Retail-Giant Sales Forecasting Case Study

Group Name:

1. Mohit Rathore
2. Mahesh Kulkarni
3. Abdulrahiman Bangara
4. Abhinav Marathe

Abstract

We identified 2 most profitable market segment for “Global Mart” as APAC Consumer & EU Consumer. We forecasted future 6 months sales & quantity for these using time series forecasting methods. Total 8 forecasting models for top 2 segments out of which 4 best models were selected for forecasting future 6 months sales & quantity from January 2015 to June 2015. Models were created using Liner Regression & Auto ARIMA techniques on time series data.

Problem solving methodology

Data Preparation

- Create 21 Market-Segment buckets
- Aggregate buckets by Sales, Profit, Quantity
- Calculate Coefficient of Variation(CV)
- Pick Top 2 Segments based on CV & Profit
- Aggregate data by Market-Segment & Order Month on subset of Top 2 segments

Modeling

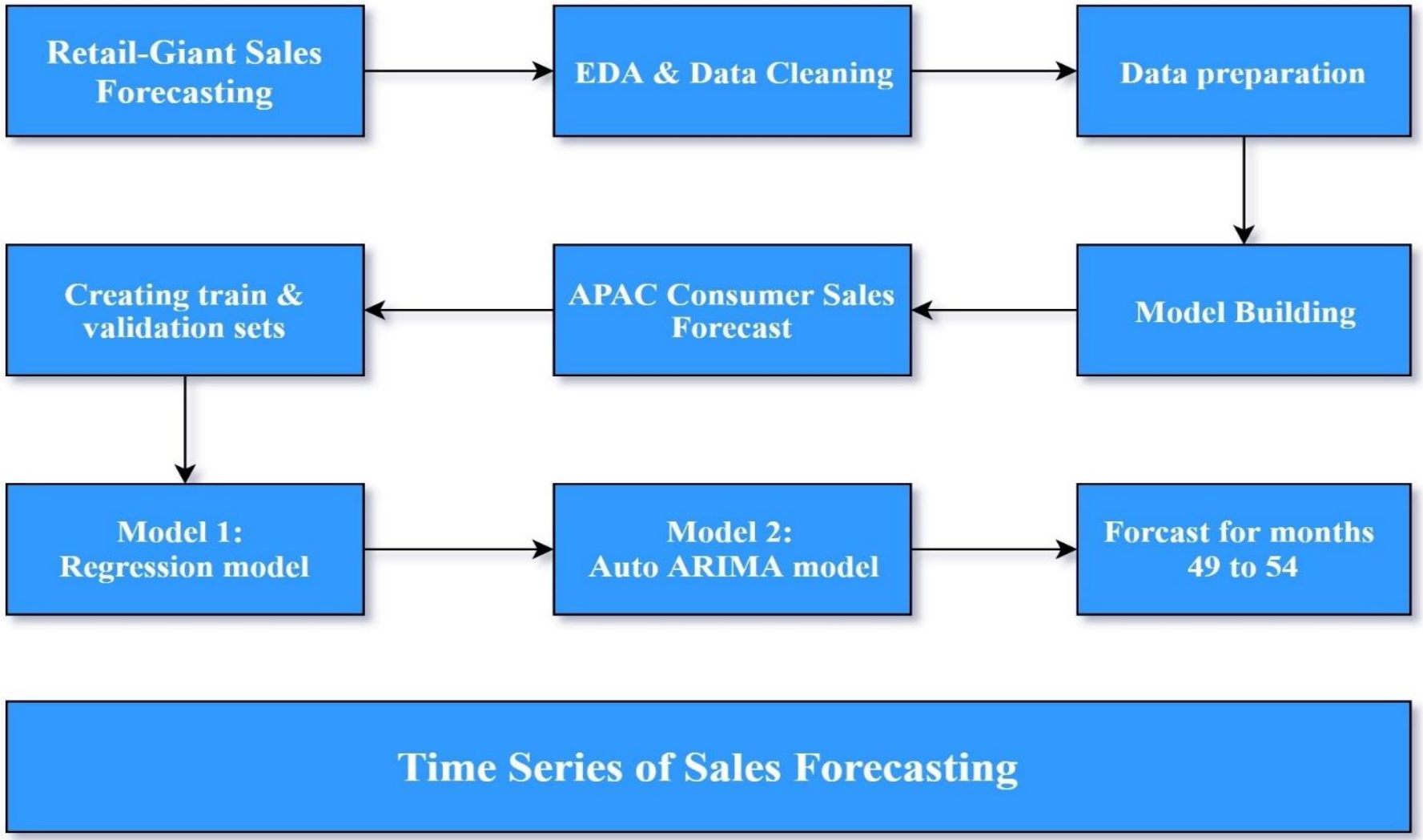
- Create time series of top Aggregated Data
- Smoothen time series to identify trend & seasonality
- Creating train & validation sets of size 42 & 6 months
- Build Model 1: Regression Model
- Build Model 2: Auto ARMA Model

Model Evaluation

- Evaluate model on validation set using MPSE
- Choose best model out of Model 1 & Model 2 using MPSE
- For Auto ARIMA plot ACF of residuals to check it resembles white noise

Forecasting

- Use best model to forecast future 6 months Sales.
- Repeat same for both Market-Segments separately for Sales & Quantity
- Prepare 4 forecasts for top 2 segments by Sales & Quantity


Retail-Giant Sales Forecasting
EDA & Data Cleaning
Data preparation
Model Building
APAC Consumer Sales Forecast
Creating train & validation sets
Model 1: Regression model
Model 2: Auto ARIMA model
Forcast for months 49 to 54
Time Series of Sales Forecasting

Analysis: Data Preparation & EDA

1. Created 21 data subset buckets based on Market & Segment they belong.
2. Aggregated data in each bucket by Sales, Quantity & Profit.
3. Calculated Coefficient of Variation(CV) using aggregated Profit for each Market-Segment using below:

   CV = sd(Profit)*100/mean(Profit)

4. Using CV & Profit found Top 2 most profitable Market-Segments as APAC_Consumer & EU_Consumer with below values:

Market	Segment	Sales	Profit	CV
EU	Consumer	1529716.24	188687.707	471.8084
APAC	Consumer	1816753.70	222817.560	420.6702

5. Aggregated data by Market-Segment & Order Month.

Analysis: Building Regression and Auto ARIMA Models

1. Created time series for aggregated data of EU_Consumer & APAC_Consumer subsets for first 48 months:
2. **ts(APAC_Consumer_Agg$Sales,frequency=12,start=c(2011,1),end=c(2014,12))**
2. Smoothened time series using Moving Average method, also tested Holt Winters smoothing.
3. Time series data was divided into train(1-42 month), validation(43-48 month) & test sets(49-54 month).
4. Model 1: Linear model : Created using tslm() function from forecast package in R on train data.
5. Model 2: Auto Arima Model: Created using auto.arima() function from forecast package in R train data
6. Both models were evaluated using Mean Absolute Percentage Error(MAPE).
7. ACF plots of residuals were used check that it resembles white noise.

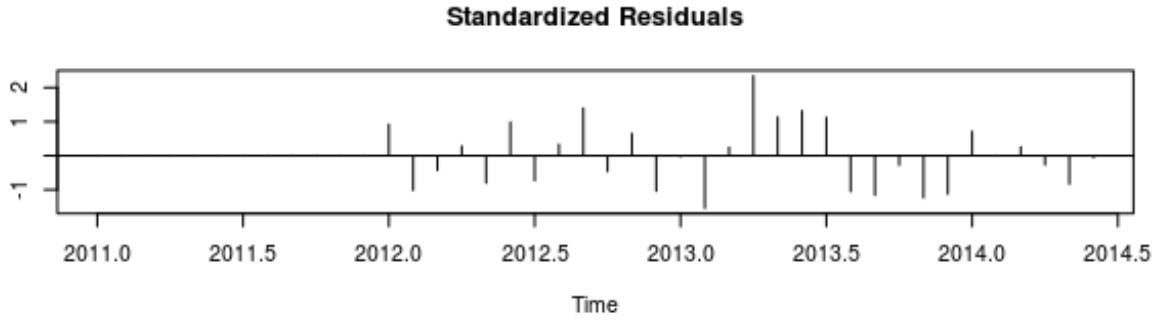
Analysis: Model Evaluation

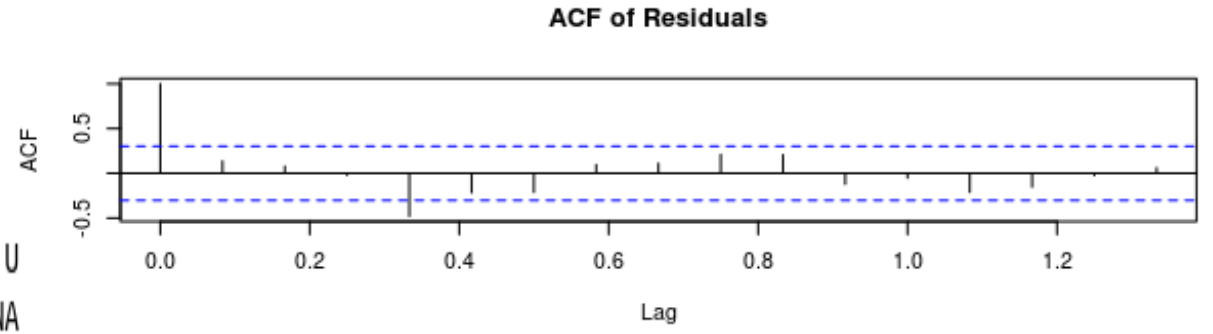
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	8940.623	13561.35	9883.642	11.61695	13.54141	0.3648822	1.900914

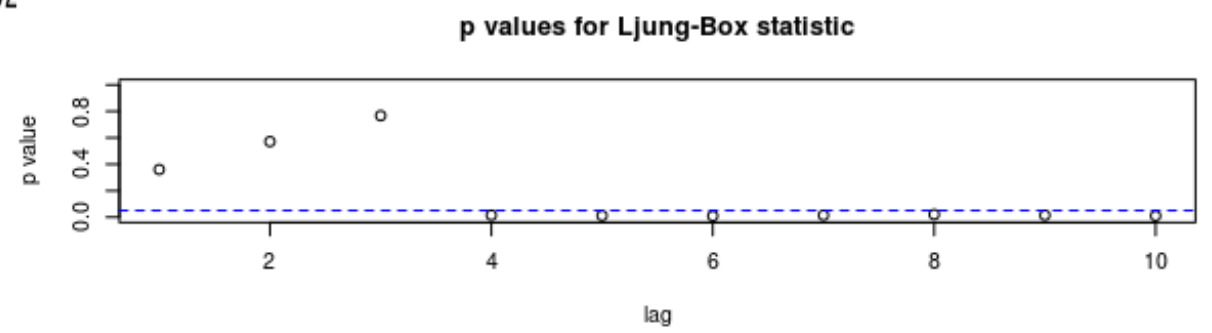
a)MAPE and other measures for APAC Consumer time series Linear Model

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-25.89145	2883.803	2041.564	-0.4648203	5.52011	0.2567916	0.1363166	NA
Test set	9518.65946	14630.005	10576.355	12.3253908	14.48387	1.3303127	0.3522145	2.046502

b)MAPE and other measures for APAC Consumer time series Auto ARIMA Model

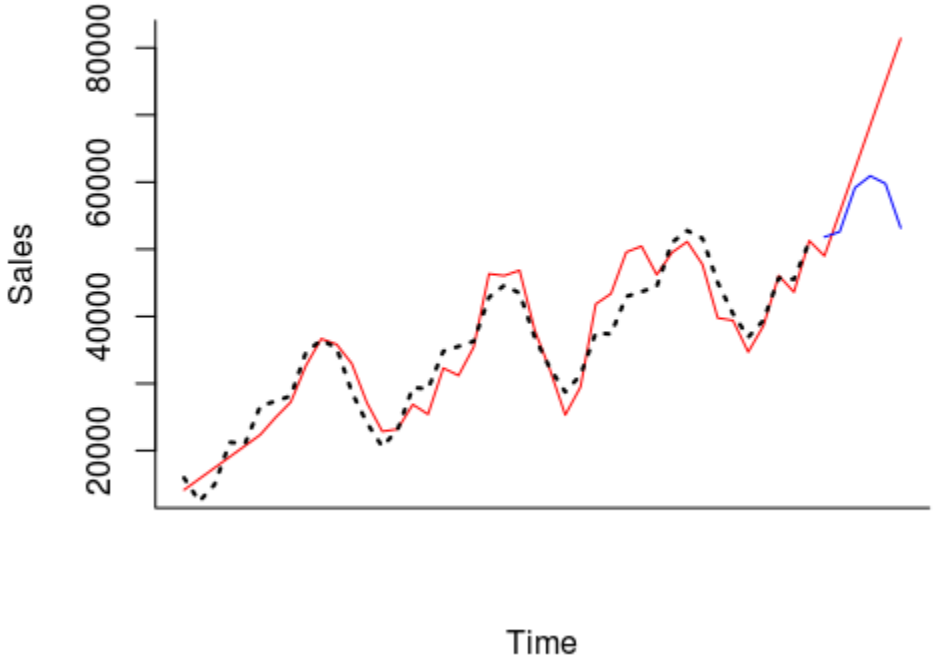




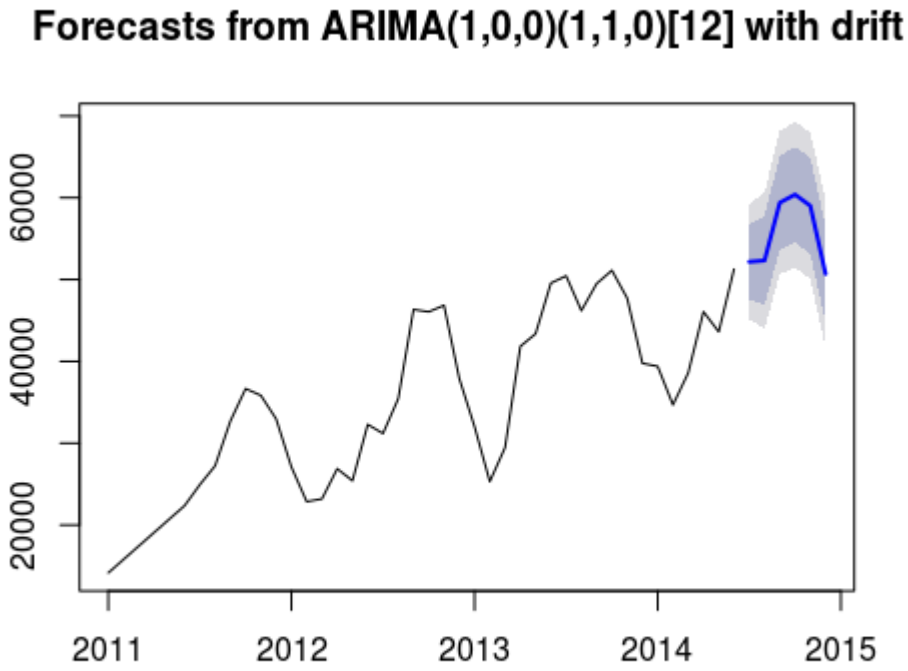


c) ACF of Residuals for APAC Consumer Sales time series

Results: 1a. APAC Consumer Sales Forecast on validation set

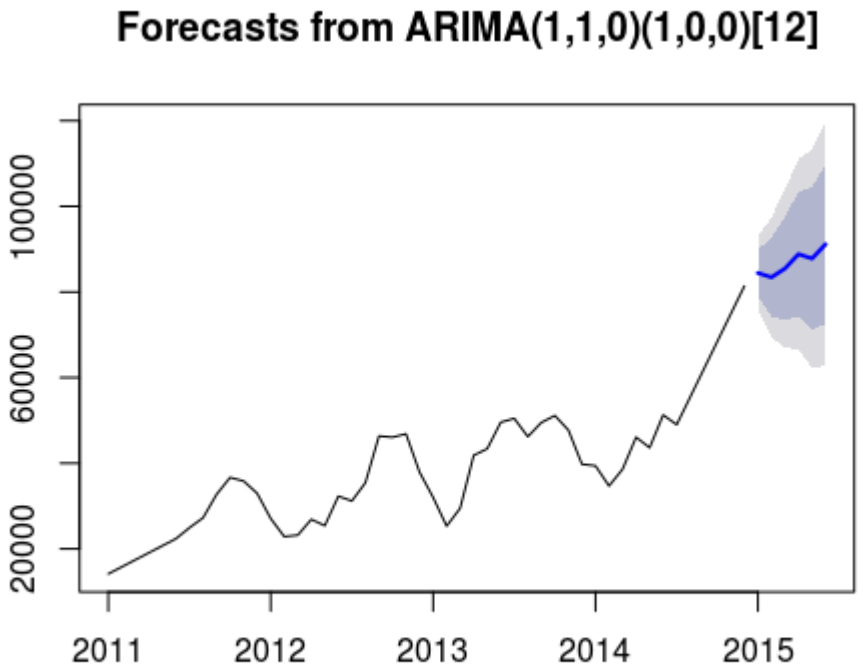


a) Linear Model Forecast



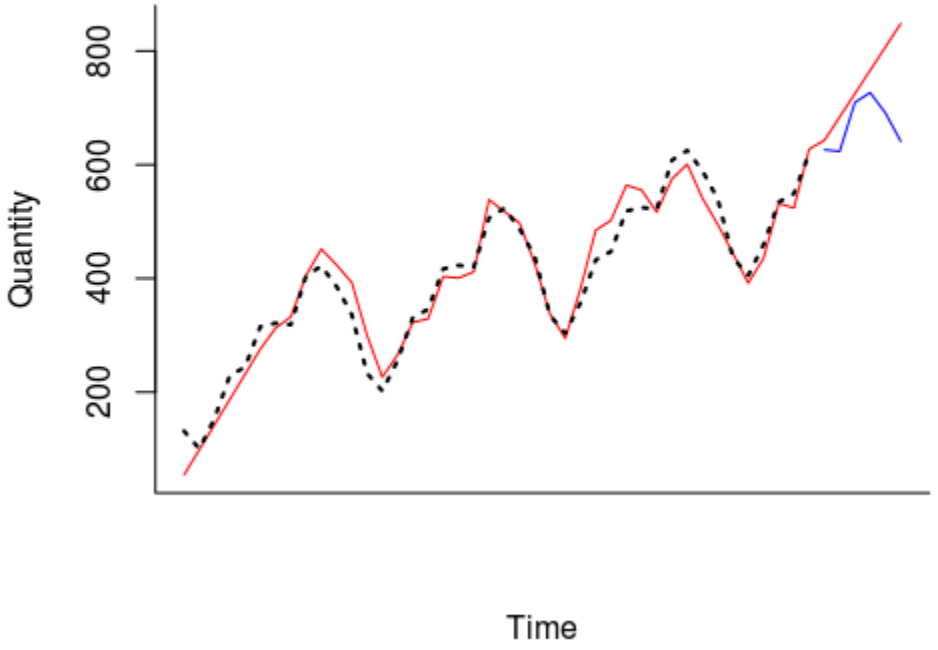
b) Auto ARIMA Model Forecast

Results: 1b. APAC Consumer Sales Forecast on test set

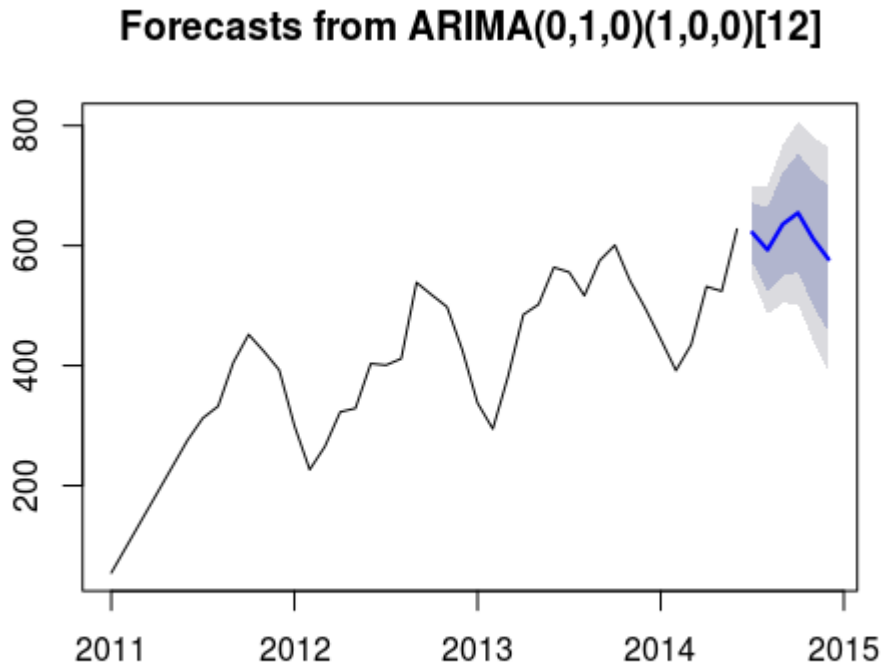


a)) Auto ARIMA Model Forecast

Results: 2a. APAC Consumer Quantity Forecast on validation set

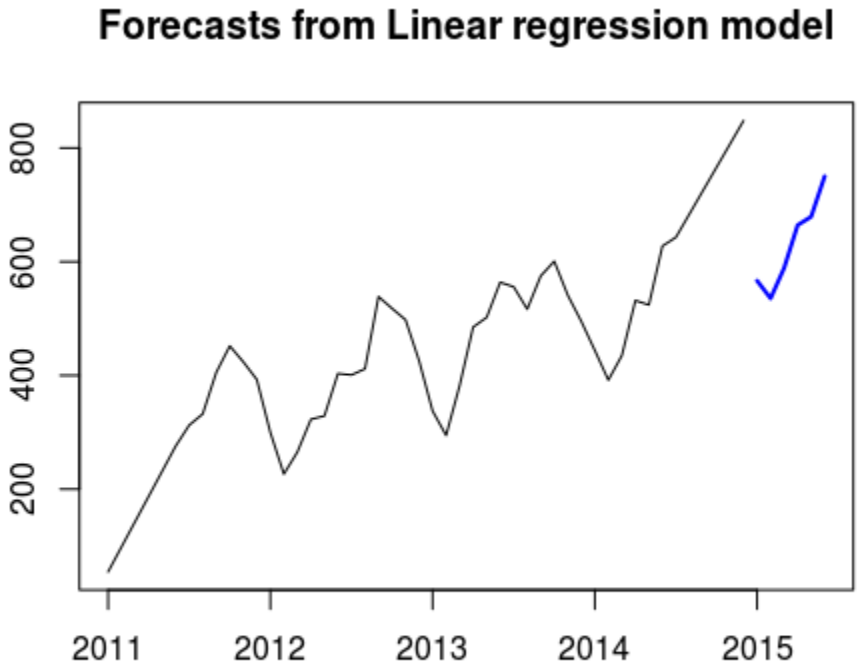


a) Linear Model Forecast



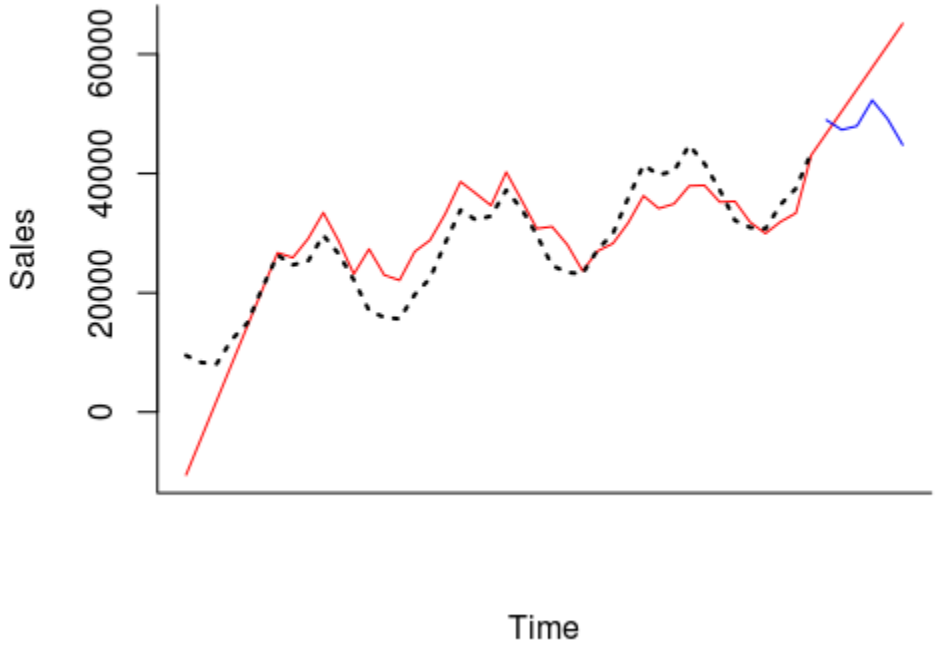
b) Auto ARIMA Model Forecast

Results: 2b. APAC Consumer Quantity Forecast on test set

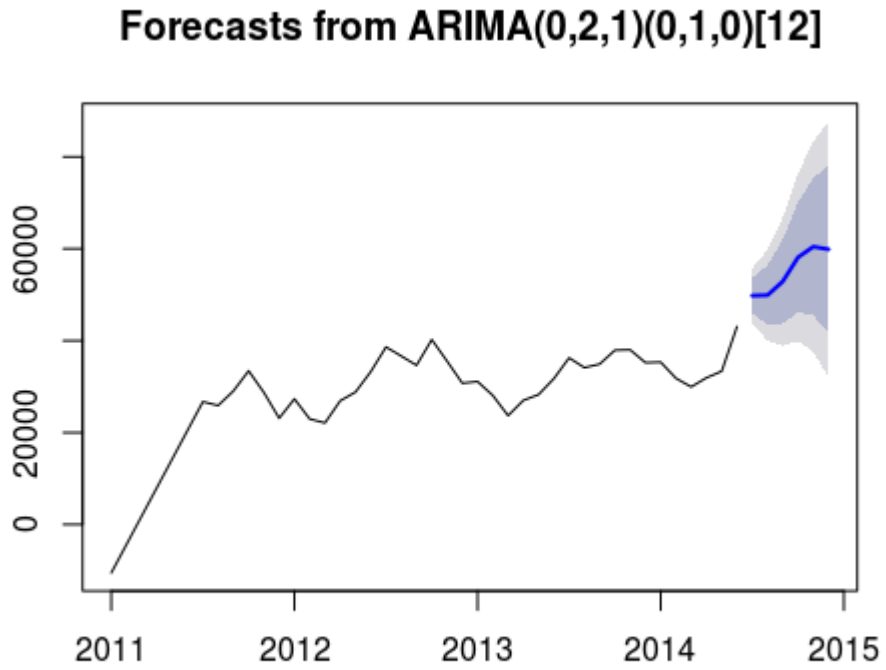


a) Linear Model Forecast

Results: 3a. EU Consumer Sales Forecast on validation set

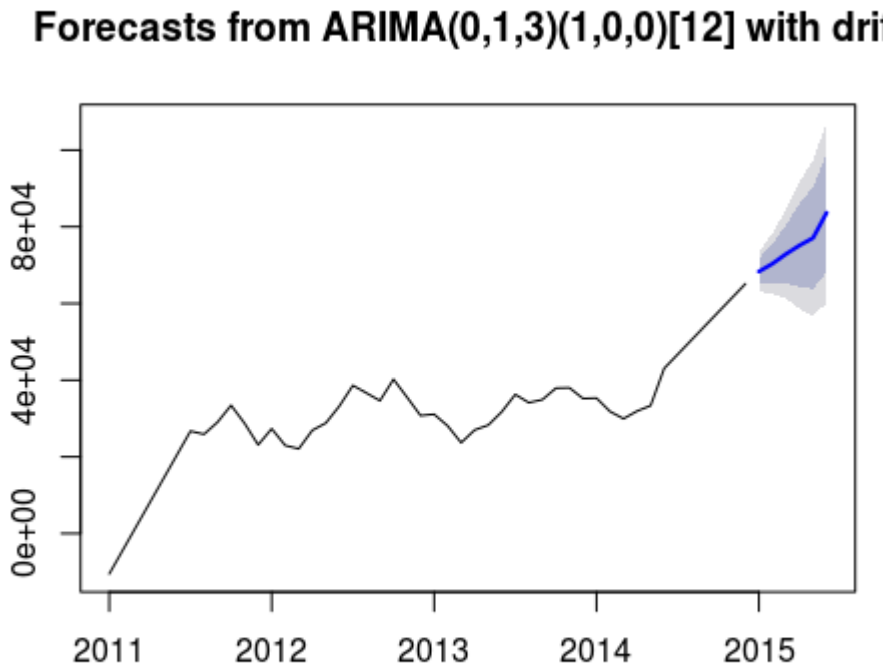


a) Linear Model Forecast



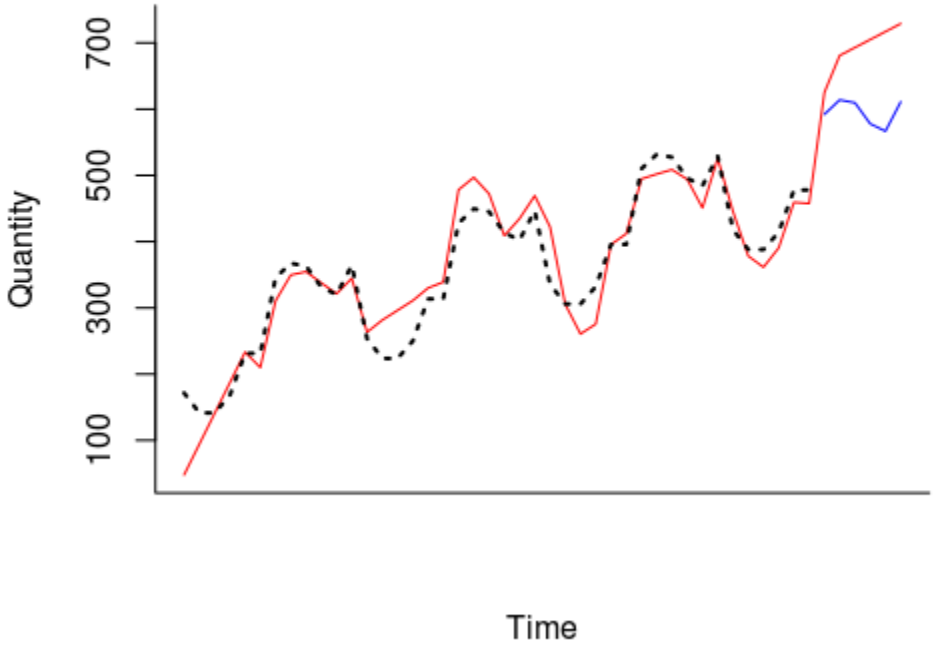
b) Auto ARIMA Model Forecast

Results: 3b. EU Consumer Sales Forecast on test set

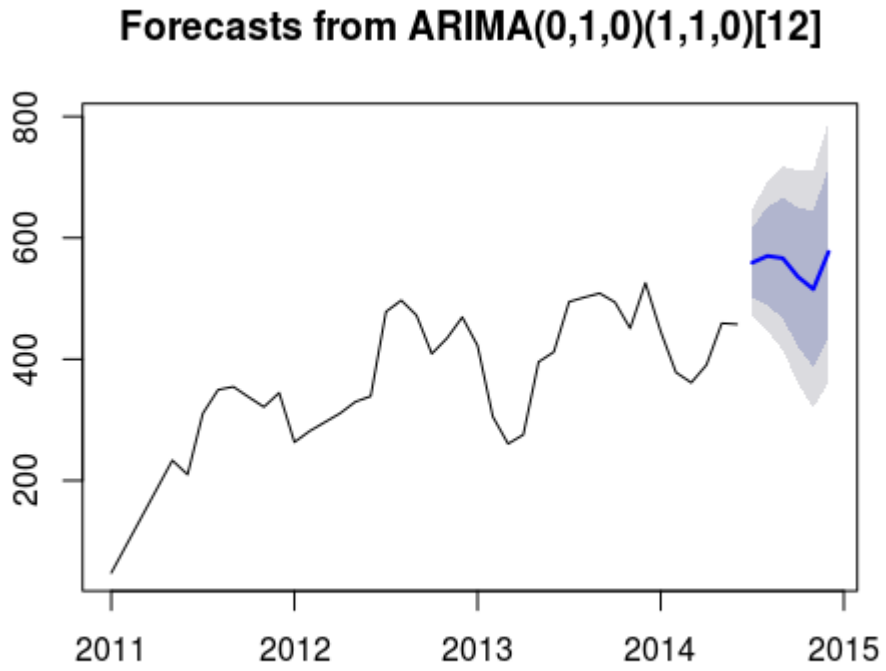


a)) Auto ARIMA Model Forecast

Results: 4a. EU Consumer Quantity Forecast on validation set

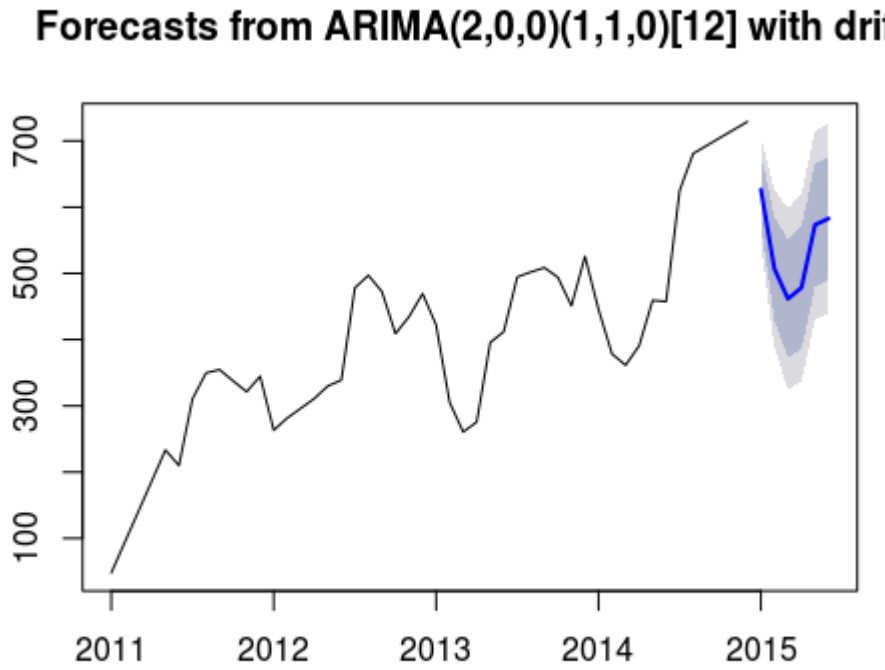


a) Linear Model Forecast



b) Auto ARIMA Model Forecast

Results: 4b. EU Consumer Quantity Forecast on validation set



a)) Auto ARIMA Model Forecast

Conclusions

1. Based on data provided we helped “Global Mart” in identifying 2 most profitable market segments as APAC Consumer and EU Consumer.
2. We created total 8 forecasting models for top 2 segments out of which 4 best were selected for forecasting future 6 months sales & quantity for months January 2015 to June 2015.
3. Below is summary of 4 key forecasts on test data(Jan – June 2015):
   a) APAC Consumer Sales is likely to rise in next 6 months with small fluctuations.
   b) APAC Consumer is also likely to rise steeply in coming 6 months.
   c) EU Consumer Sales may show slow rise in coming months.
   d) EU Consumer Quantity is likely to drop during initial 1 or 2 months & then rise rapidly in next 3 months, eventually reaching a plateau.