality which repeats after every 12 months.

Additive Seasonal Decomposition

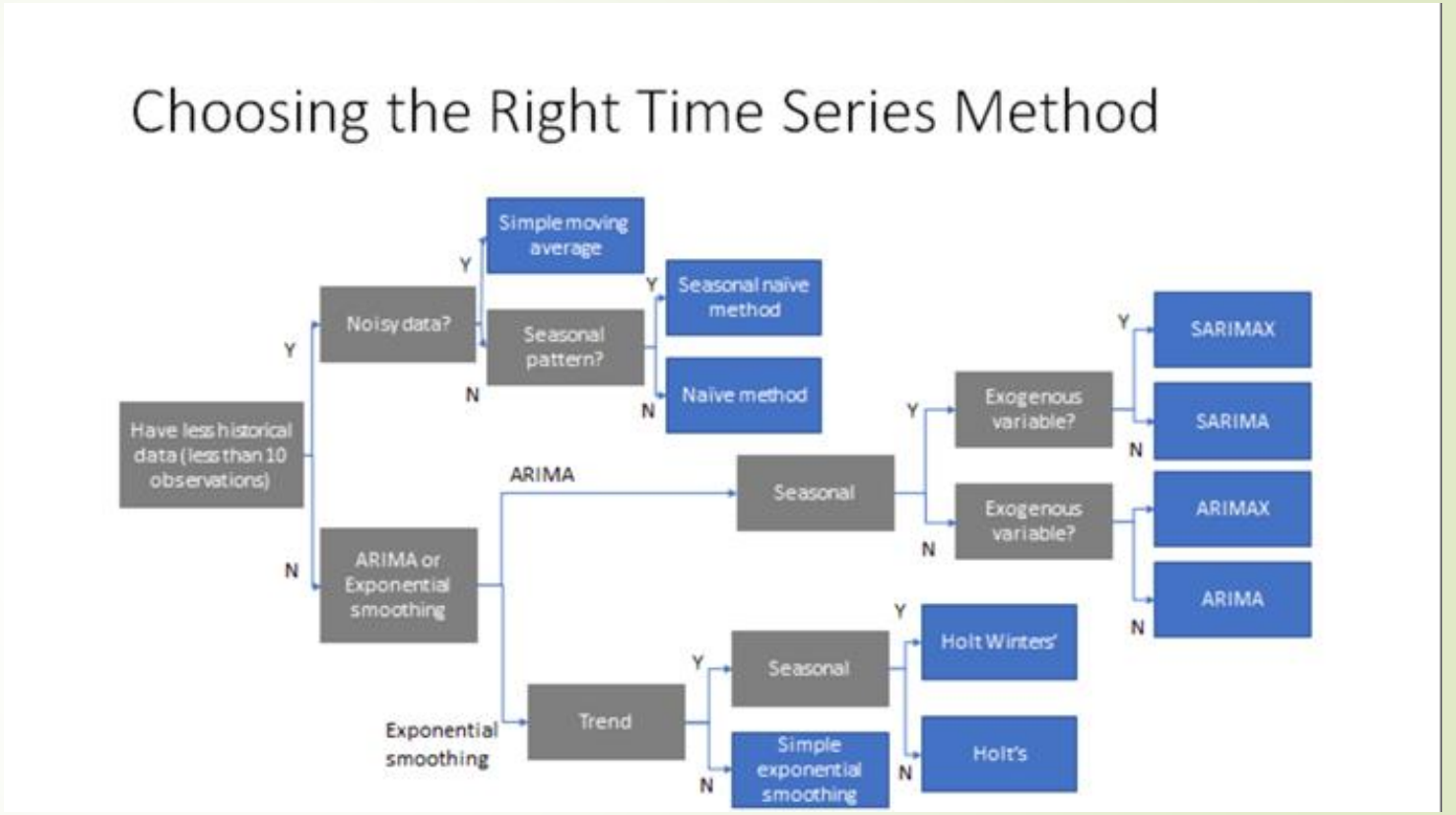


Multiplicative Seasonal Decomposition



Forecasting Model Selection Flow Chart

- with the help of the model selection flow chart, we understood what all Smoothing & ARIMA set of forecasting models would be best suited for our dataset.
- We have more than 10 observations in out historical(train) dataset, so would go and select models from the bottom half of this flow chart.
- Our dataset has both trend & seasonality. So we are going to select models accordingly from the flow chart.
- From the Exponential Smoothing set of models, the best suited forecasting model would be - **"Holt Winters"**
- From the ARIMA set of models, the best suited forecasting model would be - **"SARIMA"**.



Forecasting Models: Smoothing Techniques

- Of all the Smoothing techniques, we can conclude that Holt Winters' multiplicative method has performed the best with the least RMSE & MAPE scores.
- This model is the same as what we had found out initially with the help of model selection flow chart.

	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	23383.65	28.15
0	Simple exponential smoothing forecast	22991.87	27.73
0	Holt's exponential smoothing method	17204.15	24.99
0	Holt Winters' additive method	12971.01	17.61
0	Holt Winters' multiplicative method	11753.42	19.62

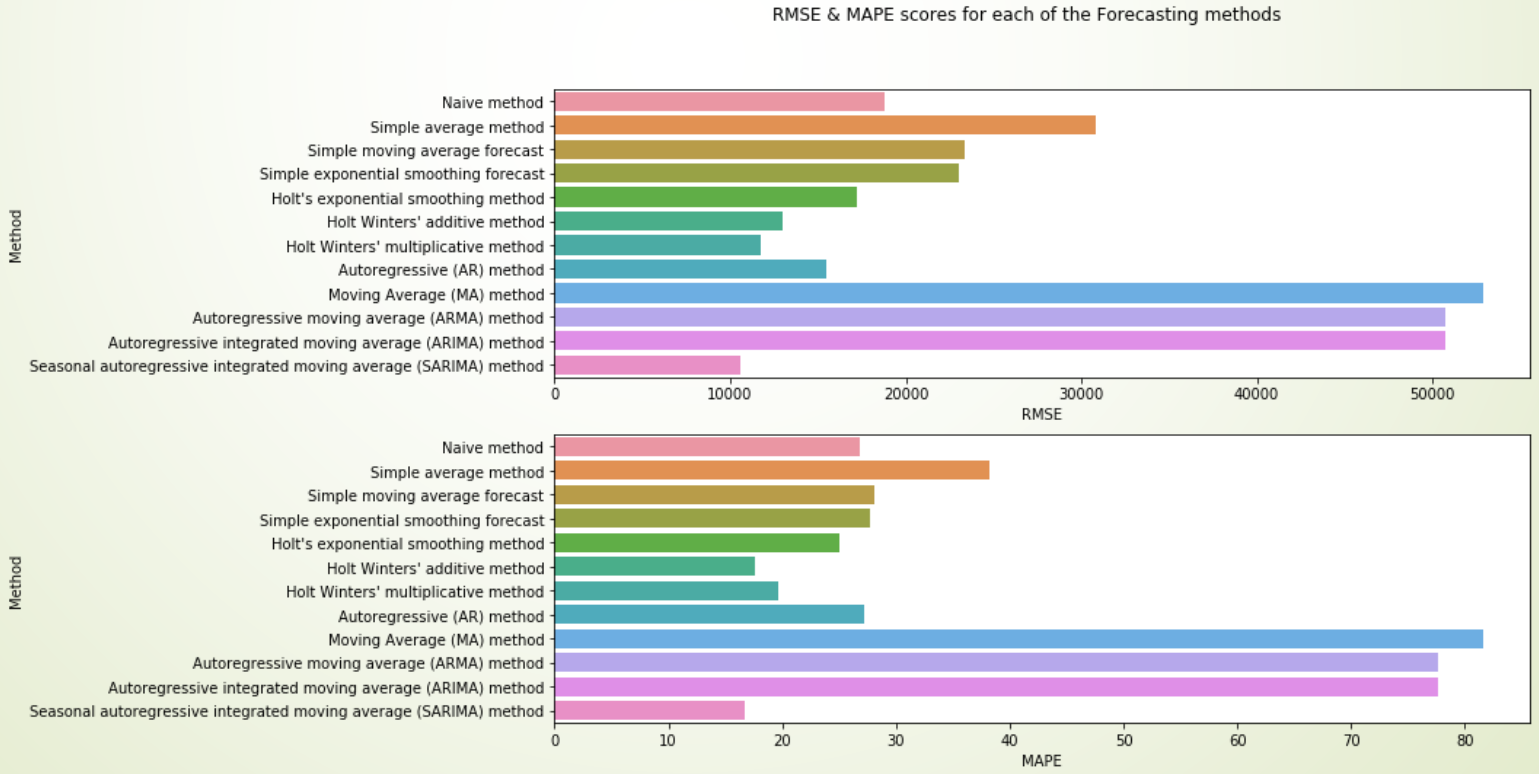
Forecasting Models: Auto Regression Methods

- Of all the Auto Regressive forecasting techniques, we can conclude that Seasonal Auto Regressive Integrated Moving Average (SARIMA) method has performed the best with the least RMSE & MAPE scores.
- This model is the same as what we had found out initially with the help of model selection flow chart.

	Method	RMSE	MAPE
0	Autoregressive (AR) method	15505.02	27.27
0	Moving Average (MA) method	52903.35	81.64
0	Autoregressive moving average (ARMA) method	50757.93	77.66
0	Autoregressive integrated moving average (ARIM...	50757.93	77.66
0	Seasonal autoregressive integrated moving aver...	10568.99	16.77

Summary & Key Takeaway

- Overall, comparing all the above 12 models(Smoothing & Auto Regressive) that we have built & evaluated so far & looking at their respective RMSE & MAPE scores, we can conclude that Seasonal Auto Regressive Integrated Moving Average (SARIMA) has performed the best in forecasting the future Sales.
- The sales for APAC-Consumer market segment is likely to rise in the coming 6 months. Hence, it is recommended to invest more in this market segment.



Thank You! ☺