Retail Giant Sales Forecasting Case Study

By Vatsal Gohel

Global Market Online Supergiant Store

- Global Mart Store takes online orders and delivers across the 7 global regions and deals with Consumer, Corporate and home office product categories

Market	Segment
Africa	Consumer
APAC (Asia Pacific)	Corporate
Canada	Home Office
EMEA(Middle East)	
EU (European Union)	
LATAM (Latin America)	
US (United States)	

Problem Statement

- The store wants to finalize the inventory management plan for the next 6 months. Hence the objective of the analysis are:

  - Find out the most profitable (and consistent) market segments for the company.
  - For these segments, forecast the sales and the demand for the next 6 months, so that the revenue and inventory may be managed accordingly

- The analysis has been divided into four parts:

  - Data Understanding
  - Finding the most profitable segments
  - Forecasting sales and demand for each of the profitable segments
  - Recommendations for inventory management

Understanding the Data

Data Preparation

- Create 21 Market Segments Buckets
- Aggregate buckets by Sales, Profit & Quantity
- Perform Train – Test split
- Calculate Coefficient of Variation for train Data
- Retain only the Market Segment with lowest Cov

Modelling

- Create time series of top aggregated data
- Smoothen time series to identify trend & Seasonality
- Creating a train & validation set of size 42 & 6 months
- Build Different Models

Model Evaluation

- Evaluate model on validation set using RMSE & MAPE
- Choose Best Model out of all the Models using MAPE

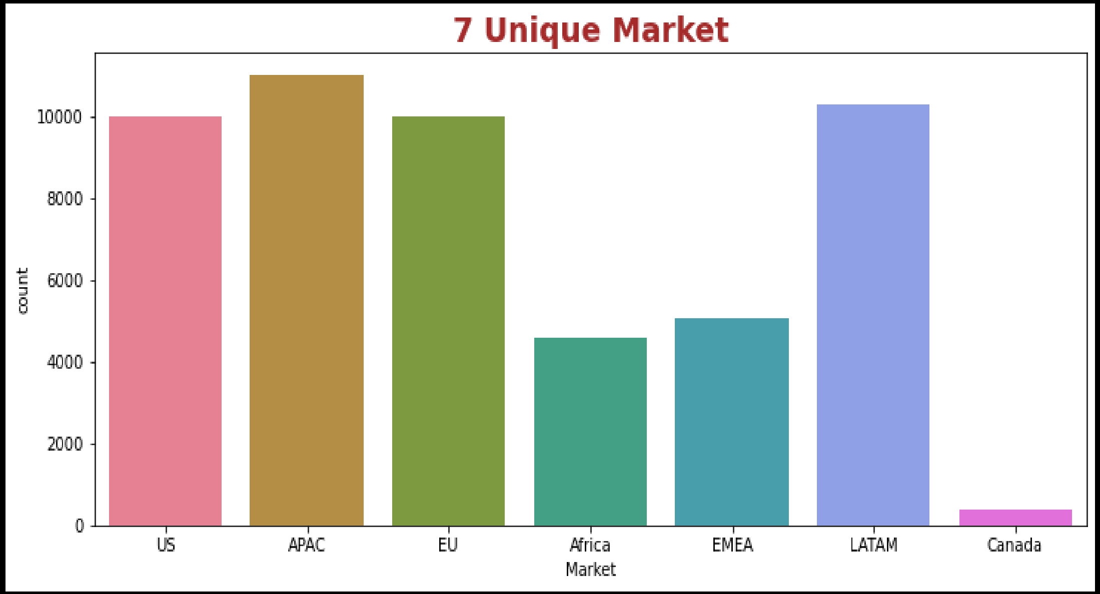
Forecasting

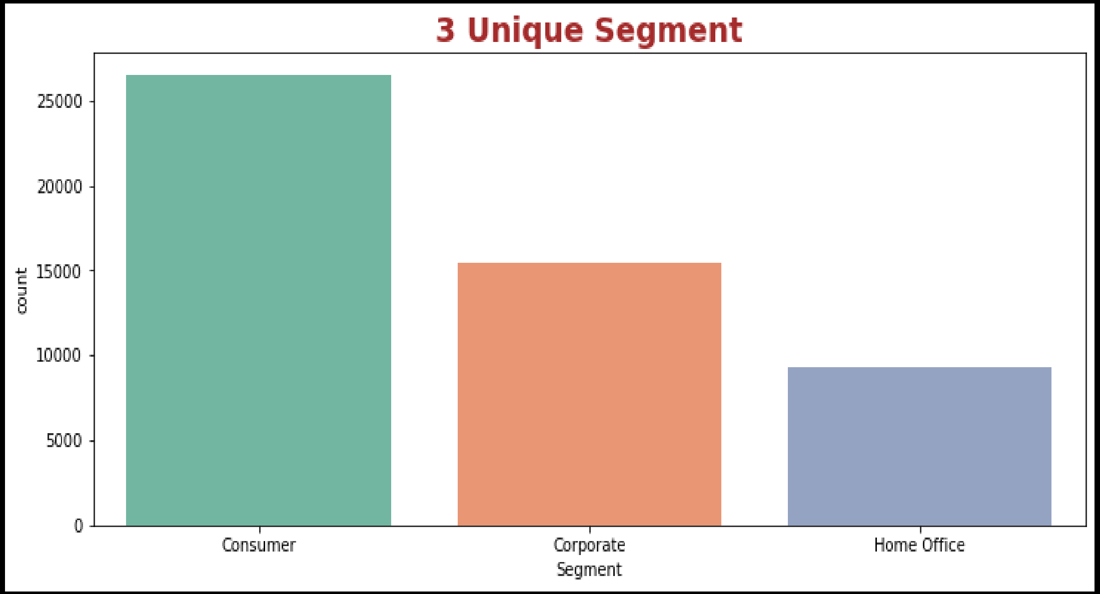
- Use Best Model to forecast future of 6 month sales for the most profitable Market Segment

Data Preparation

- First, we need to check that there is no missing values present in the data set. In the provided data set, there are no missing values thus no need to impute missing values

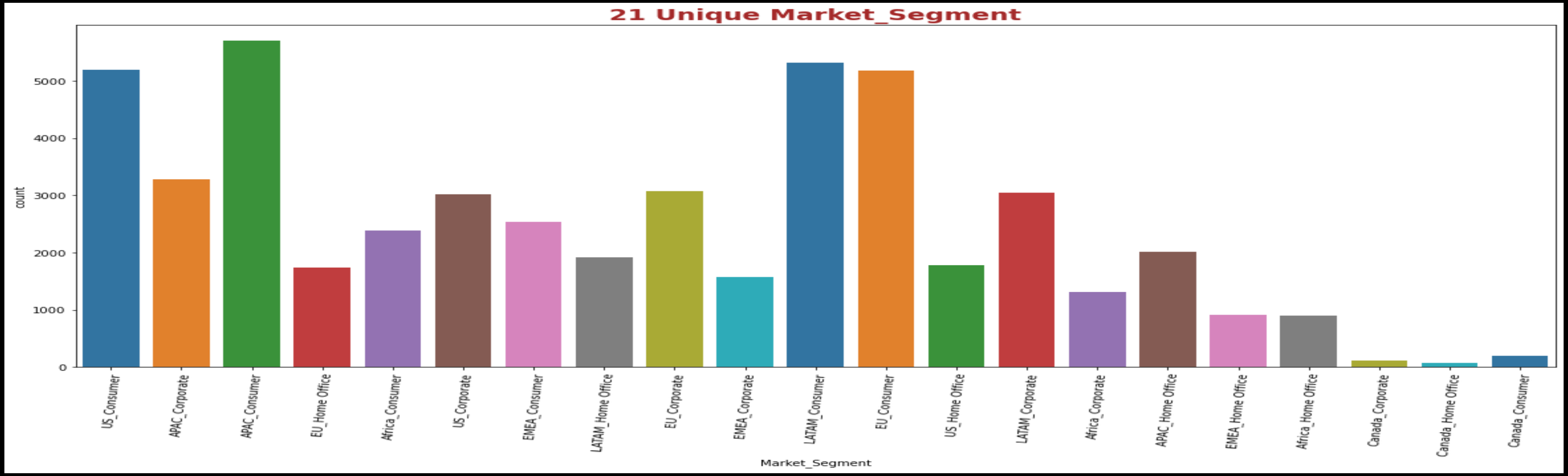
- Identifying the unique data from the data set
  - We Can see 3 Segment & 7 market which are unique in the data set
  - We will visualize the unique columns of segment and market through graph





Data Preparation

- Preparing an aggregated data with unique market & segment and name it as Market-Segment
- After aggregating the data, we will get 21 unique Market Segments
  - Visualizing the Market Segment through graph
  - We can see that APAC_Consumer has the highest count



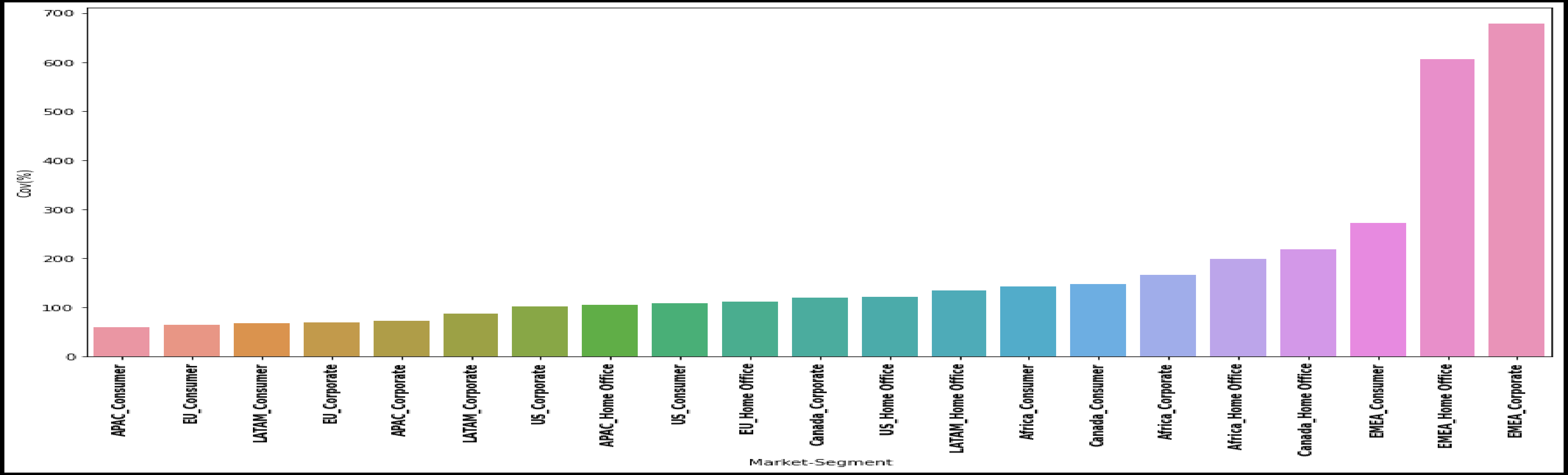
Identifying the most profitable Market-Segment

- We get 21 unique Market-Segments after combining both segments and markets
- Aggregate the data in each bucket based on Sales, Quantity & Profit
- Perform the Train – Test split in the aggregated data
- Calculate the CoV on the profit for each of the 21 market segments on the train data
  - **CoV = Standard Deviation / Mean**
- We want to forecast the sales where the market segment is reliable or in other words, there is less variation in the profits

	Market-Segment	Mean	Std	Cov(%)
0	APAC_Consumer	4223.6	2518.9	59.6
12	EU_Consumer	3627.5	2348.8	64.7
15	LATAM_Consumer	2252.7	1533.4	68.1
13	EU_Corporate	2252.0	1552.4	68.9
1	APAC_Corporate	2557.0	1871.5	73.2
16	LATAM_Corporate	1076.0	947.2	88.0
19	US_Corporate	1853.6	1904.0	102.7
2	APAC_Home Office	1379.1	1446.4	104.9
18	US_Consumer	2603.7	2851.9	109.5
14	EU_Home Office	1097.4	1223.3	111.5
7	Canada_Corporate	110.4	132.1	119.7
20	US_Home Office	1062.4	1293.1	121.7
17	LATAM_Home Office	788.5	1059.5	134.4
3	Africa_Consumer	798.9	1141.9	142.9
6	Canada_Consumer	230.1	339.6	147.6
4	Africa_Corporate	426.0	709.3	166.5
5	Africa_Home Office	333.0	662.6	199.0
8	Canada_Home Office	138.2	302.5	218.9
9	EMEA_Consumer	415.4	1128.5	271.7
11	EMEA_Home Office	123.2	747.7	606.9
10	EMEA_Corporate	172.3	1168.0	677.9

Identifying the most profitable Market-Segment

- Plotting the CoV curve to get the better idea
- According to data and graph, **'APAC_Consumer'** has the least CoV of 59.6, which indicates that it is the most stable market segment
- As a result, it makes sense to analyze the **'APAC_Consumer'** market segment

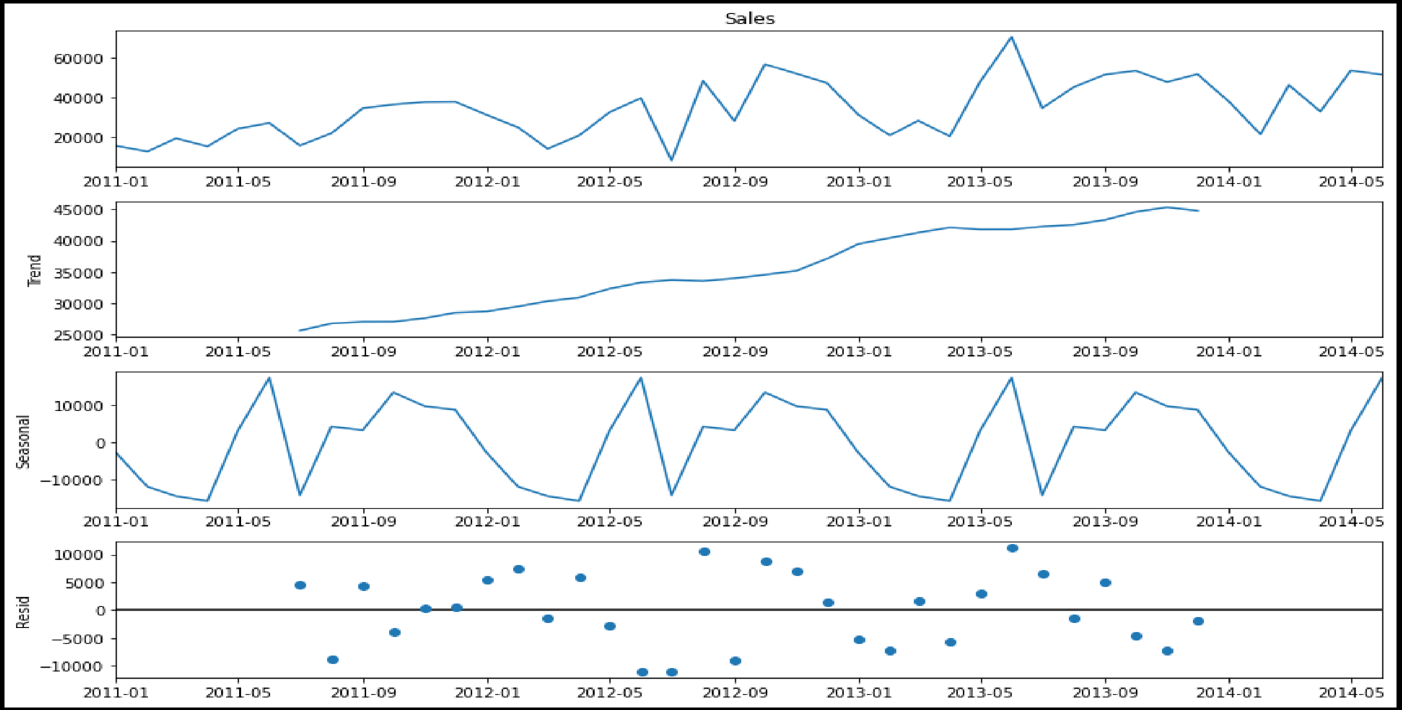


Time Series Decomposition

- Lets understand how a time series can be split into its various components that is the Trend, Seasonality, and Residuals

Seasonal Decomposition of 'APAC_Consumer' Sales Data (Additive)

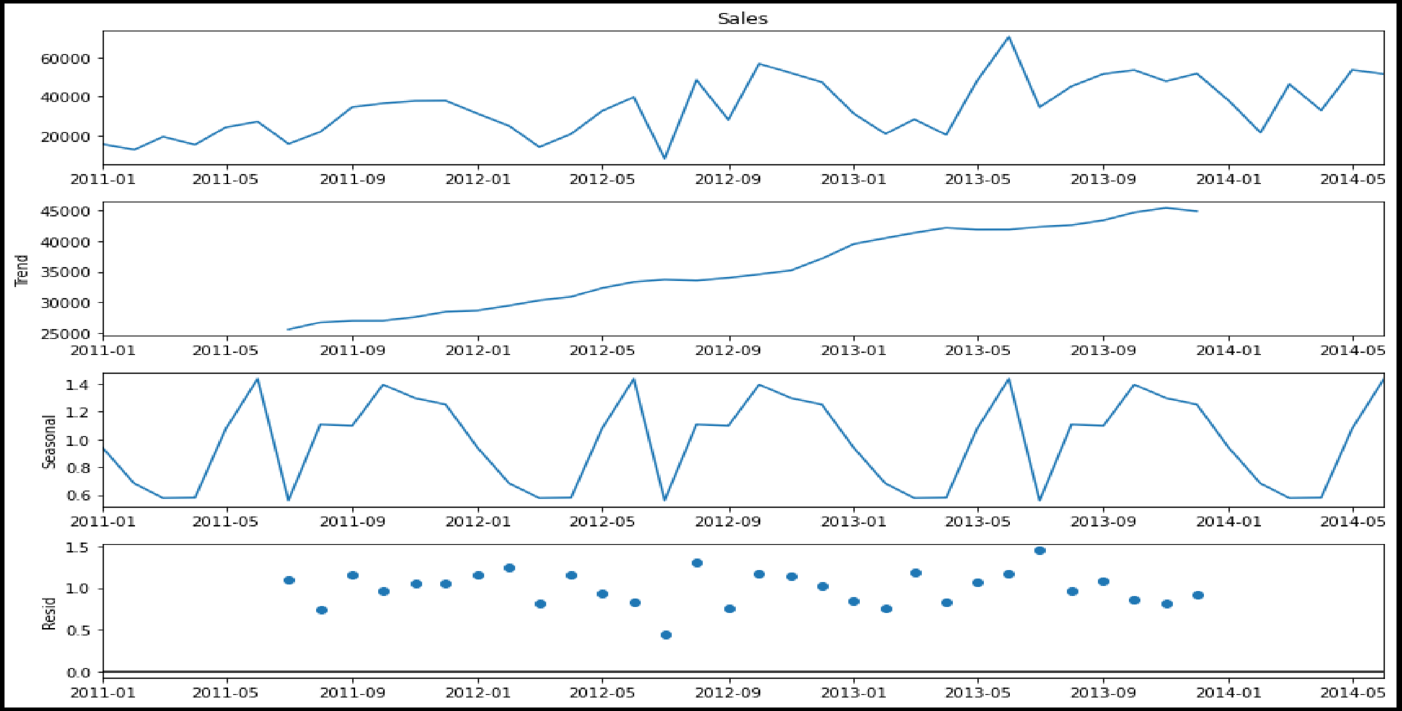
- Decomposed the data using the additive method:
  - We can see the clear upward trend and some seasonality as well in the graph
  - The residuals seem to have some pattern



Time Series Decomposition

Seasonal Decomposition of ‘APAC_Consumer’ Sales Data (Multiplicative)

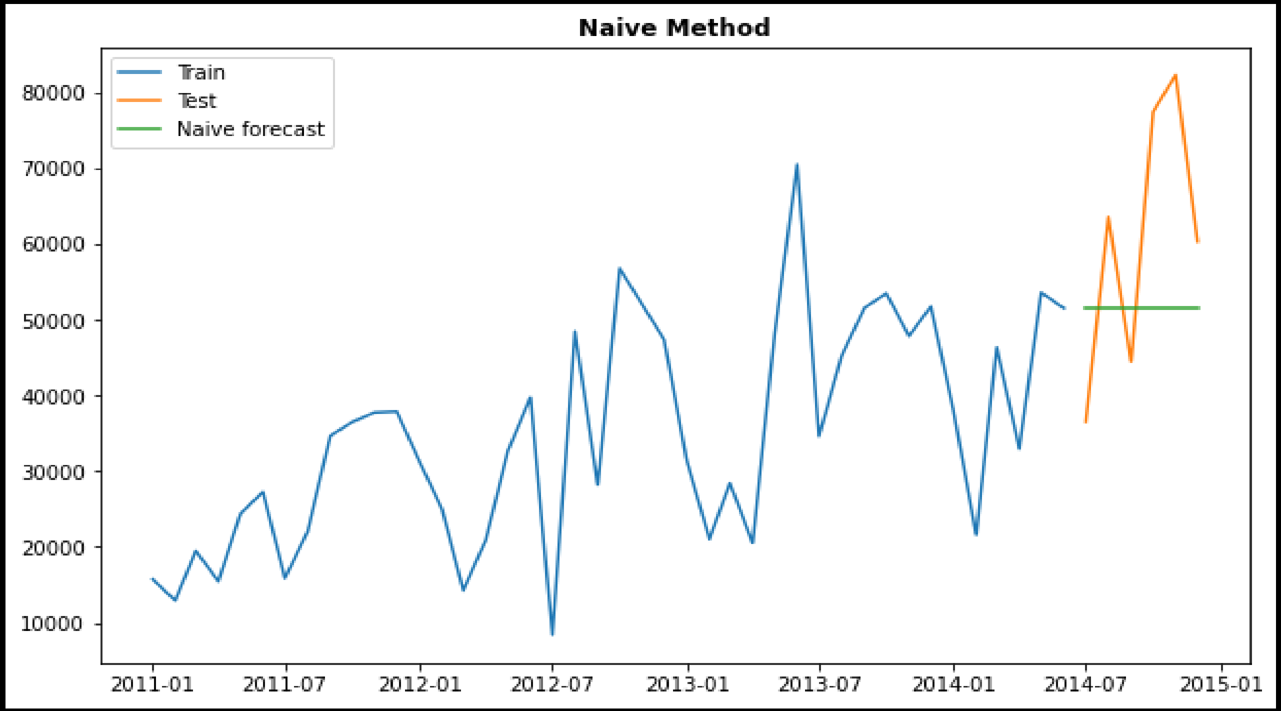
- Decomposed the data using the multiplicative method:
  - we can see that the trend is again having the upward movement
  - A certain pattern can be sensed in seasonality, and residuals seem to have been disrupted a little bit



Building and Evaluating Time Series Forecasts

Simple time series methods

- Time Series Model – Naïve Method:

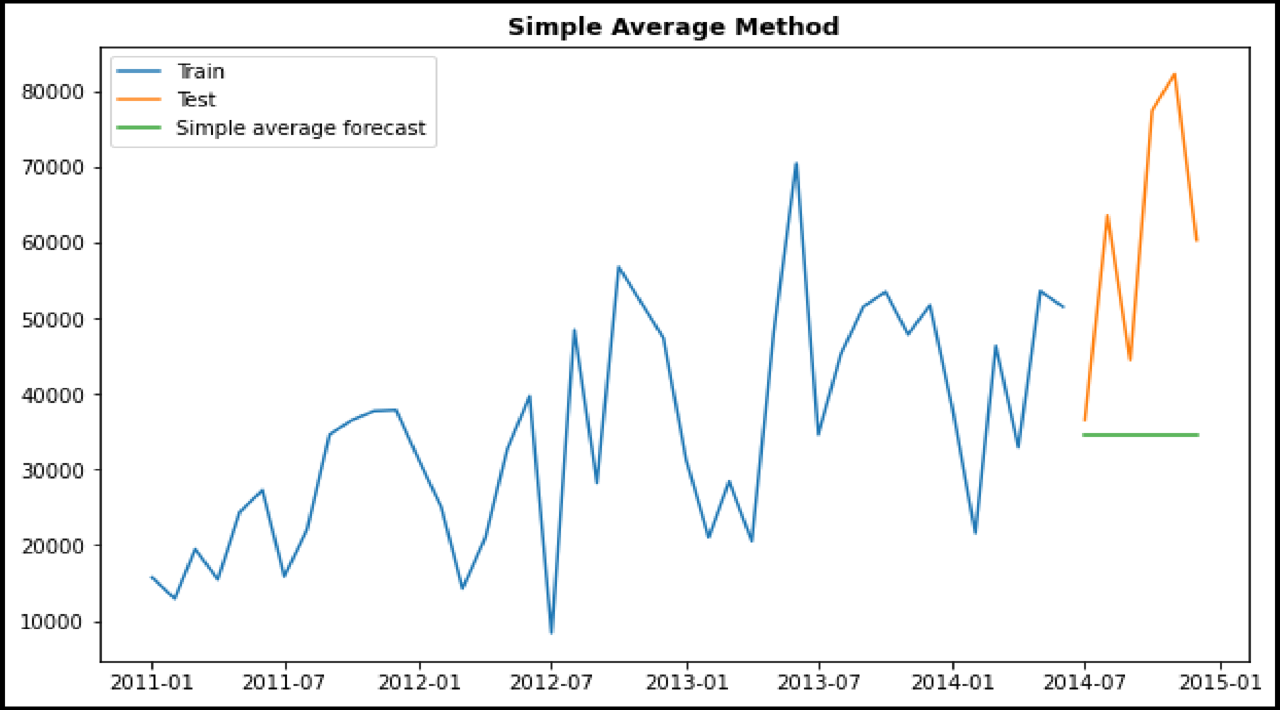


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86

Building and Evaluating Time Series Forecasts

Simple time series methods

- Time Series Model – Simple Average Method:

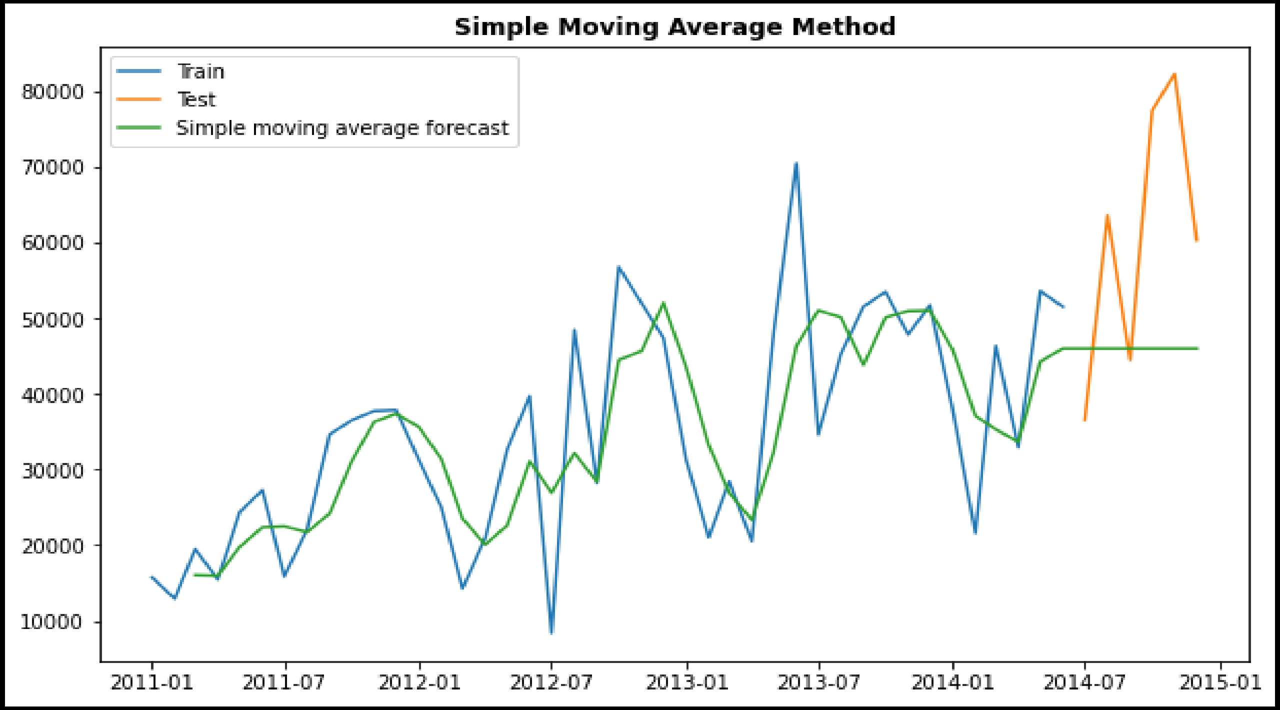


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18

Building and Evaluating Time Series Forecasts

Simple time series methods

- Time Series Model – Simple Moving Average Method:

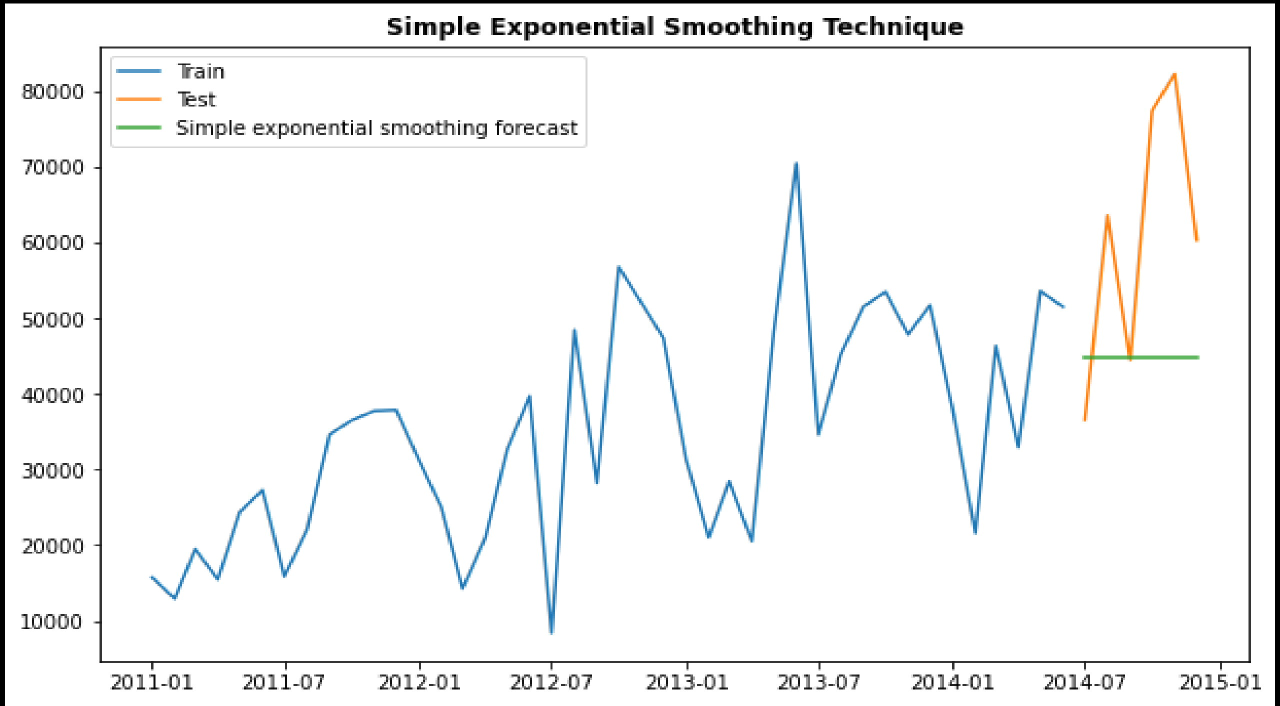


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55

Building and Evaluating Time Series Forecasts

Smoothing Techniques

- Time Series Model – Simple Exponential Smoothing:

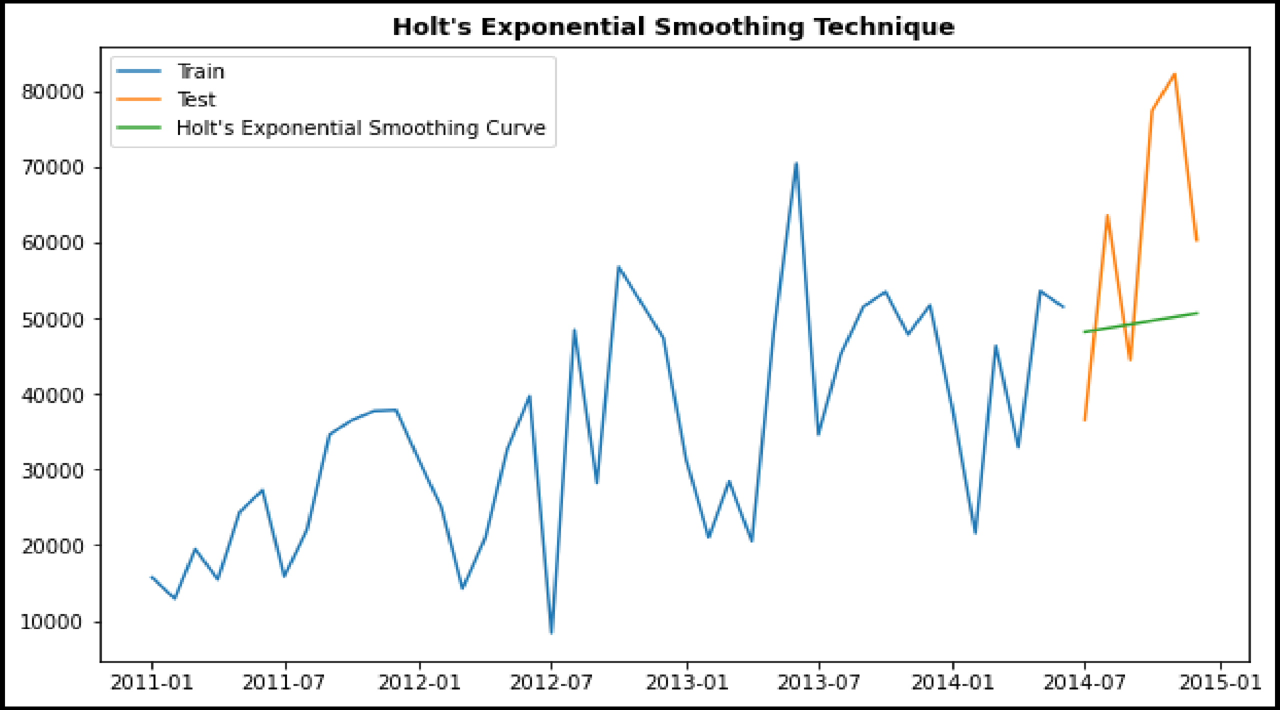


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70

Building and Evaluating Time Series Forecasts

Smoothing Techniques

- Time Series Model – Holt's Exponential Smoothing:

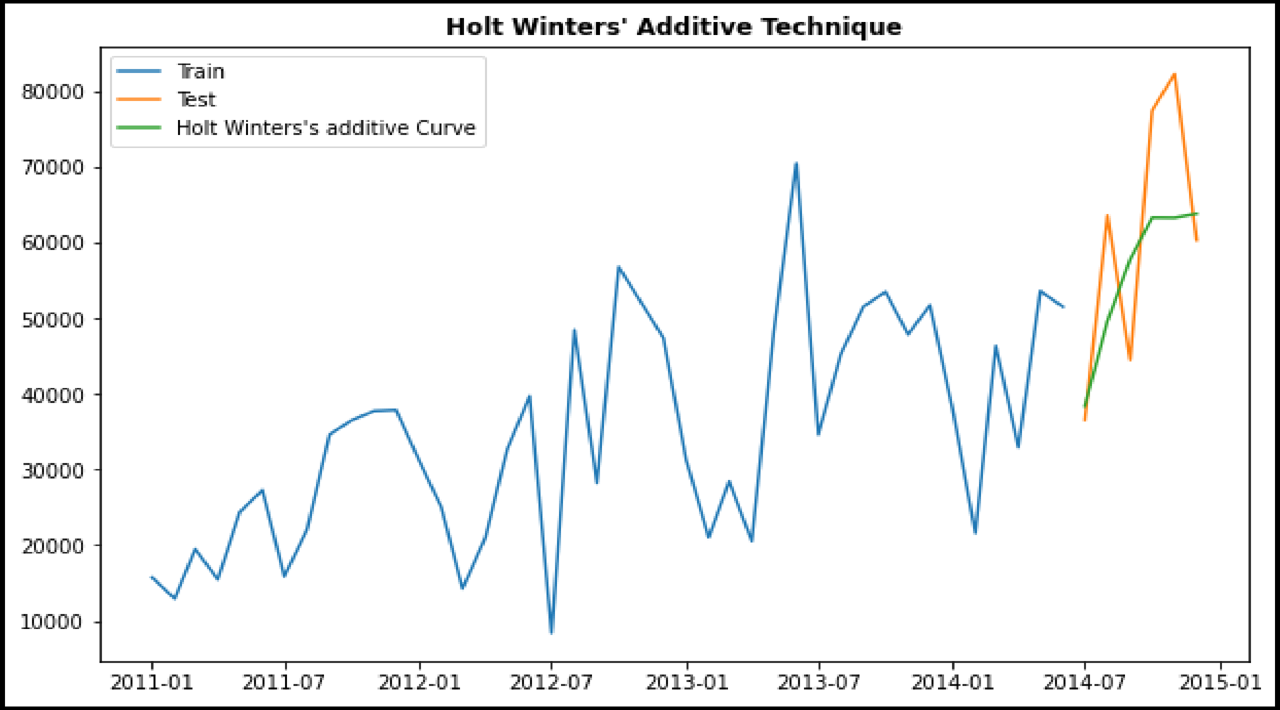


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12

Building and Evaluating Time Series Forecasts

Smoothing Techniques

- Time Series Model – Holt Winters' additive method:

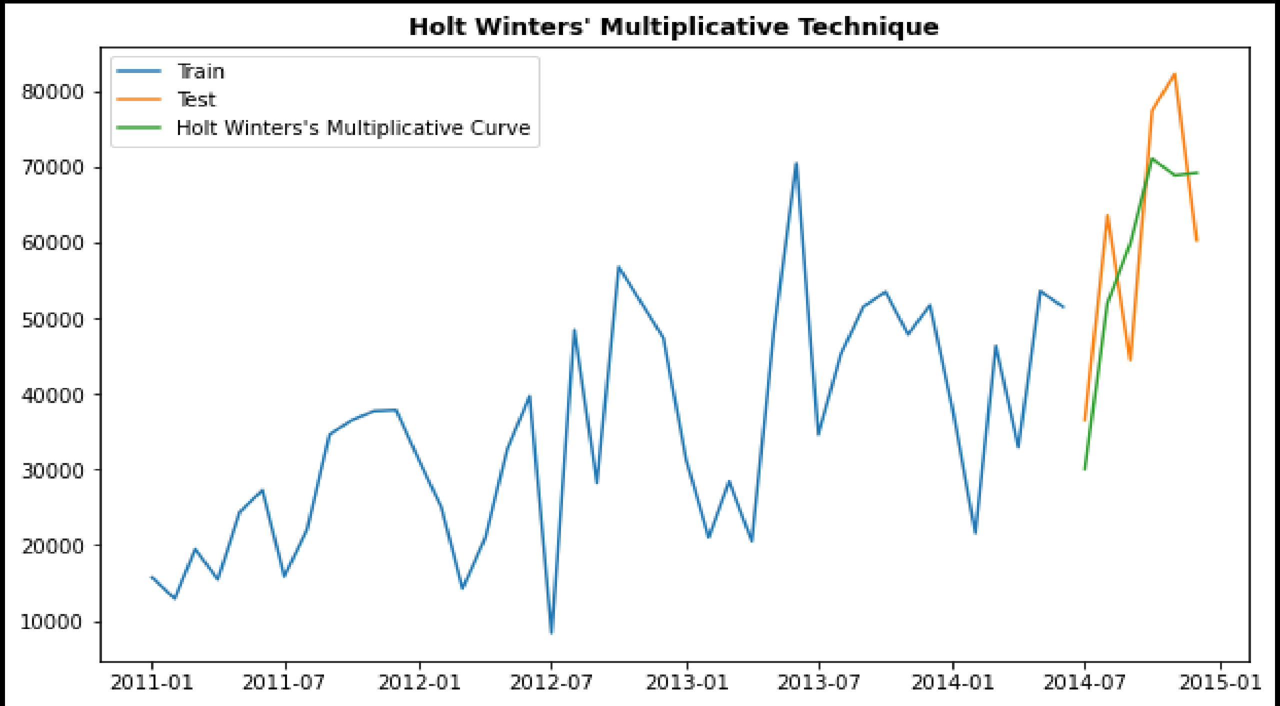


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32

Building and Evaluating Time Series Forecasts

Smoothing Techniques

- Time Series Model – Holt Winters' multiplicative method:



	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27

Building and Evaluating Time Series Forecasts

Auto Regressive methods

- First we need to check whether the time series is stationary or Non-Stationary
- If the Time-Series is not Stationary, we need to transform it in order to make it Stationary
- Two statistical tests, namely ADF and KPSS, will be performed in order to validate the Stationarity of the Time Series
- Augmented Dickey-Fuller (ADF) Test:
  - Null Hypothesis ($H0$): The series is not stationary $p\text{–}value$ > 0.05
  - Alternate Hypothesis ($H1$): The series is stationary $p\text{–}value$ <= 0.05
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test:
  - Null Hypothesis ($H0$): The series is stationary $p\text{–}value$ > 0.05
  - Alternate Hypothesis ($H1$): The series is not stationary $p\text{–}value$ <= 0.05

Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Augmented Dickey-Fuller (ADF) Test:
  - By performing the test we get the p-value of 0.19 which is greater than 0.05 so we will go with the Null Hypothesis that the series is not stationary

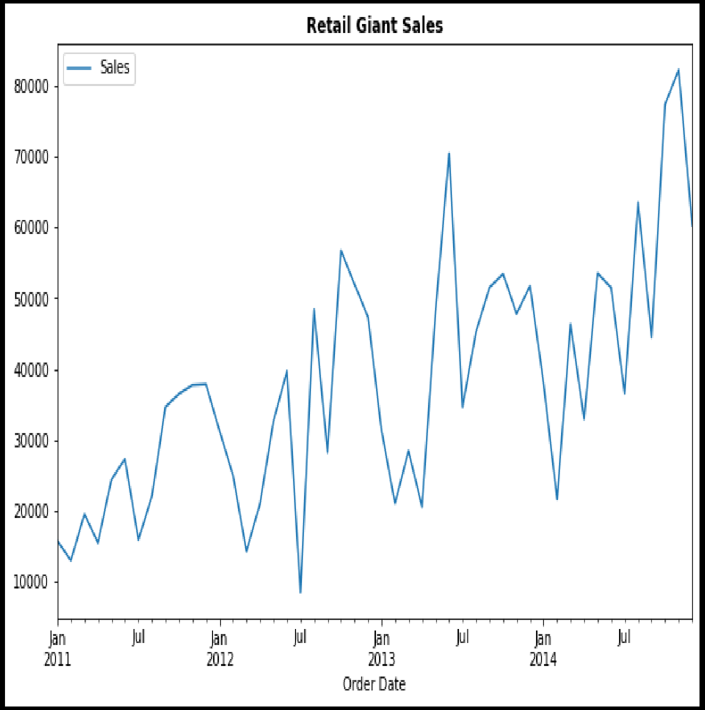
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test:
  - By performing the test we get the p-value of 0.023 which is less than 0.05 thus the Null Hypothesis that is the series is stationary gets rejected

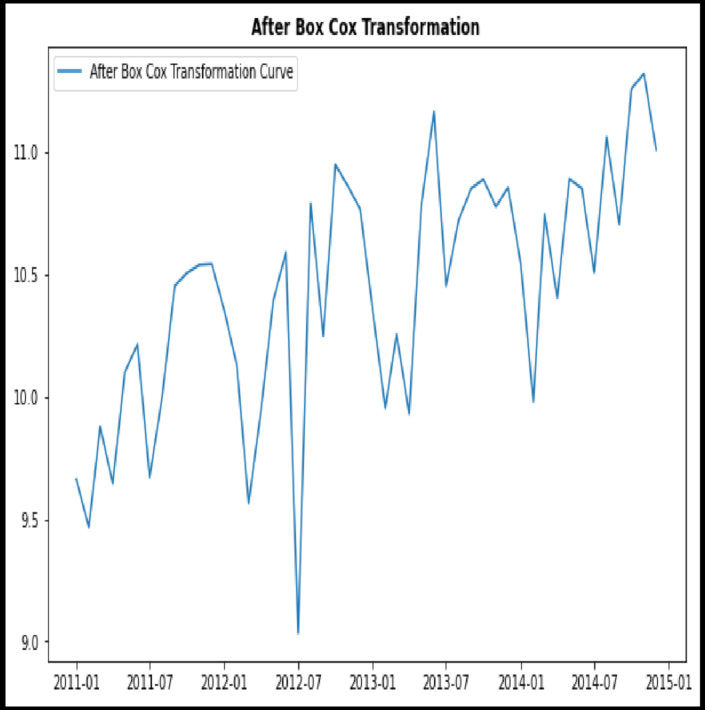
- To make the dataset stationary and suitable for Auto Regression Models, we need to take necessary steps such as differencing and Box Cox Transformation

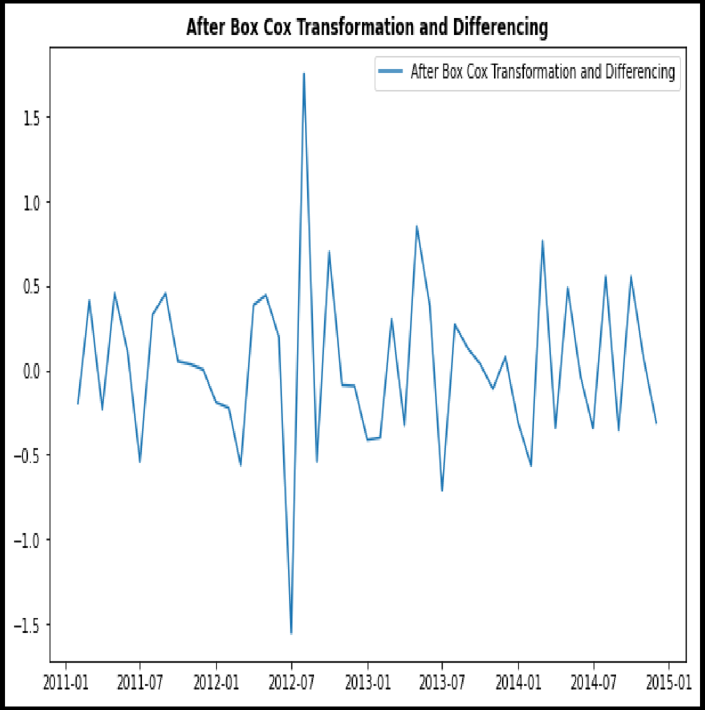
Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Box Cox Transformation & Differencing
  - To make the Non-Stationary data into Stationary data



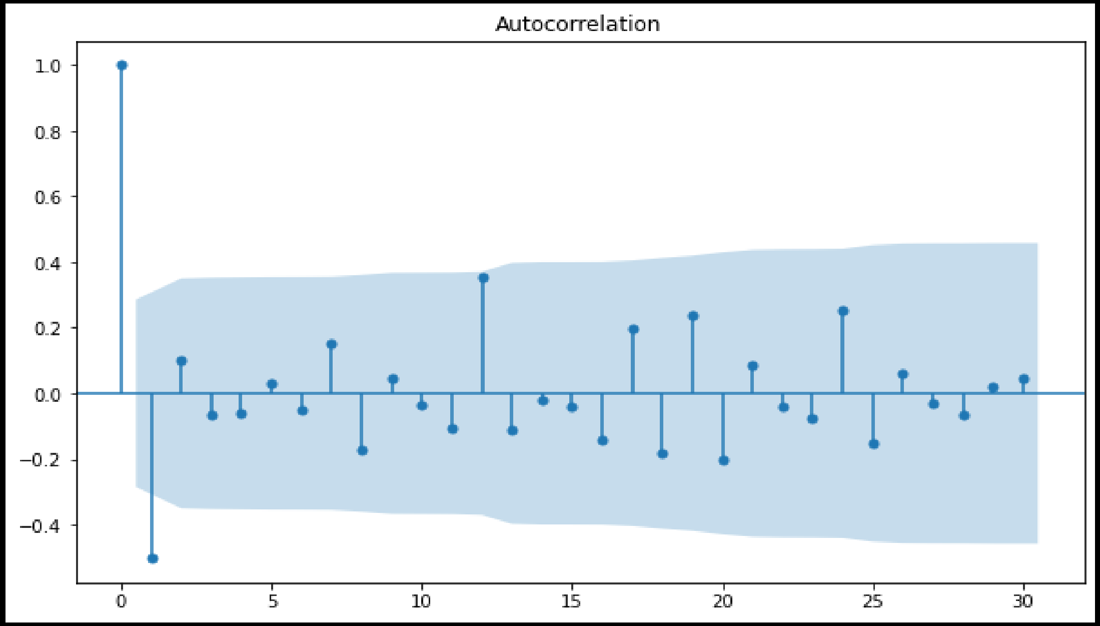


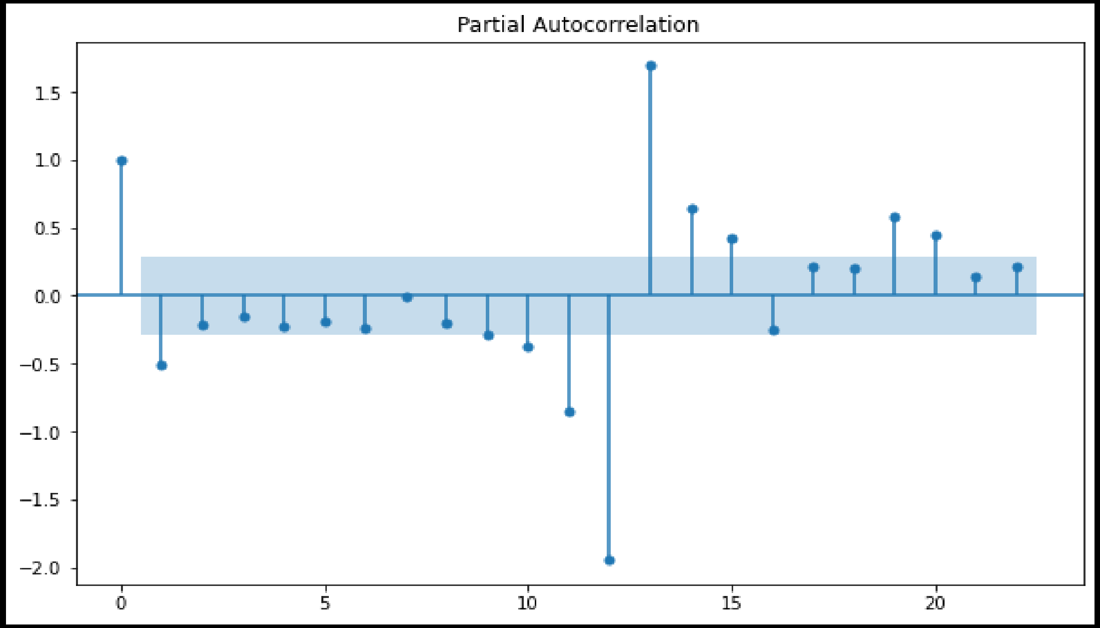


Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Autocorrelation Function (ACF) & Partial Autocorrelation (PACF) Plot
  - From ACF plot we could see the dependency on the very next node which means MA should be 1
  - From PACF plot we could see there is a seasonality in the data

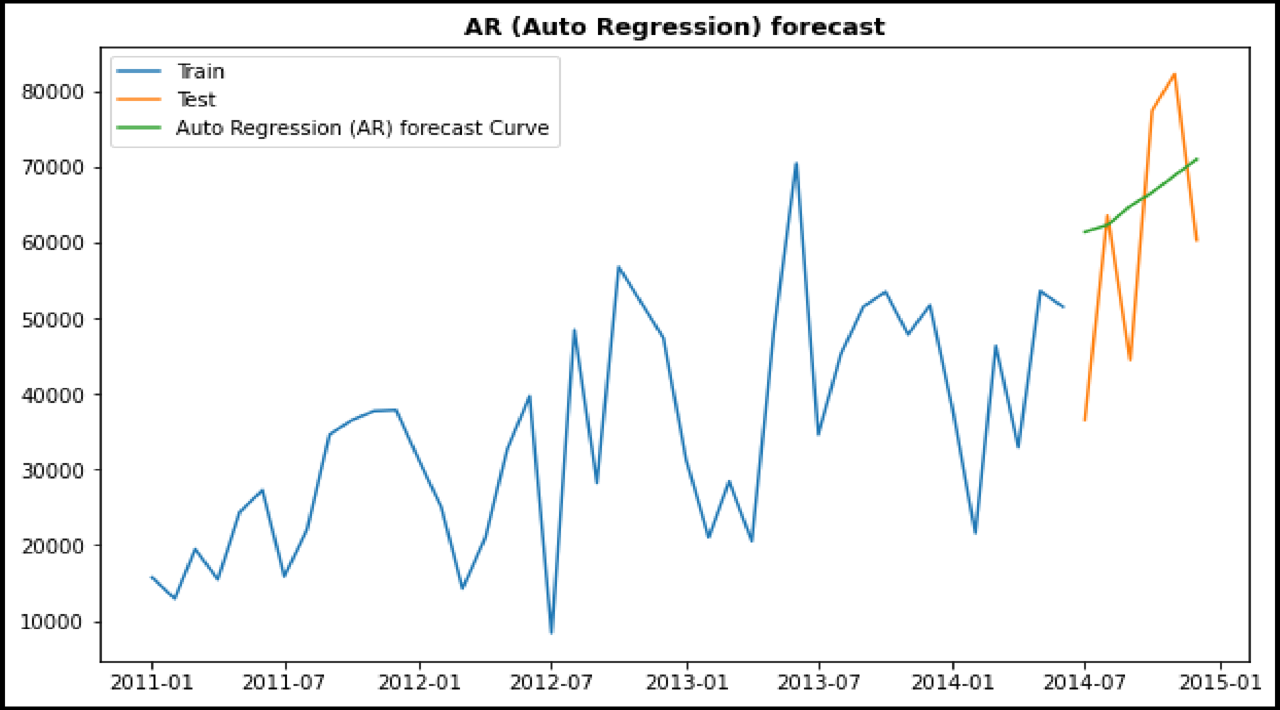




Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Time Series Model – Auto Regression (AR) Method:

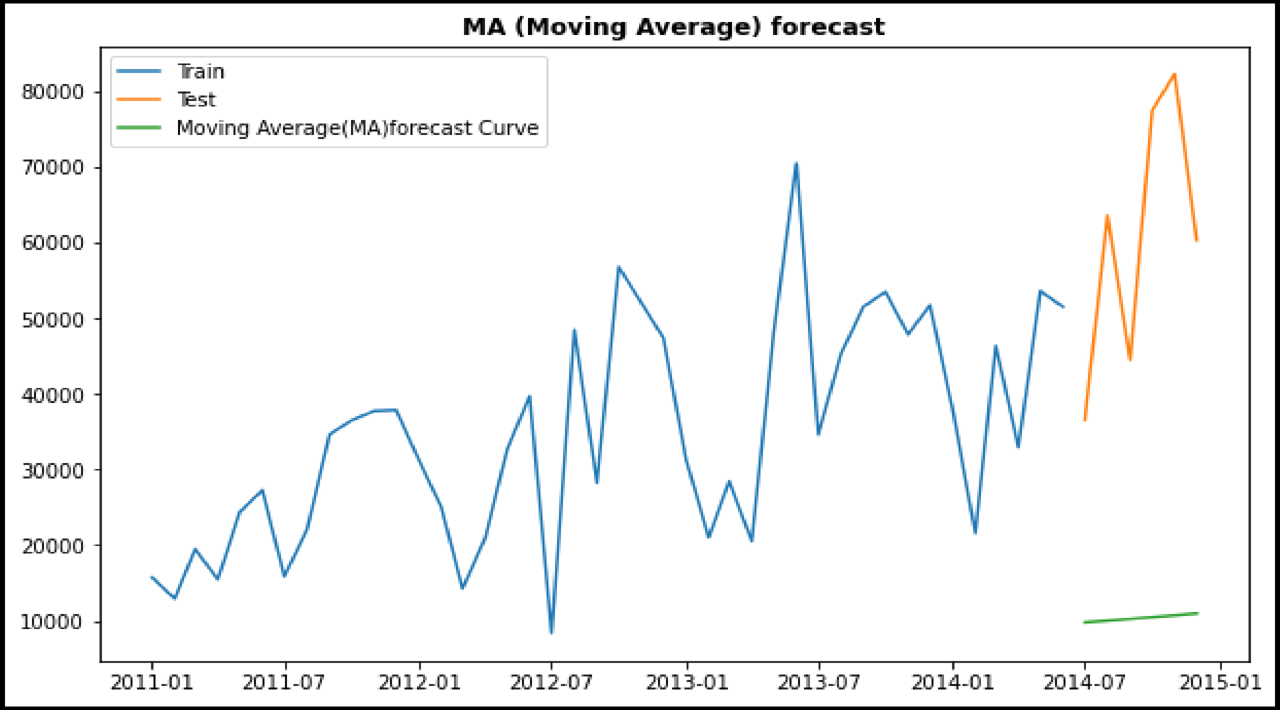


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27
0	AR (Auto Regression) Method	15505.03	27.27

Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Time Series Model – Moving Average (MA) Method:

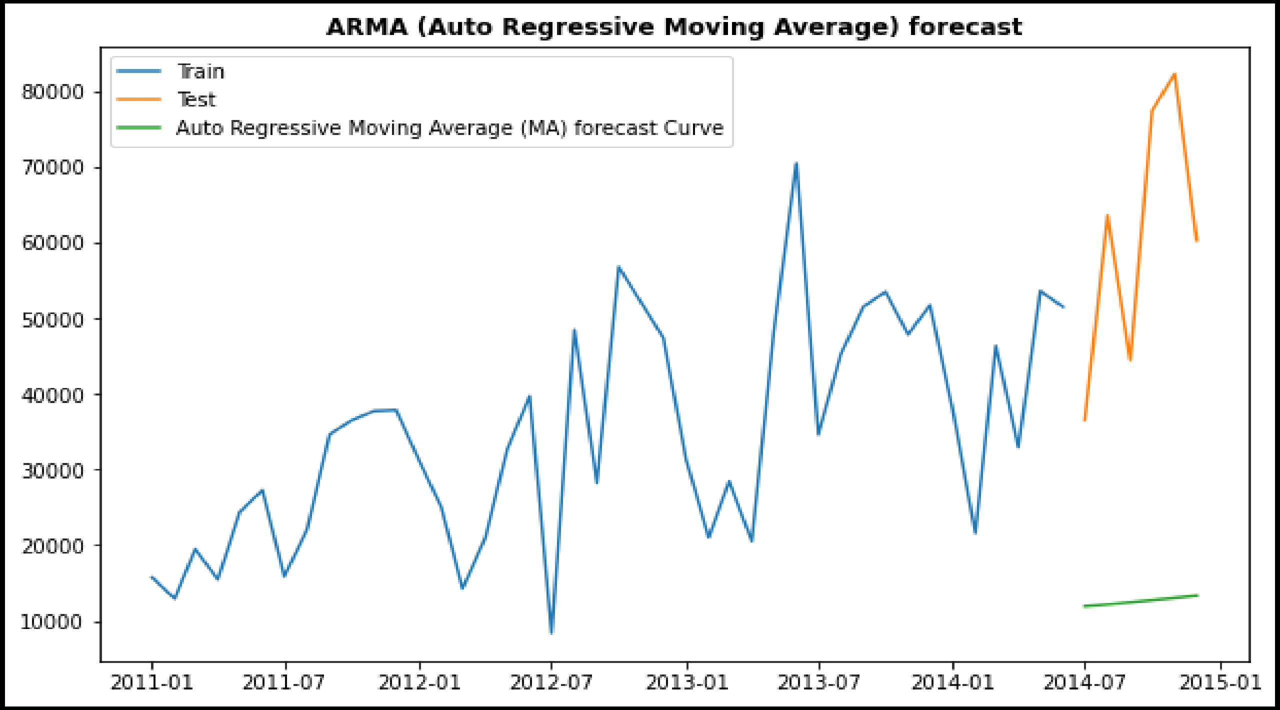


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27
0	AR (Auto Regression) Method	15505.03	27.27
0	MA (Moving Average) Method	52903.35	81.64

Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Time Series Model – Auto Regressive Moving Average (ARMA) Model:

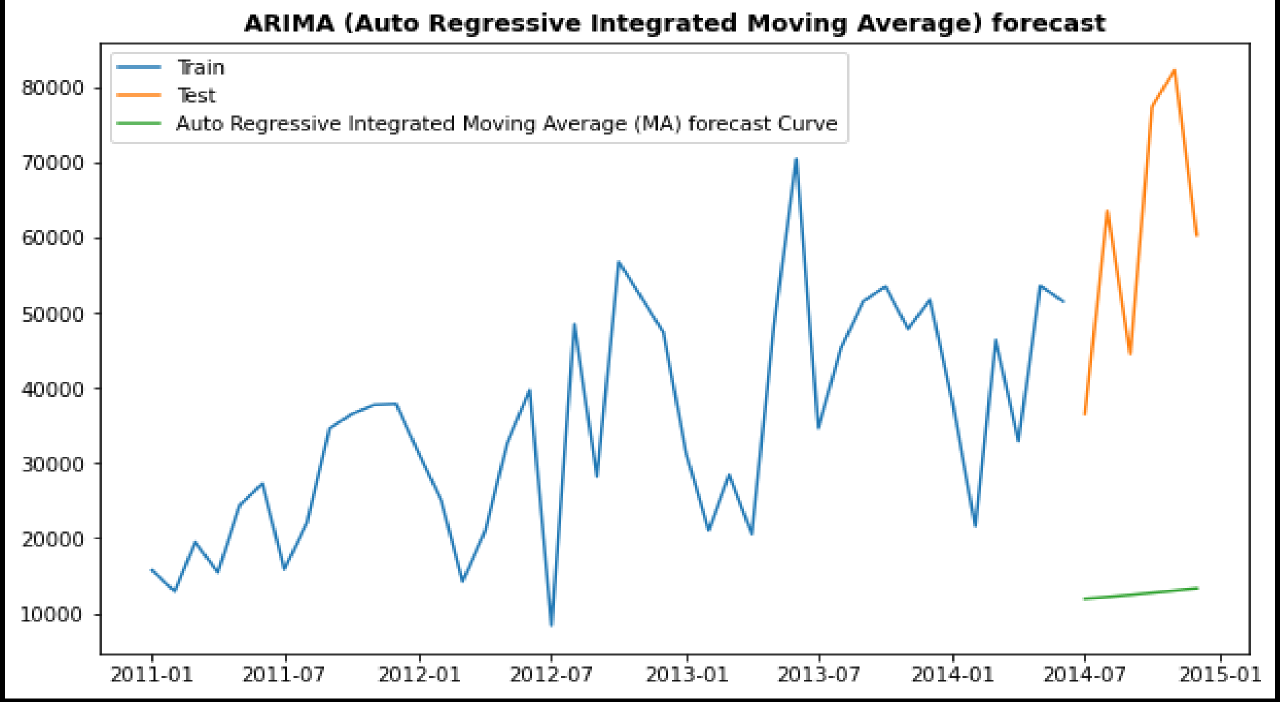


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27
0	AR (Auto Regression) Method	15505.03	27.27
0	MA (Moving Average) Method	52903.35	81.64
0	ARMA (Auto Regressive Moving Average) Method	50757.91	77.66

Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Time Series Model – Auto Regressive Integrated Moving Average (ARIMA) Model:

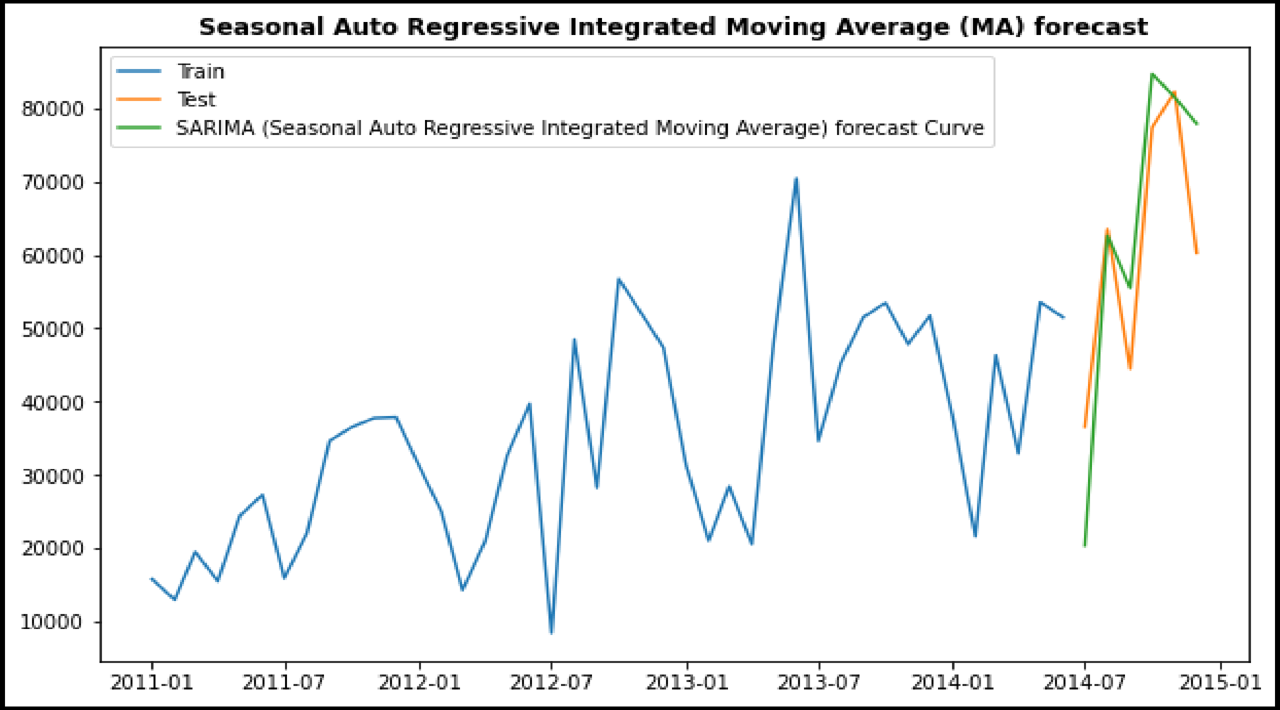


	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27
0	AR (Auto Regression) Method	15505.03	27.27
0	MA (Moving Average) Method	52903.35	81.64
0	ARMA (Auto Regressive Moving Average) Method	50757.91	77.66
0	ARIMA (Auto Regressive Integrated Moving Avera...	50757.91	77.66

Building and Evaluating Time Series Forecasts

Auto Regressive methods

- Time Series Model – Seasonal Auto Regressive Integrated Moving Average (SARIMA) Model:



	Method	RMSE	MAPE
0	Naive method	18774.05	26.86
0	Simple average method	30846.00	38.18
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27
0	AR (Auto Regression) Method	15505.03	27.27
0	MA (Moving Average) Method	52903.35	81.64
0	ARMA (Auto Regressive Moving Average) Method	50757.91	77.66
0	ARIMA (Auto Regressive Integrated Moving Avera...	50757.91	77.66
0	SARIMA (Seasonal Auto Regressive Integrated Mo...	11188.69	18.38

Model Evaluation Based on MAPE

- The best performing time series model is **Holt Winters' Additive Method** out of 12 time series models

	Method	RMSE	MAPE
0	Holt Winters's Additive Method	12565.60	17.32
0	Holt Winters's Multiplicative Method	10876.35	18.27
0	SARIMA (Seasonal Auto Regressive Integrated Mo...	11188.69	18.38
0	Holt's Exponential Trend Technique	19473.57	26.12
0	Naive method	18774.05	26.86
0	AR (Auto Regression) Method	15505.03	27.27
0	Simple moving average forecast	22019.48	27.55
0	Simple Exponential Method	22824.62	27.70
0	Simple average method	30846.00	38.18
0	ARMA (Auto Regressive Moving Average) Method	50757.91	77.66
0	ARIMA (Auto Regressive Integrated Moving Avera...	50757.91	77.66
0	MA (Moving Average) Method	52903.35	81.64

Conclusion

- We were able to help the Global Market identify the most popular market segment named **'APAC_Consumer'** based on the data provided
- There are 12 different forecasting models designed for the 'APAC_Consumer' market segment
- **Holt Winters' Additive Method** is the most accurate forecasting method among smoothing methods
- Forecasting accuracy is best achieved with **Seasonal Autoregressive Integrated Moving Average (SARIMA)** among the ARIMA set of techniques
- Analyzing the plot enabled us to realize that sales have a trend and a seasonality. Unlike other models which allow us to choose between the two, **Holt Winter's and SARIMA** allows to capture both
- Additionally, a time series of sales data may have some sort of seasonality associated with it since the sales of any item will not remain the same throughout the given period, usually a year. There may also be a positive or negative trend associated with it
- For this reason, **Holt Winter's models and SARIMA** models are the best sales forecasting models