Global Mart Case Study

SUBMISSION

Demand and Sales Forecasting for a Global Retail Chain using Time Series Algorithm

Group Name: **Code Ninjas**

1. Alvin Mark Windsor
2. Uday Balerao
3. Anand Agarwal
4. Ravi Kiran Vissa

1. Case Study Overview

CONTEXT

GLOBAL MART is an online retail giant having worldwide operations in 147 countries grouped into 7 Global Market Regions. It's customer base is of 3 major segments-consumer, corporate and home office. GLOBAL MART deals with commodities from 3 major product categories- technology, furniture and office supplies

PROBLEM

For a store operating at such a mass scale, Planning, Operations and Logistics becomes a monumental task.

The Sales/Operations department requires a finalized plan of the forecasted Demand and Sales for the next six months for the target profit Market Buckets.

This forecast would help them manage the revenue and inventory accordingly.

OBJECTIVE AND DELIVERABLES

[1] Subset the data from the transactional database to form the 21 Market Buckets [7 Global Market Regions x 3 Customer Segments]. Determine the Monthly Sales, Quantity and Profit with reference to Order Date.

[2] Identify the two most profitable and consistently profitable Market Buckets.

[3] Build a time series auto ARIMA model and classical decomposition model on sales and demand for the aforementioned Market Buckets.

[4] Evaluate the models and select the best model to forecast sales and demand for each Market Bucket. Evaluation metric considered is MAPE.

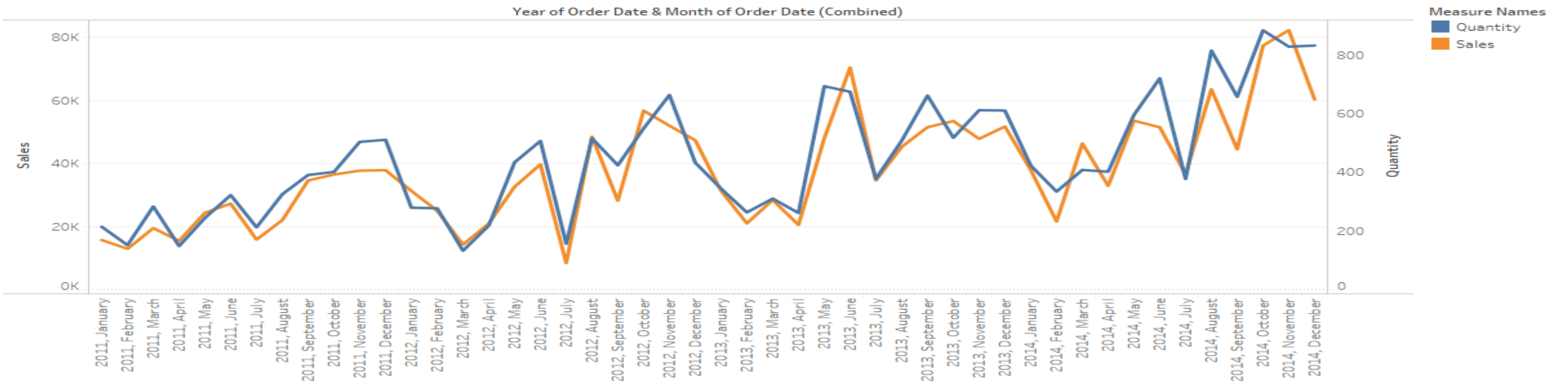
[5] Use the final models to predict the Demand and Sales of the next 6 months for the 2 Market Buckets.

2. Problem Solving Methodology

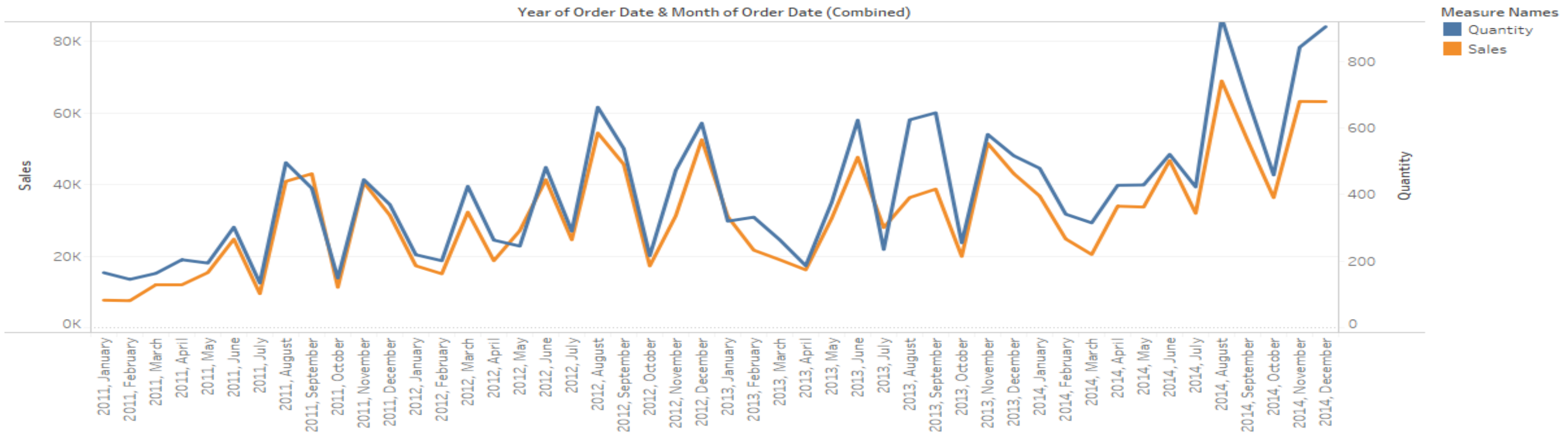
- **Business Understanding** — Understand the business context of the problem by referring online material on trends and seasonality for online retail stores.
- **Data Understanding** — Assimilating the variables available in the data file and understanding them through data dictionary and basic statistics.
  We have excluded visual plots of EDA since the data given to us is a transactional data which has trends and seasonality. This data is subsetted and aggregated into a format suitable for modelling.
- **Data Preparation** — Data cleaning, treating missing values and **Feature Engineering.**
  Finding the monthly Profit, Sales, Demand for all the 21 Market segments. Then computing the Net Profit, Monthly Average Profit and Coefficient of Variance[Profit] for each Market bucket.
  Identifying the two target Market Buckets.
- **Model Building Auto ARIMA and Classical Decomposition** — Using 42 months of data to create auto ARIMA model and classical decomposition model for Sales and for demand of each Market Bucket.
- **Model Evaluation** — Forecasting the 6 months of test data using auto ARIMA model and classical decomposition model created above. Then evaluating these models using MAPE matrix to chose the best one.
- **Recommendations** — The best model chosen will be used to Forecast the next six months demand and sales in each target Market Bucket.

3. Business Overview

Demand and Sales for Consumer Segment of APAC Market



Demand and Sales for Consumer Segment of EU Market



The graphs shown above represent the demand and sales for the target Market Buckets:

[1] The Consumer Segment of APAC Market [2] The Consumer Segment of EU Market

It is clear from the visualization that there are trend and seasonality fluctuations that influence the Demand and Sales. A simple Naïve Forecasting , Moving Average Forecasting or Exponential Forecasting would not give effective estimates of future demand and sales [This may directly impact GLOBAL MART as missed profit or missed opportunity]. In such situations a well tuned time series model would be ideal to forecast future Demand and Sales.

4. Data Understanding

The data set consists of order details made through Global Mart across the world. The orders are classified into 3 major customer segments - consumer, corporate & home office These come under 7 major market segments.

The dataset has been split into 21 market subsets each containing the following attributes:

[1] Month.Code – Number of months passed Jan 2011.

[2] Monthly.Profit – Sum of Profit for each month for the market segment.

[3] Monthly.Sales – Sum of Sales for each month for the market segment.

[4] Monthly.Demand – Sum of Quantity for each month for the market segment

[5] Net.Profit – Total Profit for the market segment

[6] Average.Profit – Average of monthly profit for the market segment

[7] Coeff. of Variation – Coefficient of variation of profit for the market segment

5. Assumptions and Data Handling

[1] Order Date - Order Date is used as our reference point as it is recorded at Point of Sale for GLOBAL MART. It is clear from the data that we have 48 months of transactional data ranging from 1st Jan ,2011 to 31st Dec, 2014. Hence we have **derived a new metric** called as **Month.Code** which basically determines the number of months passed from Jan, 2011.

Jan 2011 → 1
Feb 2011 → 2
Mar 2011 → 3 and so on.

[2] Duplication Checks - Data Duplication checks have been performed.

[3] Missing Value Treatment - There are no missing values found for Sales, Quantity and Profit columns. Missing values are only present in Postal Code Attribute for all countries except **United States**. Hence we will not perform and data imputation as Postal Code will not influence our model.

[4] Data Preparation – We have computed the monthly Profit, demand and sales for all 21 Market Buckets. Following which we have derived Net Profit, Average Monthly Profit and Coefficient of Variance[Monthly Profit].

The coefficient of variation for each of the market segment is calculated as below:

Coefficient of variation[Profit] = sd (Monthly Profit) / mean(Monthly Profit)

Here, sd means standard deviation.

Based on the above values, we need to chose the two segments which have high Net Profit, high Average Monthly Profit and low Coefficient of variation[Profit].

[5] Custom User Defined Functions- We have defined two user defined function as follows:

[1]parameter_aggregation: It aggregates the data for each market bucket and computes the monthly metrics as stated in point#4.

[2] ts_movavg_smoother: It performs moving average smoothing for an input time series based on an input window. Smoothing is performed on both sides of the window.

[6] Hypothesis Testing for Stationary Series at 95% Confidence Interval.

adf.test(): H0 – Series is not stationary | H1 – Series is Stationary

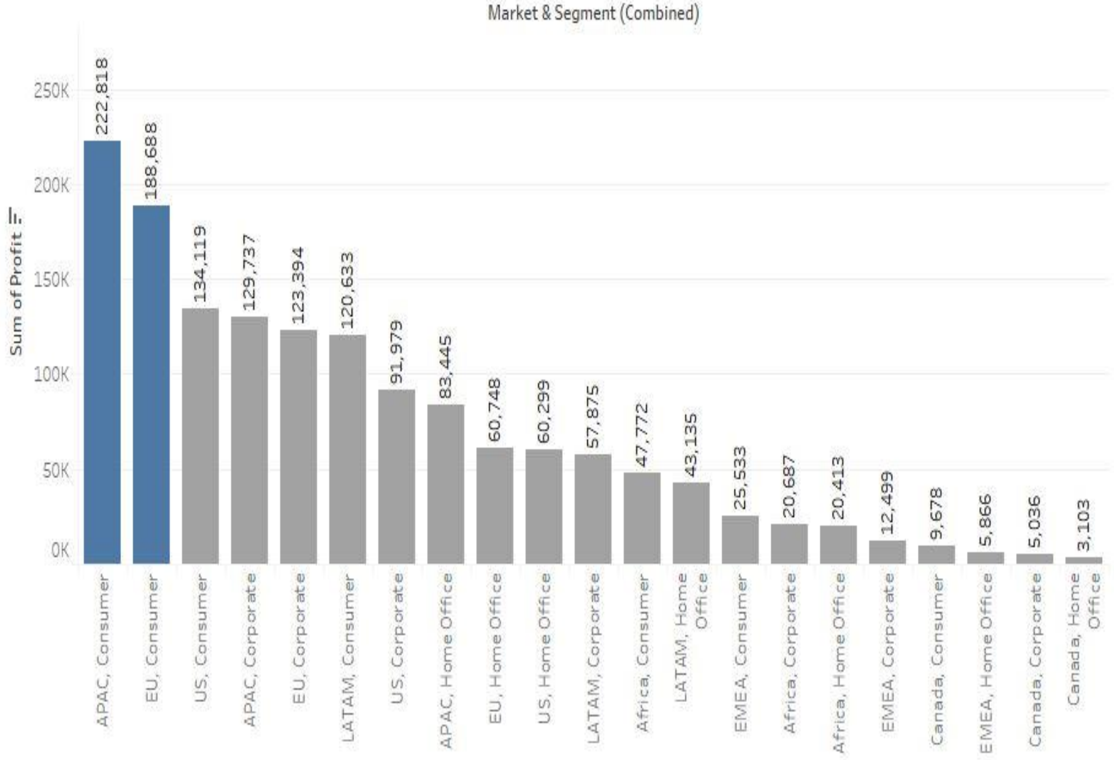
Kpss.test(): H0 – Series is stationary | H1 – Series is not Stationary

6. Finding the Top 2 Market segments

UpGrad

The graph on the left shows the Net Profit for the each market segment. The graph on the right shows the average of monthly profit and coefficient of variation (Profit) for each market segment. Considering these factor the below best two market segments are chosen (highlighted in green in the graph).

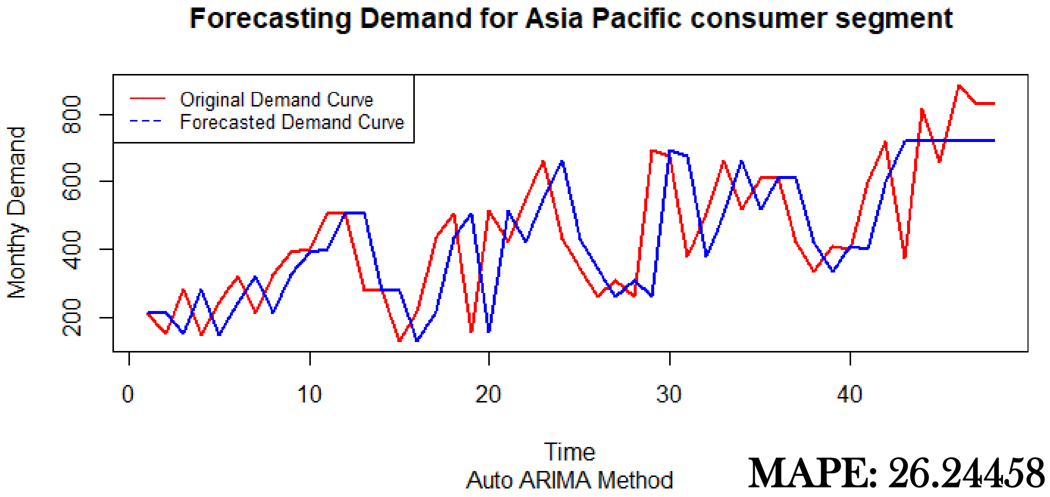
[1] The Consumer Segment of APAC Market
[2] The Consumer Segment of EU Market



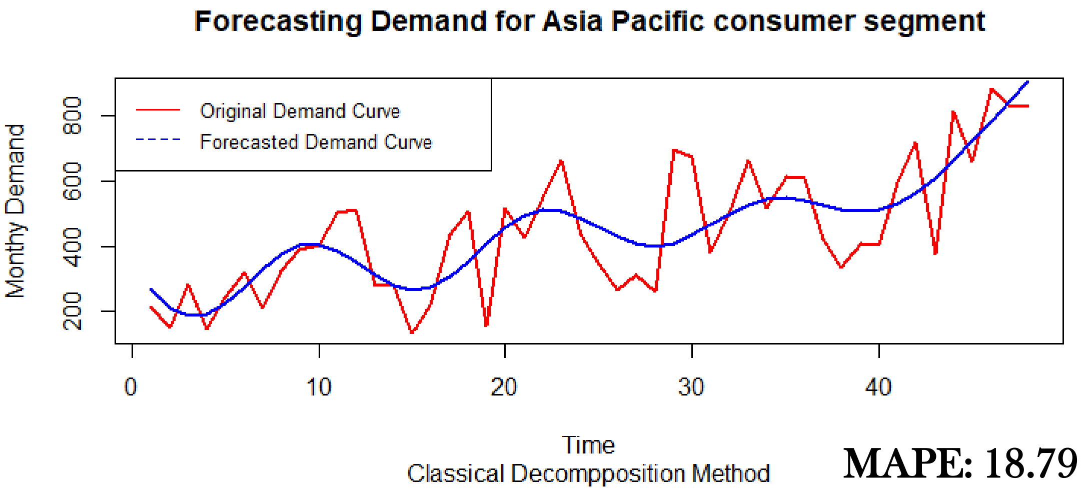
Net Profit for each market segment

Avg. of Monthly Profit and Coefficient of variation[Profit] for each market segment

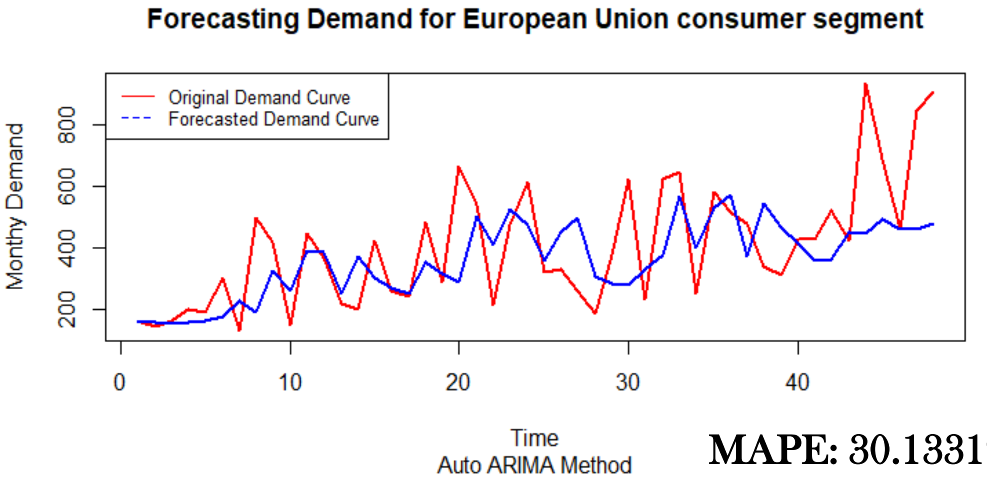
7. Forecasting Demand - Comparing Auto ARIMA and Classical Decomposition Models



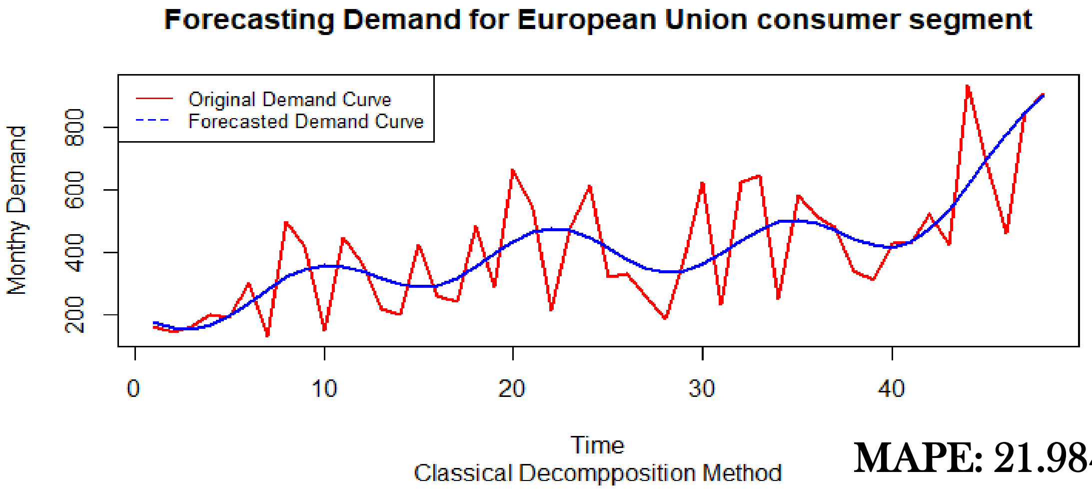
MAPE: 26.24458



MAPE: 18.79196

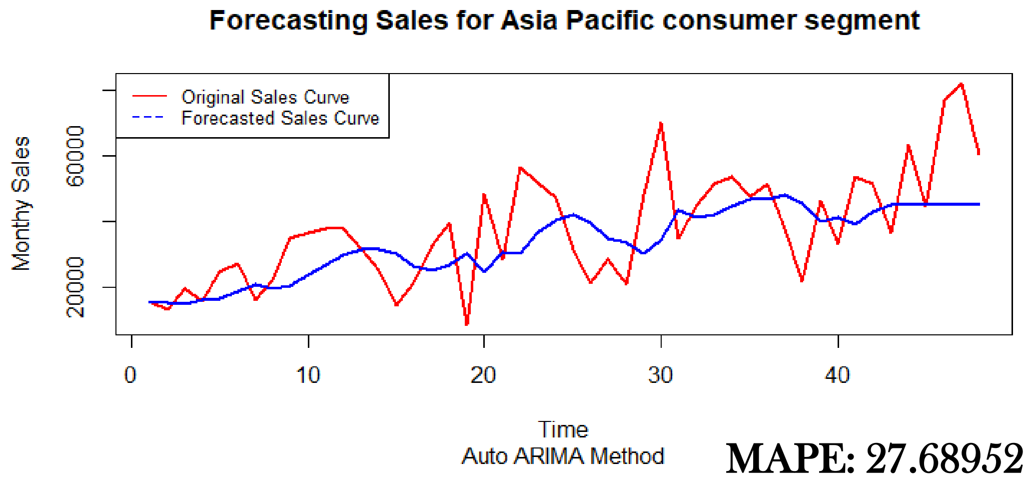


MAPE: 30.13319

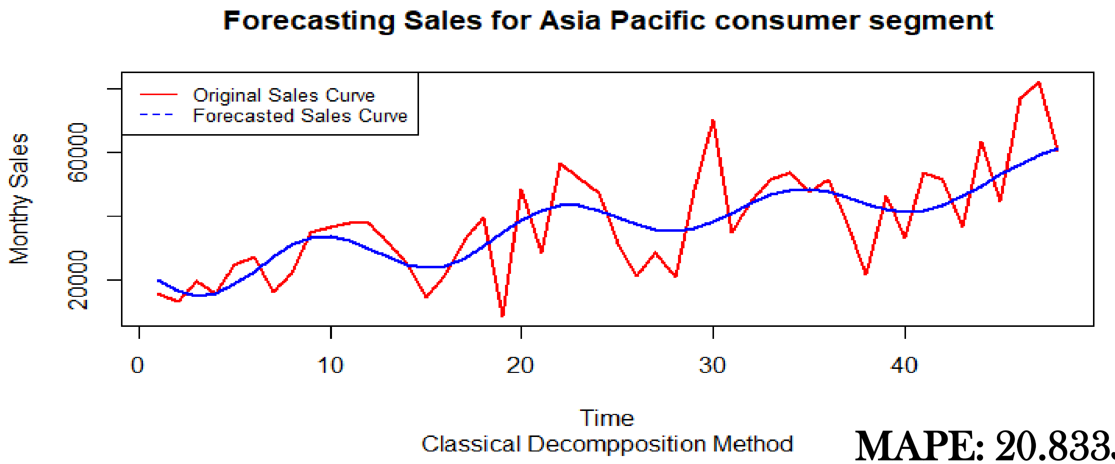


MAPE: 21.98432

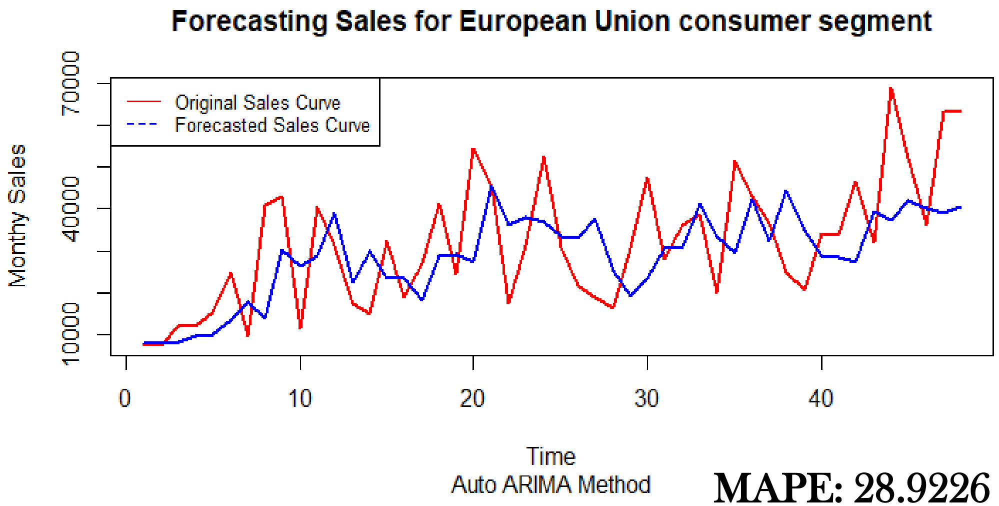
8. Forecasting Sales - Comparing Auto ARIMA and Classical Decomposition Models



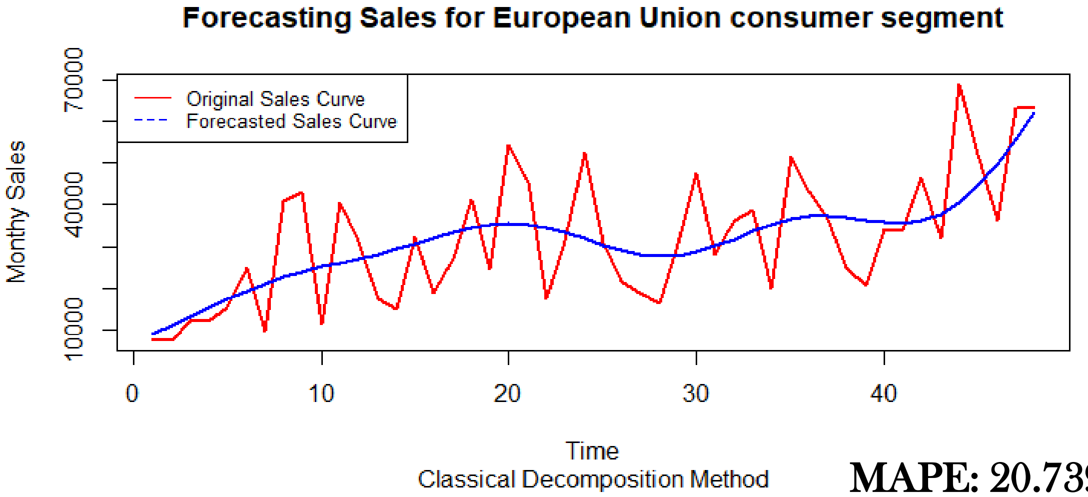
MAPE: 27.68952



MAPE: 20.83351



MAPE: 28.9226



MAPE: 20.73958

9. APAC Consumer Demand and Sales Forecasting Model Evaluation

Model	Parameter	Classical Decomposition Model	Auto Arima Model
Demand Forecasting Models APAC Consumer Market Bucket	Global Component	lm(Monthly.Demand~ sin(0.5*Month.Code)*poly(Month.Code,1)+ cos(0.1*Month.Code)*poly(Month.Code,1)+ tan(0.02*Month.Code), data = smothed_demand.apac_consmr)	ARIMA(0,1,0) ; This implies that 1 stage differencing was performed. sigma^2 estimated as 25366 | log likelihood=-266.07 | AIC=534.14 | AICc=534.24 | BIC=535.85
	Local Component	ARIMA(0,0,0) with zero mean	NA
	Residual Component	Stationary	Stationary
	ADF TEST [Residual]	The Augmented Dickey-Fuller Test shows a p-value<0.01	The Augmented Dickey-Fuller Test shows a p-value<0.01
	KPSS TEST [Residual]	The KPSS Test for Level Stationarity shows a p-value>0.1	The KPSS Test for Level Stationarity shows a p-value>0.1
	Test Forecast RMSE	127.49	174.37
	Test Forecast MAPE [%]	18.79	26.24
Sales Forecasting Models APAC Consumer Market Bucket	Global Component	lm(Monthly.Sales~ sin(0.5*Month.Code)*poly(Month.Code,1)+ cos(0.05*Month.Code)*poly(Month.Code,1)+ tan(0.02*Month.Code), data = smothed_sales.apac_consmr)	ARIMA(0,1,1). This implies that 1 stage differencing was performed and the resulting timeseries was modeled as MA(1). ma1 Coeff: -0.7559 | Std.Error. 0.1381 sigma^2 estimated as 174361546 | log likelihood=-447.11 | AIC=898.23 | AICc=898.55 | BIC=901.66
	Local Component	ARIMA(0,0,0) with zero mean	
	Residual Component	Stationary	Stationary
	ADF TEST [Residual]	The Augmented Dickey-Fuller Test shows a p-value<0.01	The Augmented Dickey-Fuller Test shows a p-value<0.01
	KPSS TEST [Residual]	The KPSS Test for Level Stationarity shows a p-value>0.1	The KPSS Test for Level Stationarity shows a p-value>0.1
	Test Forecast RMSE	14997.64	22755.75
	Test Forecast MAPE [%]	20.83	27.69

Insights

- APAC-CONSUMER MARKET BUCKET [DEMAND Model]-From the above results it is clear that the Classical Decomposition Forecasting model performs better than the Auto Arima model. It provides a MAPE reduction of 7.45% and a Root Mean Square Error reduction of 26.8% in comparison to the Auto Arima model.
- APAC-CONSUMER MARKET BUCKET [SALES Model]- From the above results it is clear that the Classical Decomposition Forecasting model performs better than the Auto Arima model. It provides a MAPE reduction of 6.86% and a Root Mean Square Error reduction of 34.1% in comparison to the Auto Arima model.

10. EU Consumer Demand and Sales Forecasting Model Evaluation

Model	Parameter	Classical Decomposition Model	Auto Arima Model
Demand Forecasting Models EU Consumer Market Bucket	Global Component	lm(Monthly.Demand~ sin(0.5*Month.Code)*poly(Month.Code,1)+ cos(0.09*Month.Code)*poly(Month.Code,1)+ tan(0.02*Month.Code), data = smothed_demand.eu_consmr)	ARIMA(2,1,0) ; This implies that 1 stage differencing was performed. Following which it was modeled as an AR(2) model. ar1 Coeff: -0.7359 | Std. Error: 0.1224 | ar2 Coeff: -0.5879 | Std. Error: 0.1185. sigma^2 estimated as 21185 | log likelihood=-261.9 | AIC=529.8 | AICc=530.44 | BIC=534.94
	Local Component	ARIMA(0,0,0) with zero mean	
	Residual Component	Stationary	Stationary
	ADF TEST [Residual]	The Augmented Dickey-Fuller Test shows a p-value=0.02453	The Augmented Dickey-Fuller Test shows a p-value=0.04521
	KPSS TEST [Residual]	The KPSS Test for Level Stationarity shows a p-value>0.1	The KPSS Test for Level Stationarity shows a p-value>0.1
	Test Forecast RMSE	189.32	316.76
	Test Forecast MAPE [%]	21.98	30.13
Sales Forecasting Models EU Consumer Market Bucket	Global Component	lm(Monthly.Sales~ sin(0.4*Month.Code)*poly(Month.Code,1)+ cos(0.09*Month.Code)*poly(Month.Code,1), data = smothed_sales.eu_consmr)	ARIMA(2,1,0) ; This implies that 1 stage differencing was performed and the resulting timeseries was modeled as AR(2). ar1 Coeff: -0.5796 | Std.Error: 0.1346 | ar2 Coeff: -0.4906 | Std.Error: 0.1310. sigma^2 estimated as 168564657 | log likelihood=-445.84 | AIC=897.67 | AICc=898.32 | BIC=902.81
	Local Component	ARIMA(0,0,0) with zero mean	
	Residual Component	Stationary	Stationary
	ADF TEST [Residual]	The Augmented Dickey-Fuller Test shows a p-value=0.01311.	The Augmented Dickey-Fuller Test shows a p-value<0.01
	KPSS TEST [Residual]	The KPSS Test for Level Stationarity shows a p-value>0.1	The KPSS Test for Level Stationarity shows a p-value>0.1
	Test Forecast RMSE	13736.59	19499.13
	Test Forecast MAPE [%]	20.74	28.92

Insights

- EU-CONSUMER MARKET BUCKET [DEMAND Model]-From the above results it is clear that the Classical Decomposition Forecasting model performs better than the Auto Arima model. It provides a MAPE reduction of 8.15% and a Root Mean Square Error reduction of 40.23% in comparison to the Auto Arima model.

- EU-CONSUMER MARKET BUCKET [SALES Model]- From the above results it is clear that the Classical Decomposition Forecasting model performs better than the Auto Arima model. It provides a MAPE reduction of 8.19% and a Root Mean Square Error reduction of 29.55% in comparison to the Auto Arima model.

11. Results and Suggestions

1. From the Market Bucket Analysis it is clear that :

- [1] The Consumer Segment of APAC Market
- [2] The Consumer Segment of EU Market

Are our target Market Buckets as they are the Most Profitable and Consistently Profitable Market Buckets.

2. APAC Consumer Market Bucket

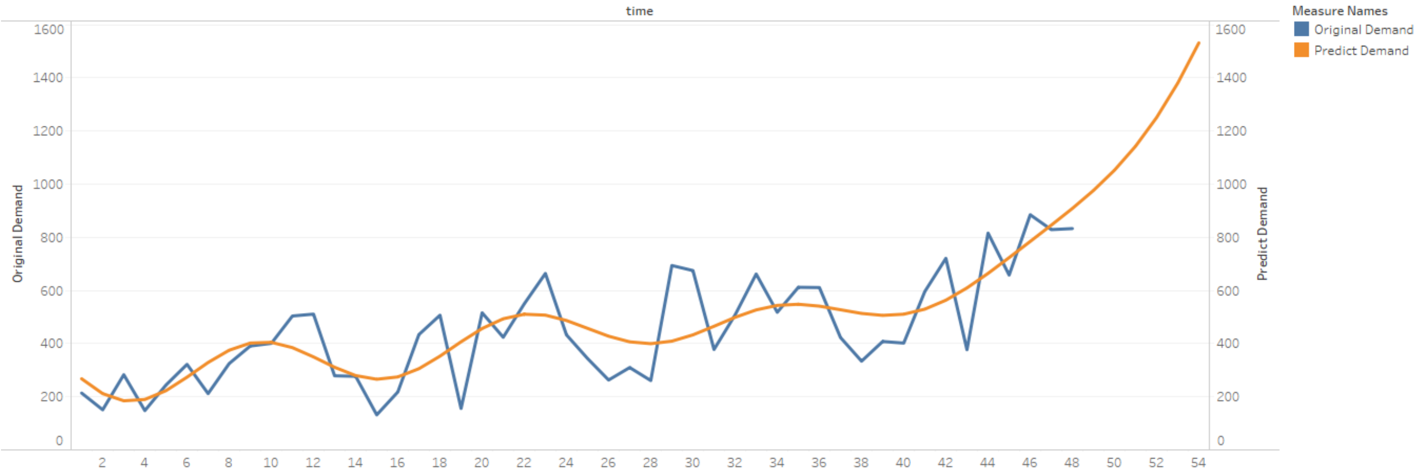
Demand Forecasting

From the model evaluation table shown in slide9. The classical decomposition forecasting model performs better than the Auto Arima model. **It provides a MAPE reduction of 7.45% and a Root Mean Square Error reduction of 26.8%** in comparison to the Auto Arima model. The graph to the right shows the Forecasted Demand vs. Actual Demand along with the Forecasted Demand for the future 6-month period.

APAC Consumer Demand Forecast	
Future Month	Demand Forecast
Jan-15	976.3252
Feb-15	1052.902
Mar-15	1143.2185
Apr-15	1251.3994
May-15	1380.3938
Jun-15	1531.8233

Avg. of Monthly Profit and Coefficient of variation[Profit] for each market segment

Demand Forecast for Asia Pacific Consumer Segment

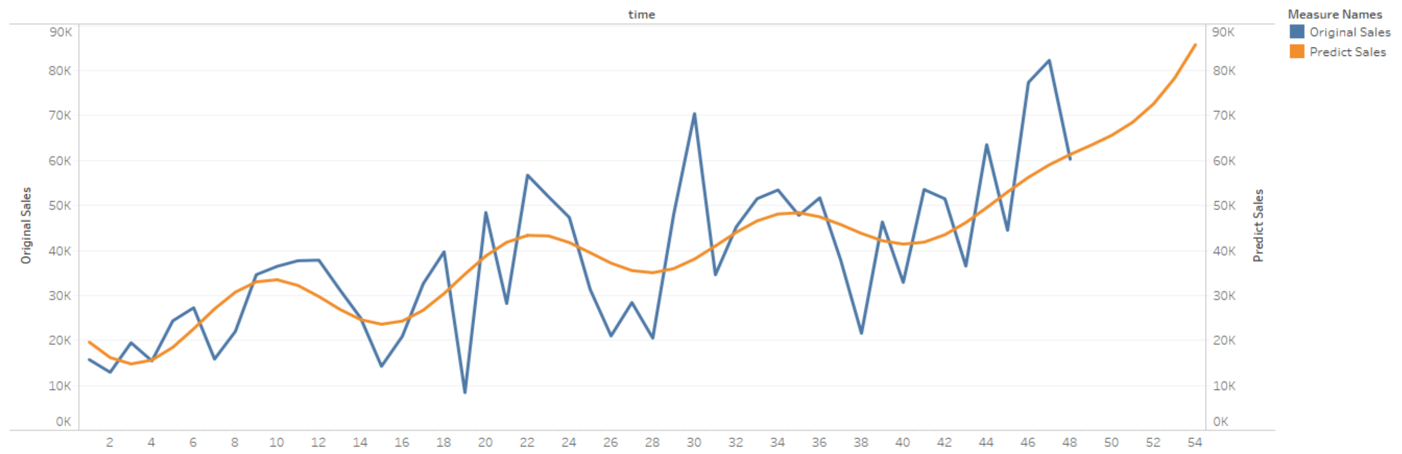


3. APAC Consumer Market Bucket

Sales Forecasting

From the model evaluation table shown in slide9. The Classical Decomposition Forecasting model performs better than the Auto Arima model. **It provides a MAPE reduction of 8.19% and a Root Mean Square Error reduction of 29.55%** in comparison to the Auto Arima model. The graph to the below shows the Forecasted Sales vs. Actual Sales along with the Forecasted Sales for the future 6-month period.

Sales Forecast for Asia Pacific Consumer Segment



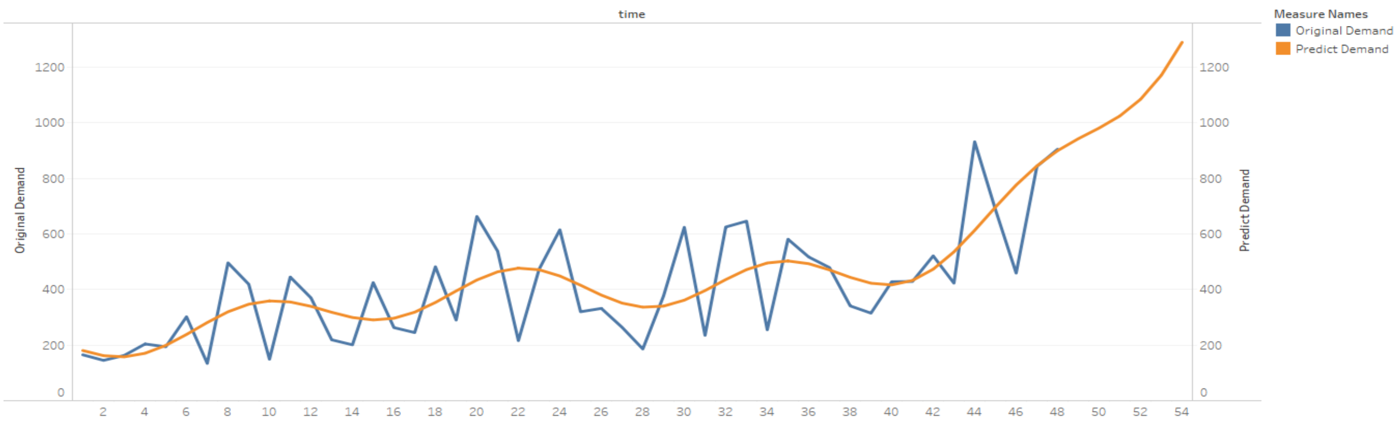
APAC Consumer Sales Forecast	
Future Month	Sales Forecast
Jan-15	63405.05
Feb-15	65629.26
Mar-15	68534.04
Apr-15	72620.16
May-15	78283.06
Jun-15	85754.44

4. EU Consumer Market Bucket

Demand Forecasting

From the model evaluation table shown in slide10. The Classical Decomposition Forecasting model performs better than the Auto Arima model. **It provides a MAPE reduction of 8.15% and a Root Mean Square Error reduction of 40.23%** in comparison to the Auto Arima model. The graph to the below shows the Forecasted Demand vs. Actual Demand along with the Forecasted Demand for the future 6-month period.

Demand Forecast for EU Consumer Segment

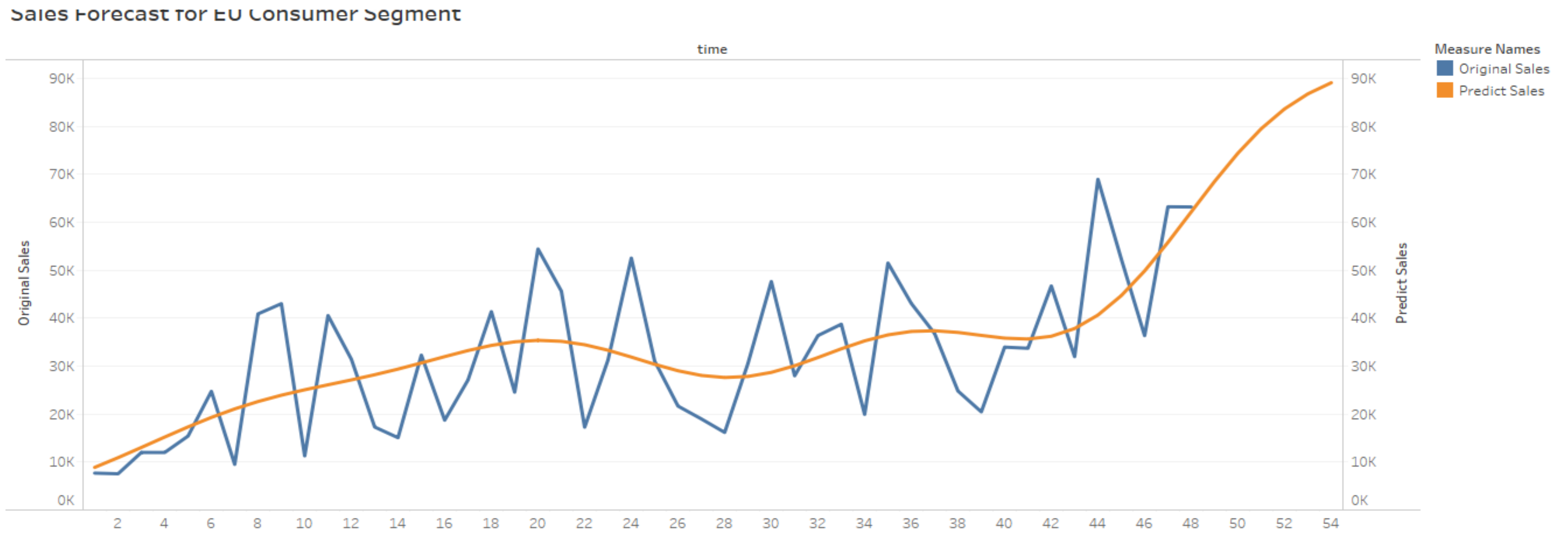


EU Consumer Demand Forecast	
Future Month	Demand Forecast
Jan-15	942.8787
Feb-15	981.0561
Mar-15	1024.5725
Apr-15	1084.6744
May-15	1171.0478
Jun-15	1289.6338

5. EU Consumer Market Bucket

Sales Forecasting

From the model evaluation table shown in slide9. The Classical Decomposition Forecasting model performs better than the Auto Arima model. **It provides a MAPE reduction of 8.19% and a Root Mean Square Error reduction of 29.55%** in comparison to the Auto Arima model. The graph to the below shows the Forecasted Sales vs. Actual Sales along with the Forecasted Sales for the future 6-month period.



EU Consumer Sales Forecast	
Future Month	Sales Forecast
Jan-15	68506.41
Feb-15	74397.72
Mar-15	79502.25
Apr-15	83628.35
May-15	86762.69
Jun-15	89076.09

Thank You!