Ecommerce Capstone Project

SUBMISSION

1. Mruthyunjaya Kumar

2. Dharmanandana Reddy

3. MuthuKumaran SM
4. Deepak Kumar
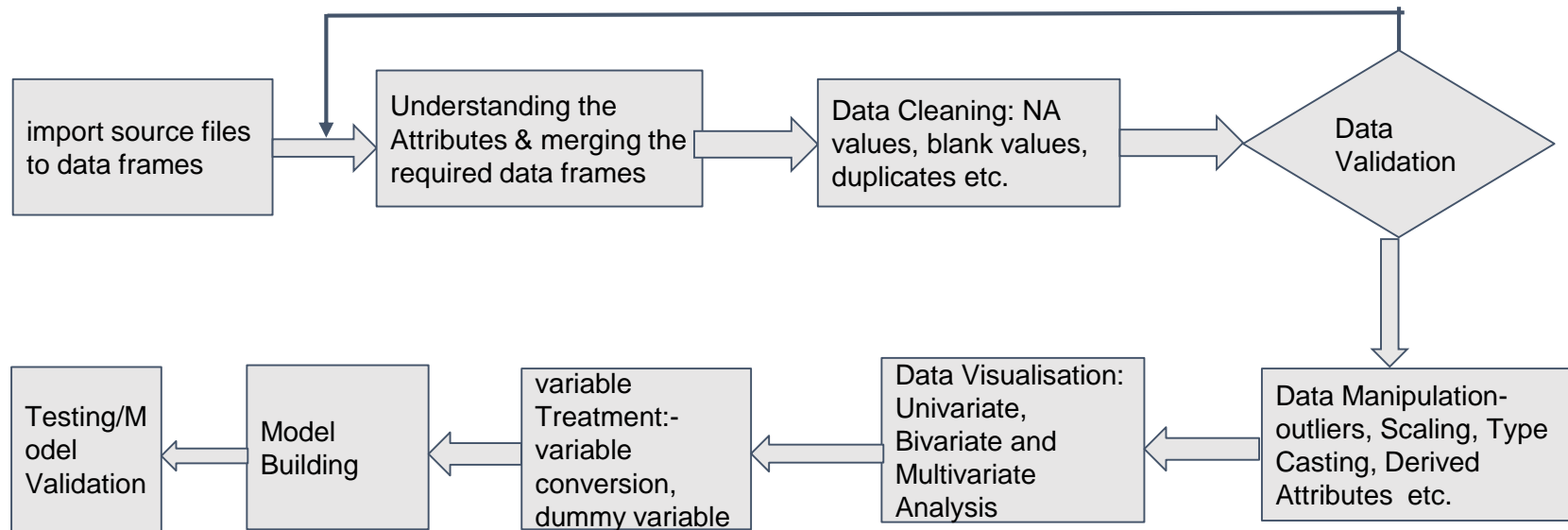
Ecommerce Capstone Project - Problem Statement

Eleckart is an e-commerce firm. Last year they spent money in marketing and they offered promotions as well but the CFO feels that marketing and promotions was not sufficiently impactful.
Hence this year they expect better budget allocation and they want improvement in revenue response.

Objective:

- So we need to develop a Market mix model to observe the actual impact of different marketing variables over the last year.
- We need to recommend the optimal budget allocation for different marketing levers for the next year.
- To create market mix models for 3 product sub-categories - camera accessory, home audio and gaming accessory. Also, the models have to be built at a weekly level

Problem solving methodology

```
┌──────────────┐      ┌──────────────────┐      ┌──────────────────┐      ◆
│ import source│      │ Understanding the│      │ Data Cleaning: NA│    Data
│ files to data│ ──►  │ Attributes &     │ ──►  │ values, blank    │ ──► Validation
│ frames       │      │ merging the      │      │ values,          │      ◆
└──────────────┘      │ required data    │      │ duplicates etc.  │
                      │ frames           │      └──────────────────┘
                      └──────────────────┘
```

import source files to data frames → Understanding the Attributes & merging the required data frames → Data Cleaning: NA values, blank values, duplicates etc. → Data Validation

Testing/Model Validation ← Model Building ← variable Treatment:- variable conversion, dummy variable ← Data Visualisation: Univariate, Bivariate and Multivariate Analysis ← Data Manipulation- outliers, Scaling, Type Casting, Derived Attributes etc.

Data Understanding

- We have got below datasets -
  Consume Electronics - This file has order details of each product from June 2015 to July 2016.
  Product List - This file contains details of product and frequency it was sold.
  Media Investment - It provides details of money spent on different marketing channels at monthly level.
  Special Sale Calendar - It provides details of sale (date and event) when promotion when promotion was given.
  Monthly NPS Score - This files contains customer satisfaction score monthwise.
  Product Details - It provide category, subcategory, vertical level details for a product.

- Collected relevant data for analysis
- Checked the outlier Values present in data
- Observed correlation between different variables
- Checked the trends of different variables (increasing/decreasing)
- Explored the data by using different plots and did univariate, bivariate and multivariate analysis
- Created plots to understand range of values

Data Preparation and Data Quality

- Identified and treated the columns having na, negative and outlier values
    - Pin, customerId, GMV had missing values.
    - Deliverybybdays, deliverybycdays, customerId, pincode, Product_procurement_sla had negative value.
- Treated gmv values where value was 0 or MRP value was higher than gmv value.
- Formatted the order_date to match our defined format for further analysis.
- Removed the records that are not in our analysis period "July 15 to June 16"
- Calculated the list price (unit price) of each product from gmv
- Aggregated the order to weekly level from daily level and monthly level
- Calculated the discount for each unit
- Frame a holiday date for analysis
- Extracted the records from the holiday period
- Extracted the records based on the three sub-category Camera_accessory, home_audio and gaming_accessory
- Merged different files into one single file for analysis
- Created derived variables

Understanding the trend of sale
(Home Audio)

For Home Audio we can see that sudden changes are happening in the total revenue in different week numbers. In few weeks revenue is very high but in other weeks it is too low. We can say due to some attributes there is sudden change in the revenue.


Total GMV vs Week for Home Audio - median GMV as horizontal
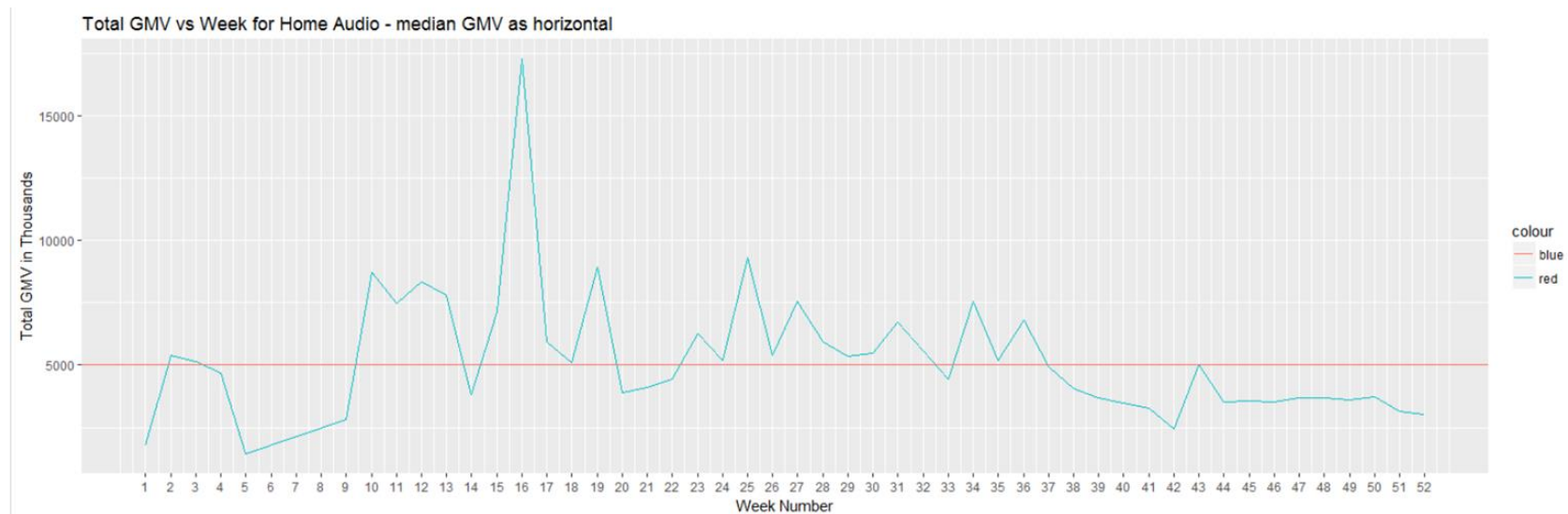
Understanding the trend of sale
(Gaming Accessories)

For Gaming Accessory we can see that sudden changes are happening in the total revenue in different week numbers. In few weeks revenue is very high but in other weeks it is too low. We can say due to some attributes there is sudden change in the revenue.

Total GMV vs Week for Gaming Accessory - median GMV as horizontal
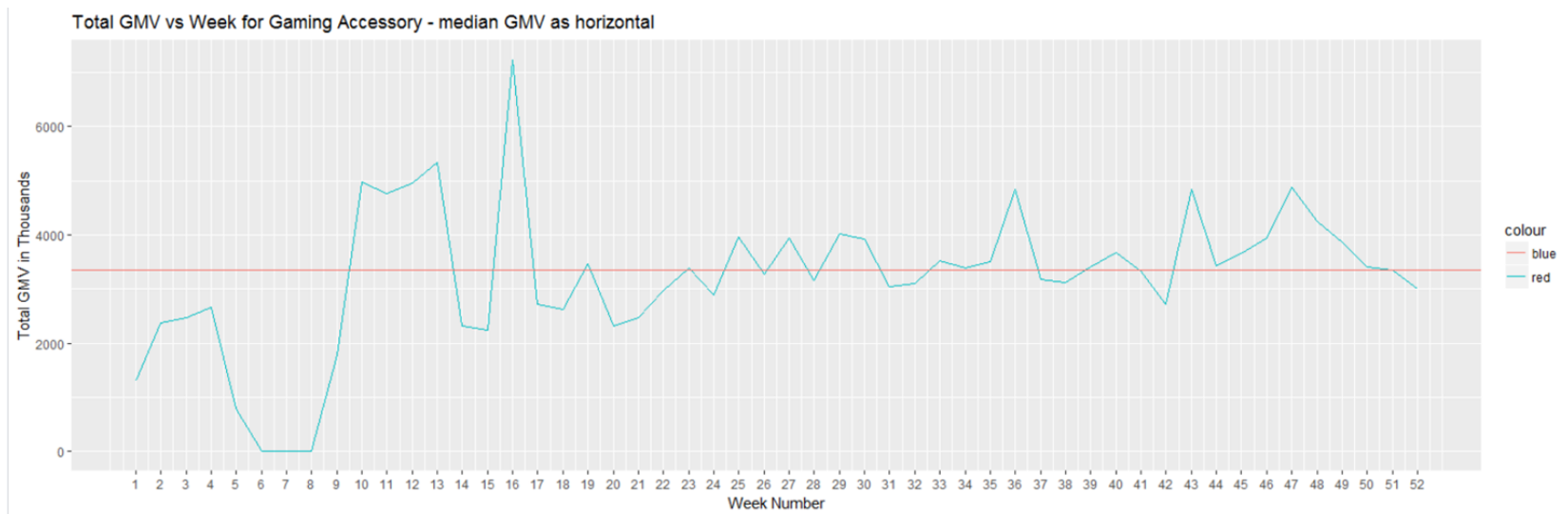
Understanding the trend of sale
(Camera Accessories)

For Camera Accessory we can see that sudden changes are happening in the total revenue in different week numbers. In few weeks revenue is very high but in other weeks it is too low. We can say due to some attributes there is sudden change in the revenue.

Total GMV vs Week for CameraAccessory - median GMV as horizontal
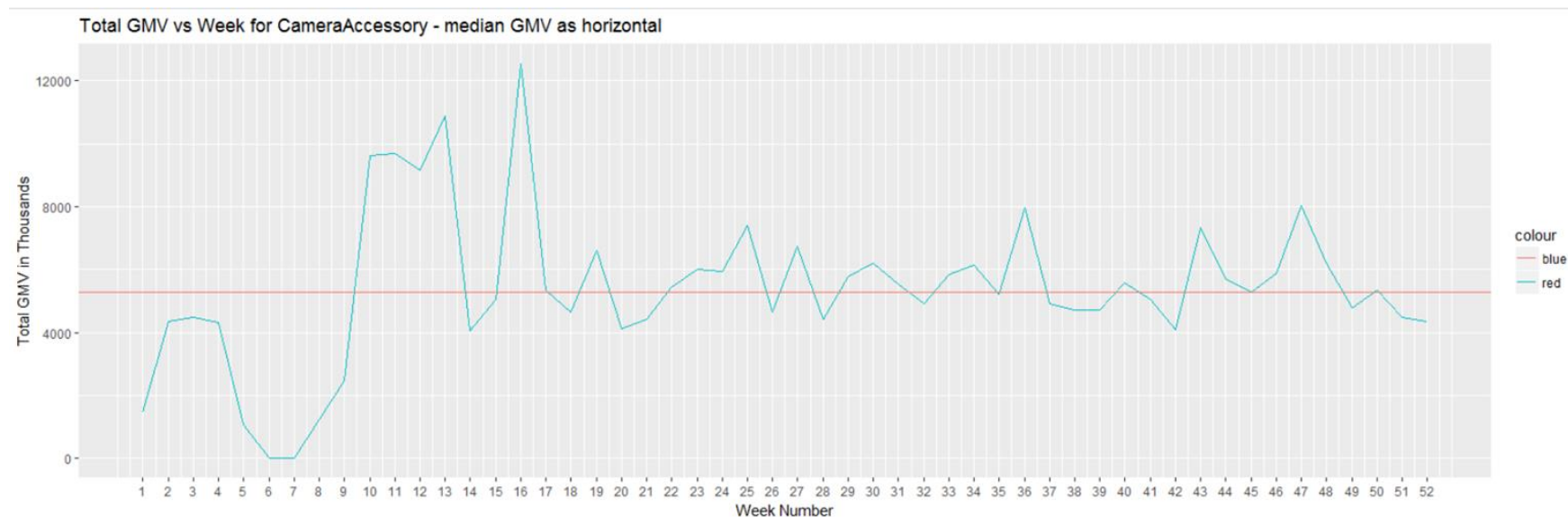
For Home Audio, We can say that whenever promotion is given during festival (Dussehra, Christmas etc.) then the revenue is very high but if the promotions are given frequently on other days then it is not much impactful. When promotion was given on 15/10/2015 then revenue is very high. Similarly on 20/1/2016, 20/2/2016 revenue is increased.

During festivals like Dussehra, Christmas people like to listen song and due to promotions they purchase it. So sale is too high.

Saleday Total GMV for HomeAudio
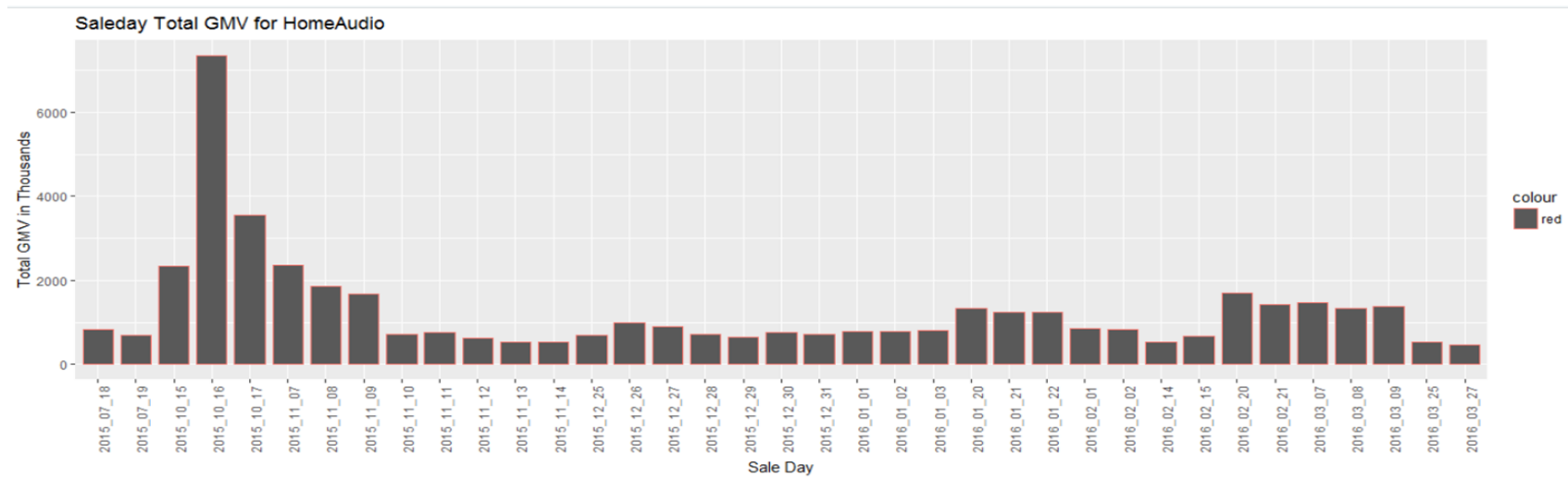
For Gaming Accessory, We can say that the revenue is very high when the promotion is given during Dussehra, republic Day, BSD but on other days it is not much impactful.

Seems too many people love (specially young age) playing games during these holidays.

Saleday Total GMV for GamingAccessory
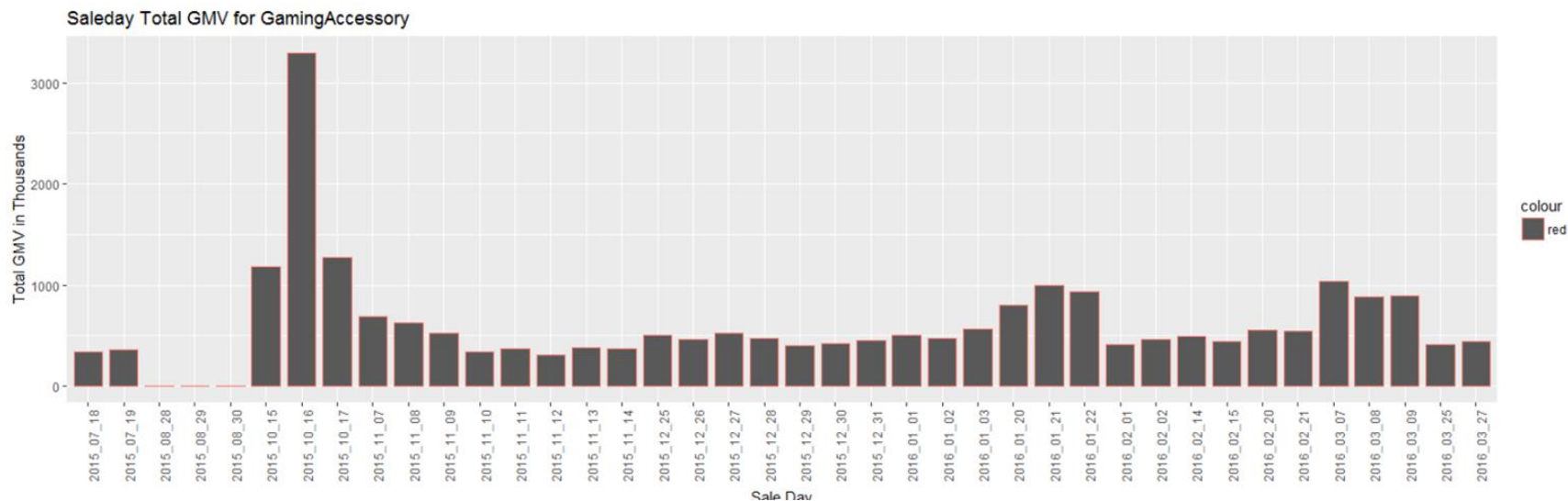
For Camera Accessory, We can see that the revenue is very high when the promotion is given during Dussehra, republic Day, BSD but on other days it is not much impactful.

So the promotions given during festival will have very high impact on the revenue.
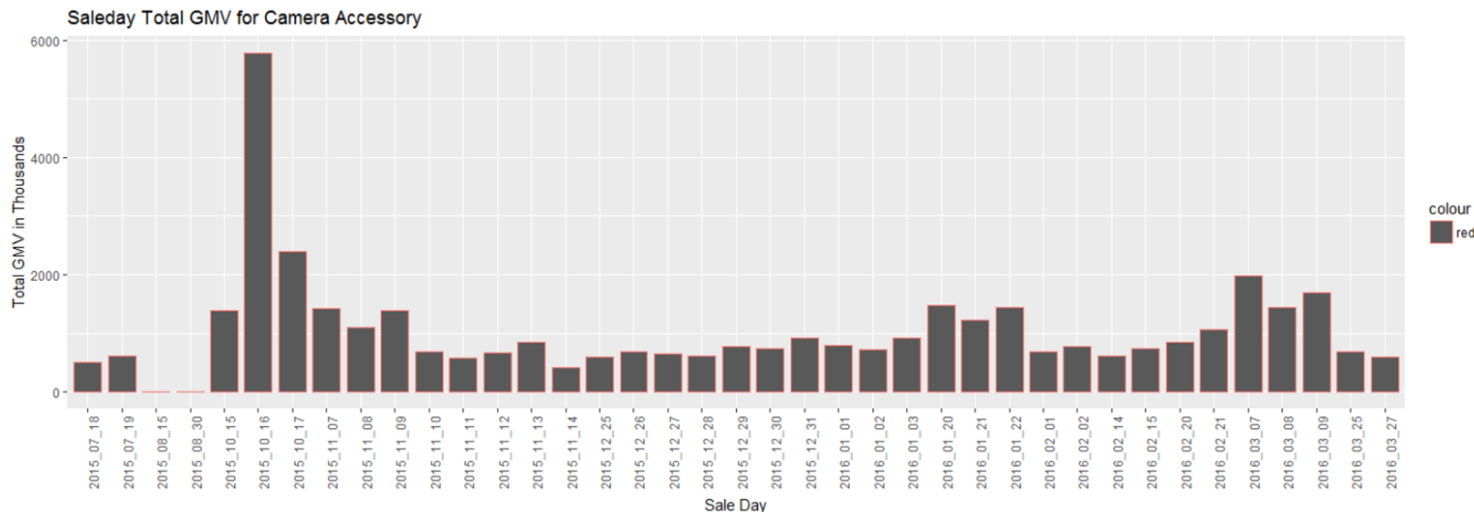
Saleday Total GMV for Camera Accessory

Looking at the plot we can say that
When discount increases the count of sale is
increasing for the "Camera Accessory".
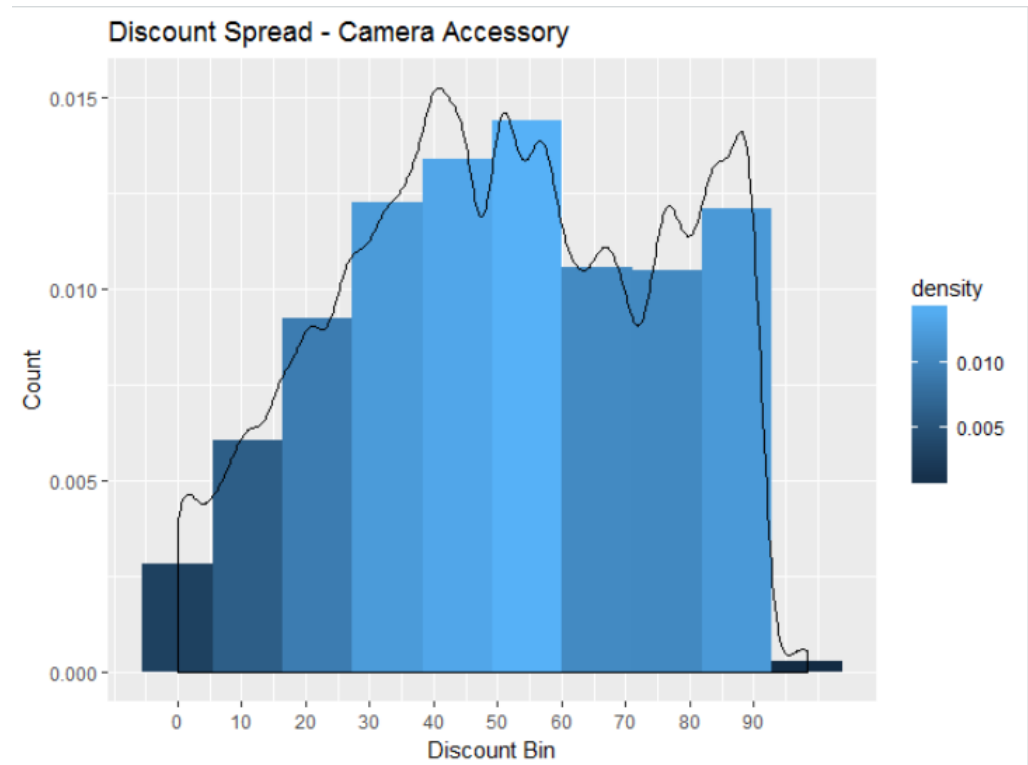
Discount Spread - Camera Accessory

Looking at the plot we can say that

When discount increases the count of sale is

increasing for the "Home Audio".
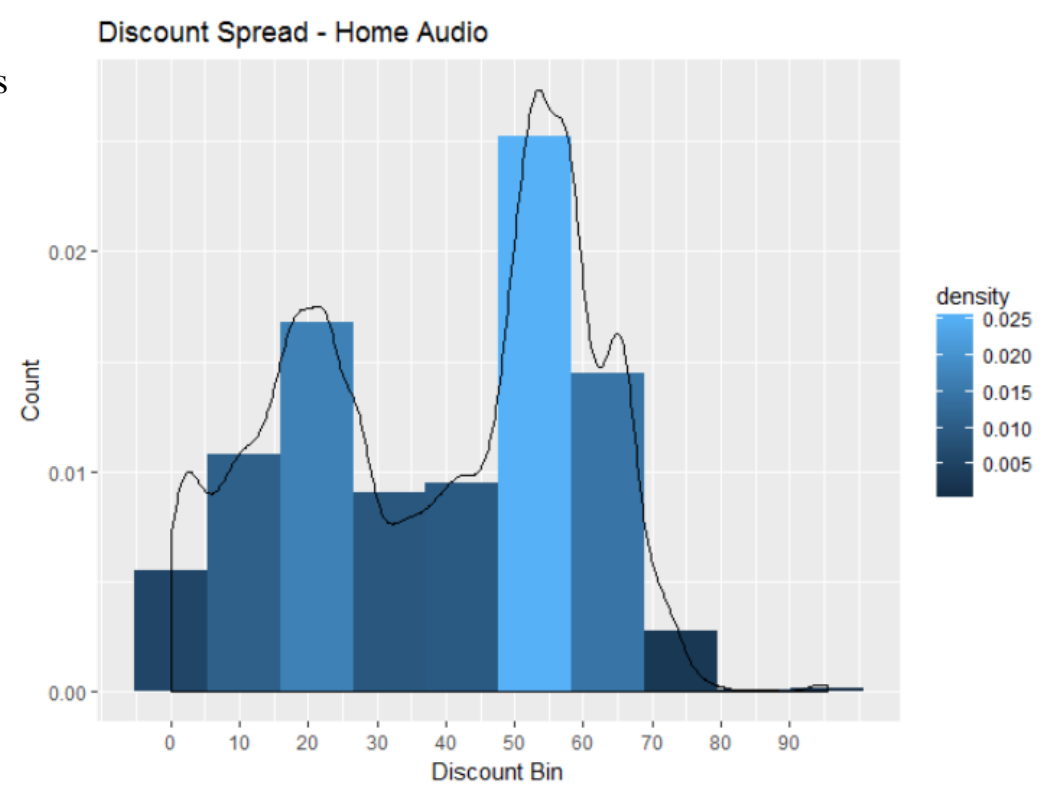
Looking at the plot we can say that
When discount increases the count of sale is
increasing for the "Gaming Accessory"

Discount Spread - Gaming Accessory
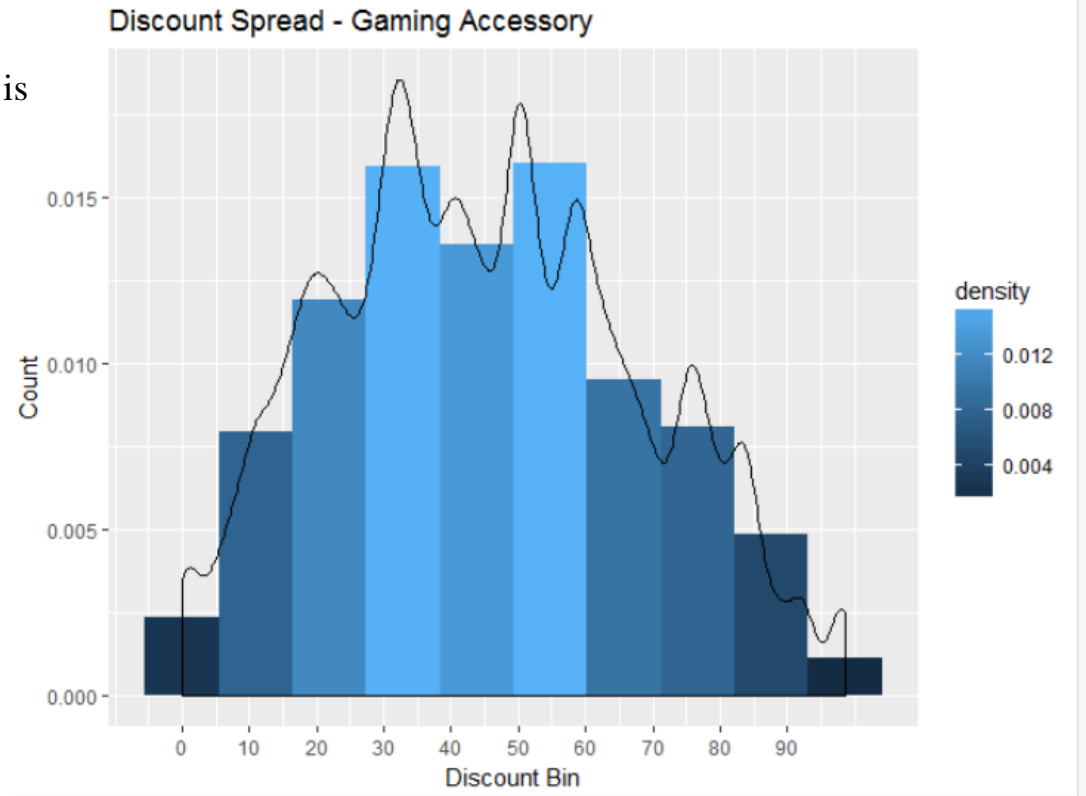
We can see that with the decorated price the revenue is higher for all the 3 categories. So price decoration has high impact on the revenue generated.

Price Decoration impact on GMV
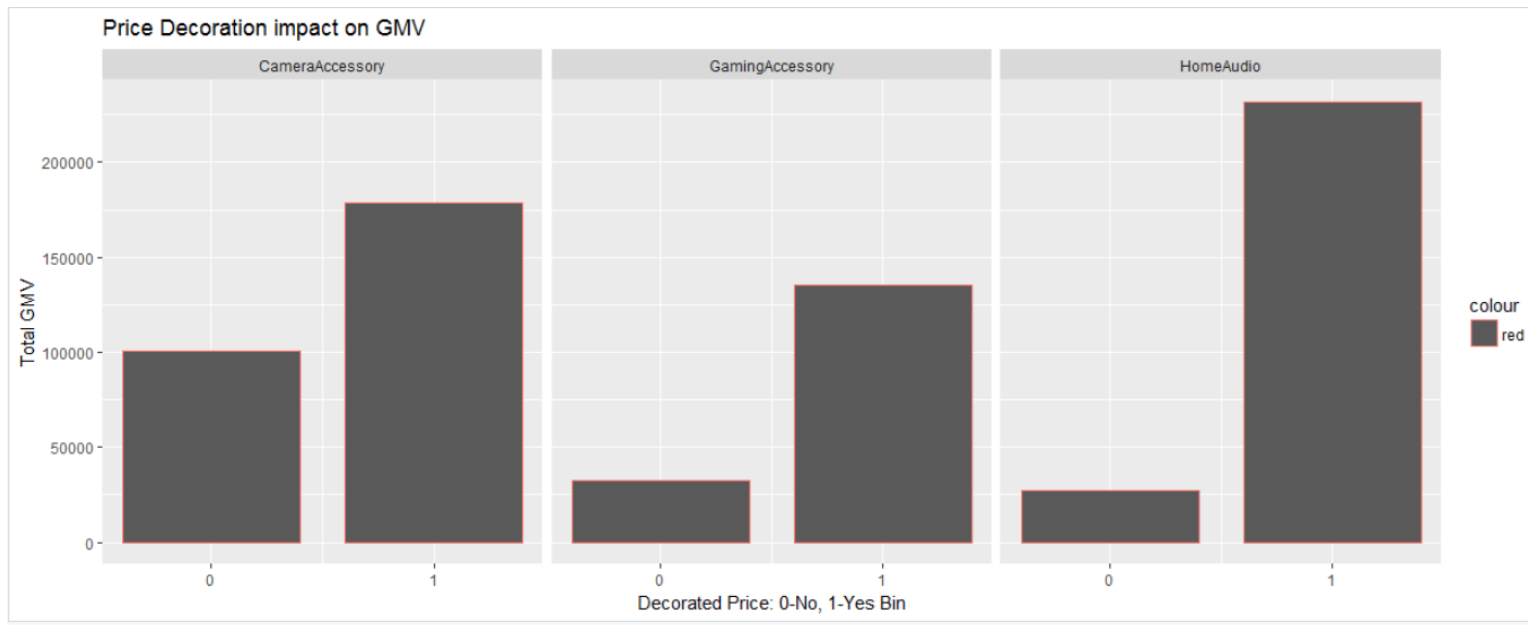
From the below plot we can say that expense is lowest for august

and expense is high in Sep,Oct,Dec,Mar
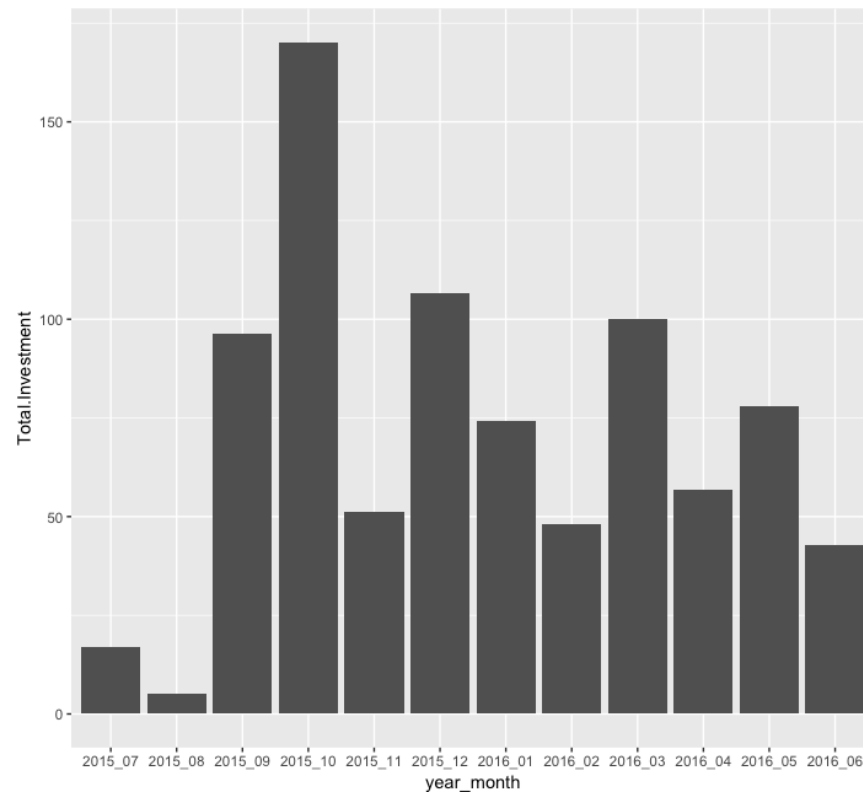
Model Building

We have multiple attributes contributing to the initial Model. Using stepAIC all insignificant attributes are removed. We have removed all the insignificant attributes one by one by checking their P value, significance of attribute and vif value.

To remove an attribute we have followed below steps

a. first check P value if it is greater than .05 then remove it
b. check whether the attribute is significant or not then remove it, 3 star (***) means it is more significant
c. check for vif value, if it higher then it can be removed
d. For Multiplicative Linear Model we have taken log of independent variables and used
e. For optimum lag value for distributed lag model, Auto regressive distributed lag Model, Koyck Model we have use finitedlmauto of dLagM library
f. For Models which are not satisfactory on multiple independent series we have done Model building and evaluation against GMV
g. For forecasting Transposed Matrix of Test data was used with multi independent series
h. We have checked the Residual value and error values for different models and have done tuning of attributes and lag values for it
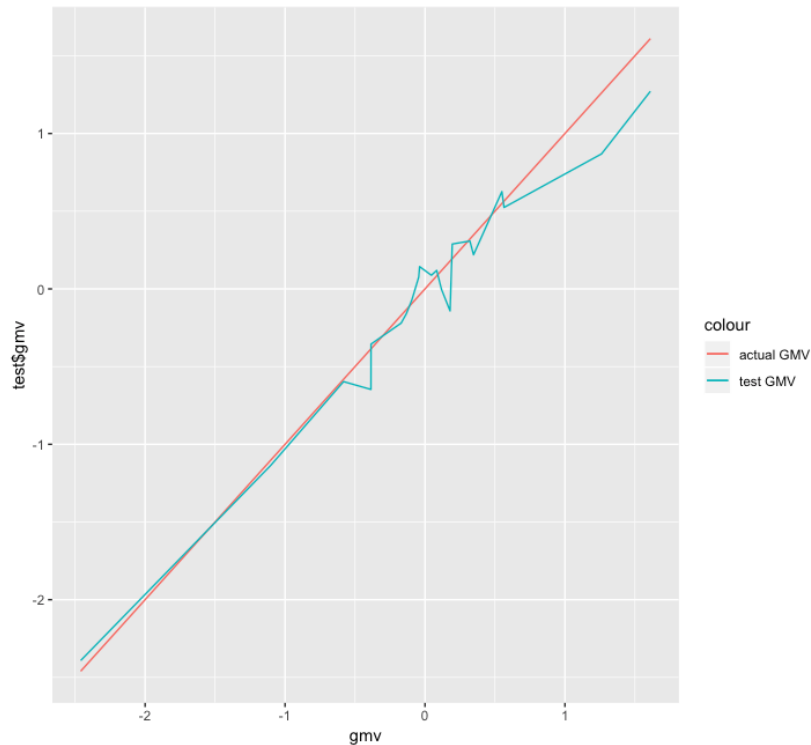
Camera Accessories – Five models outcome

Model	Significant variables	Multiple R square	Adjusted R square
1.Linear Model	weeknumber + units + discount + Percent + CameraBattery + CameraTripod + Flash + TV + Digital + Sponsorship	0.9662	0.9492
2.Multiplicative Linear Model	product_mrp + units + discount + Percent + CameraEyeCup	0.9912	0.9853
3.Distributed Lag Model	Units+CameraTripod	0.8982338	0.9372
4. Auto Regressive DL Model	Flash_CameraAccessory	0.7767	0.6985
5. Kyock Model	Affiliates, units	0.8341	0.8272

Model 1,2,3 are using multiple independent series for the dependent variable GMV
Model 4,5 are giving best values with the GMV compared against for specific features (1:1)

Linear Model : Camera Accessary

Multiplicative Linear Model : Camera Accessary

cameraAccessory - Multiplicative Model
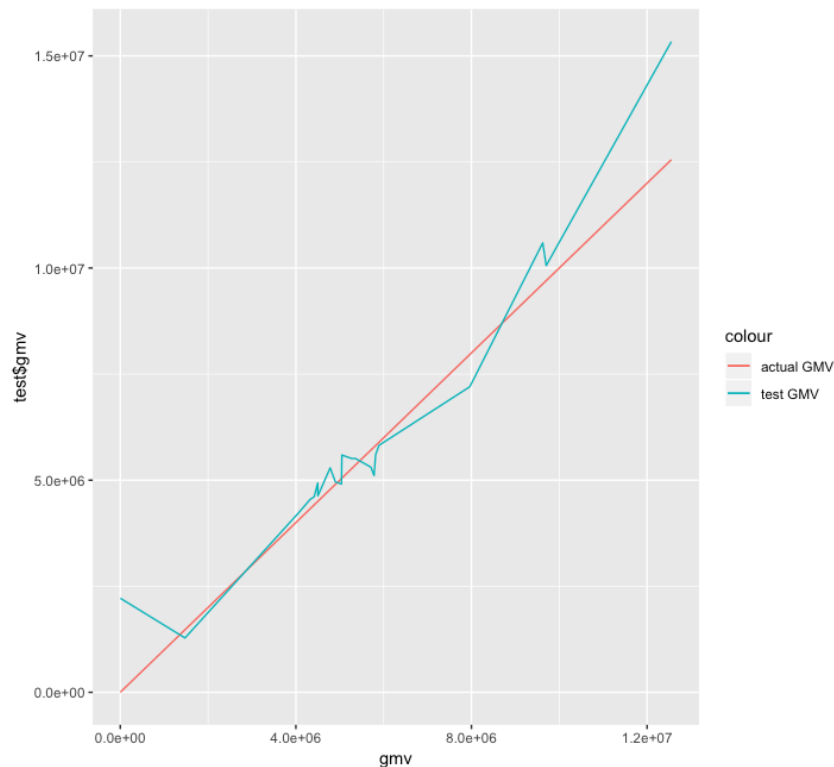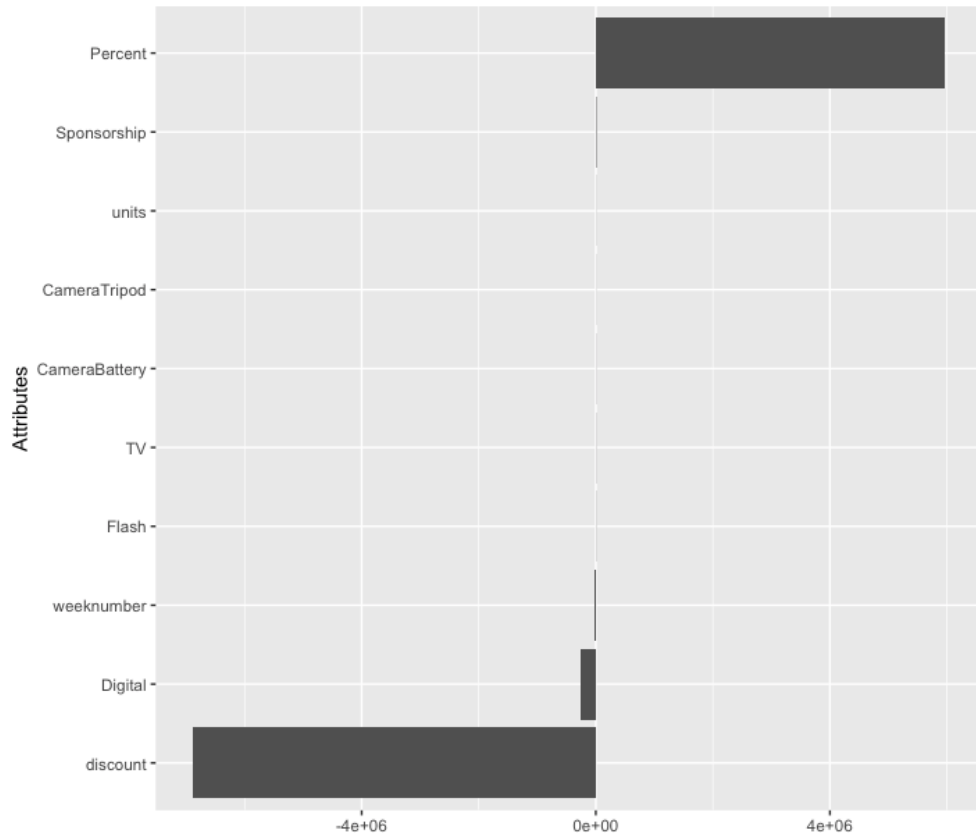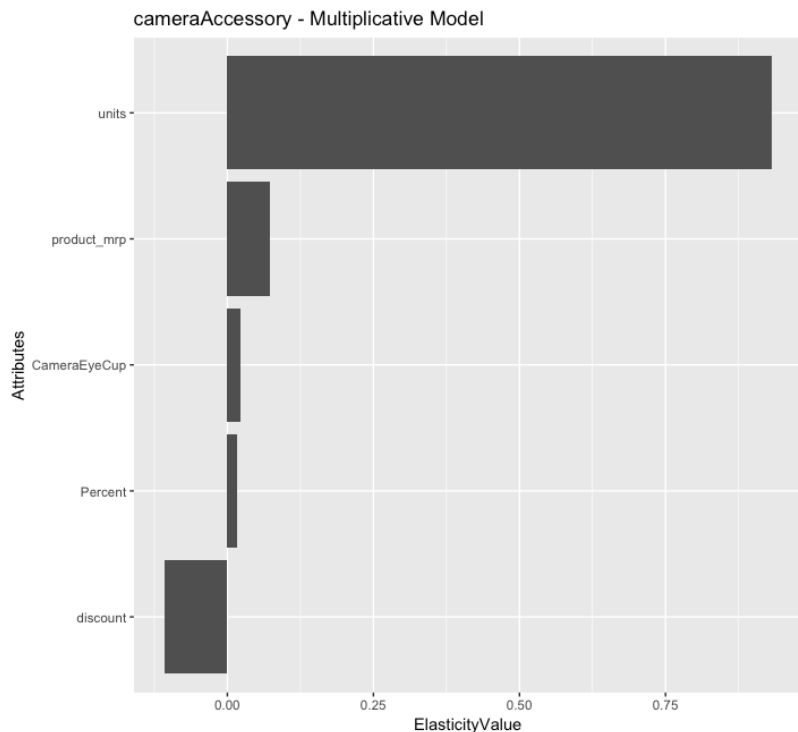
Distributed Lag Model : Camera Accessary
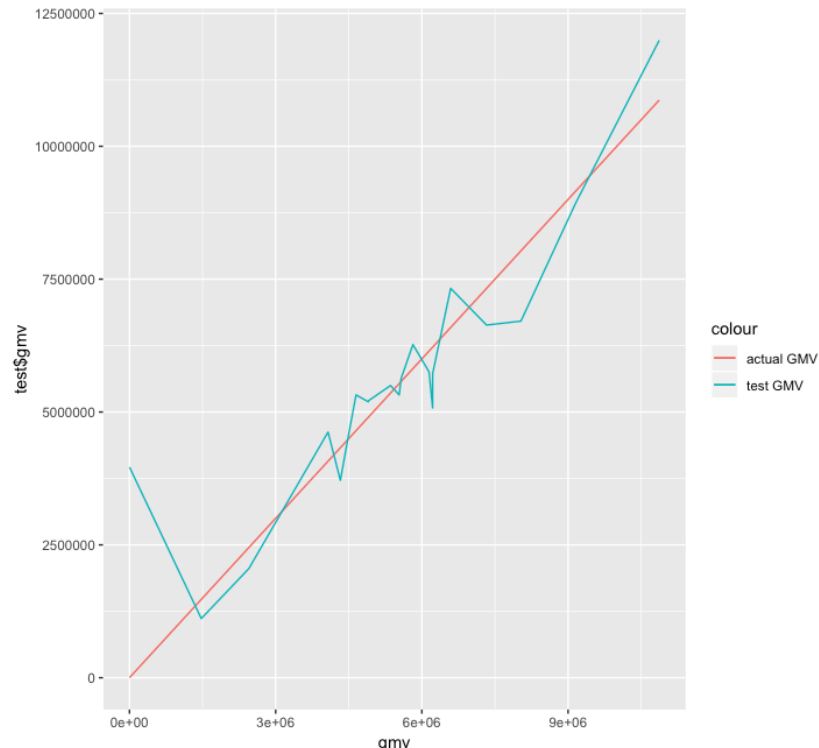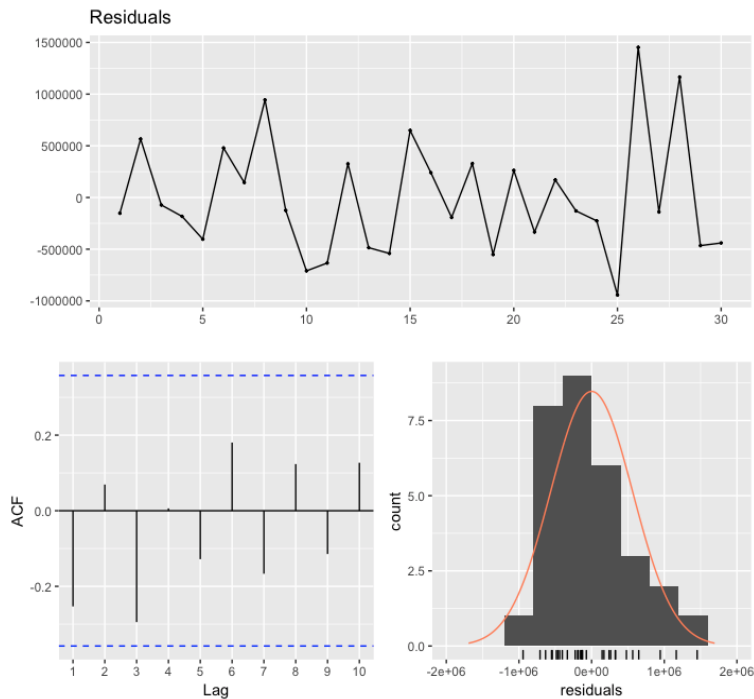
Auto Regressive DL Model : Camera Accessary
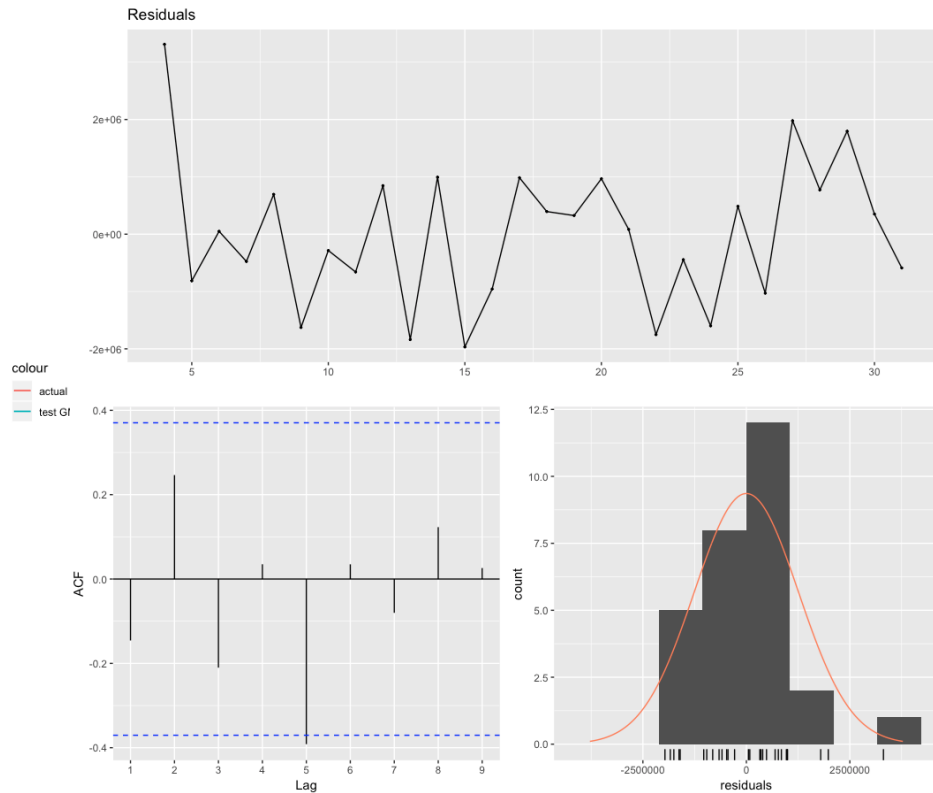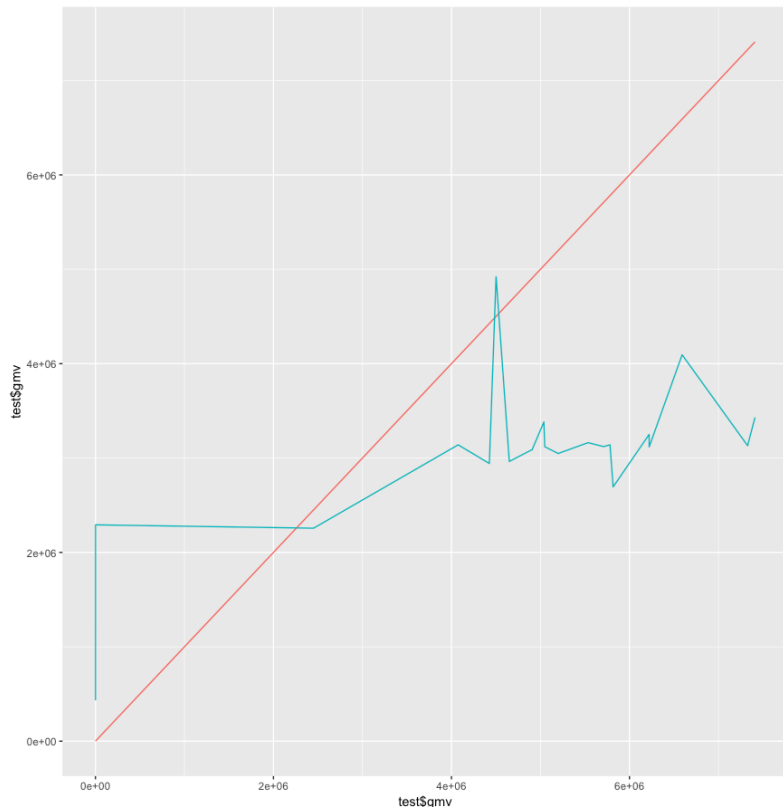
Kyock Model : Camera Accessary
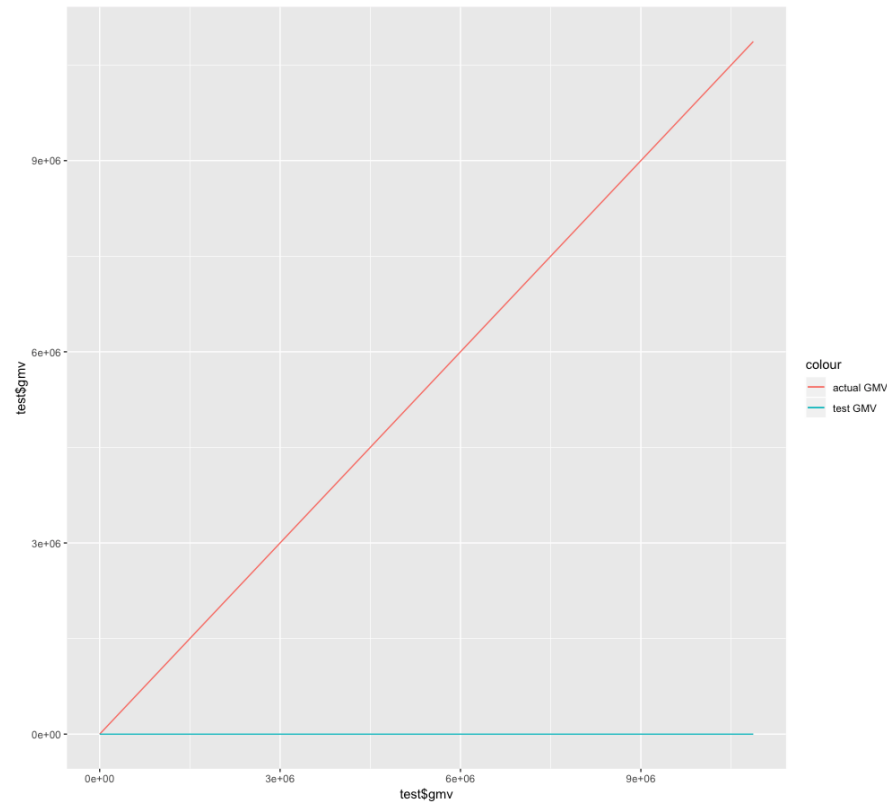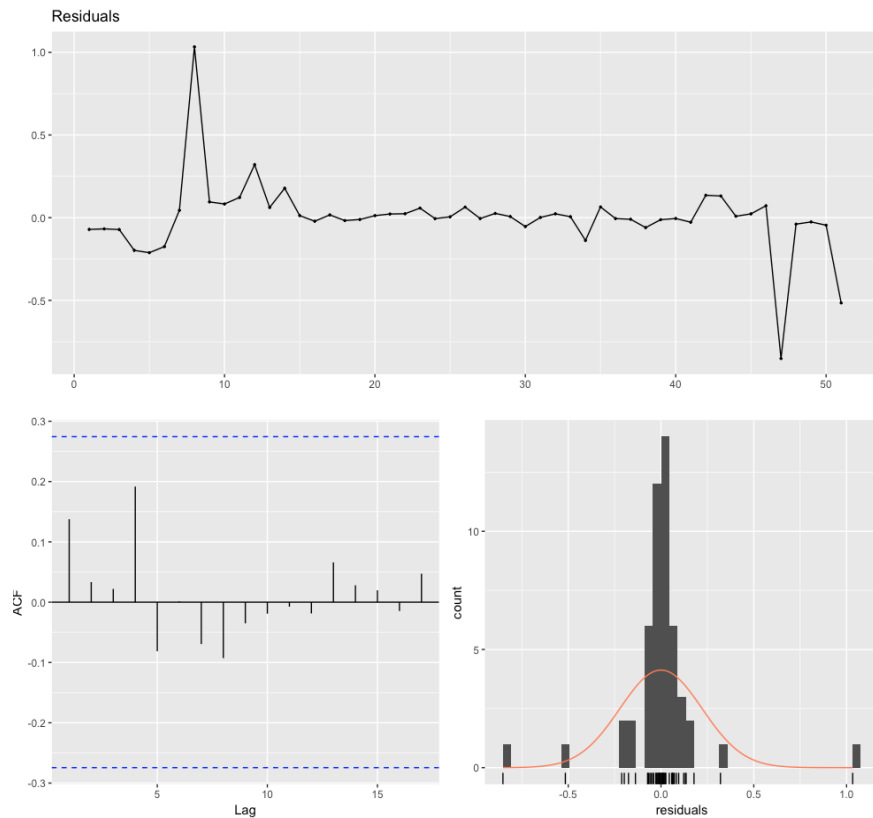
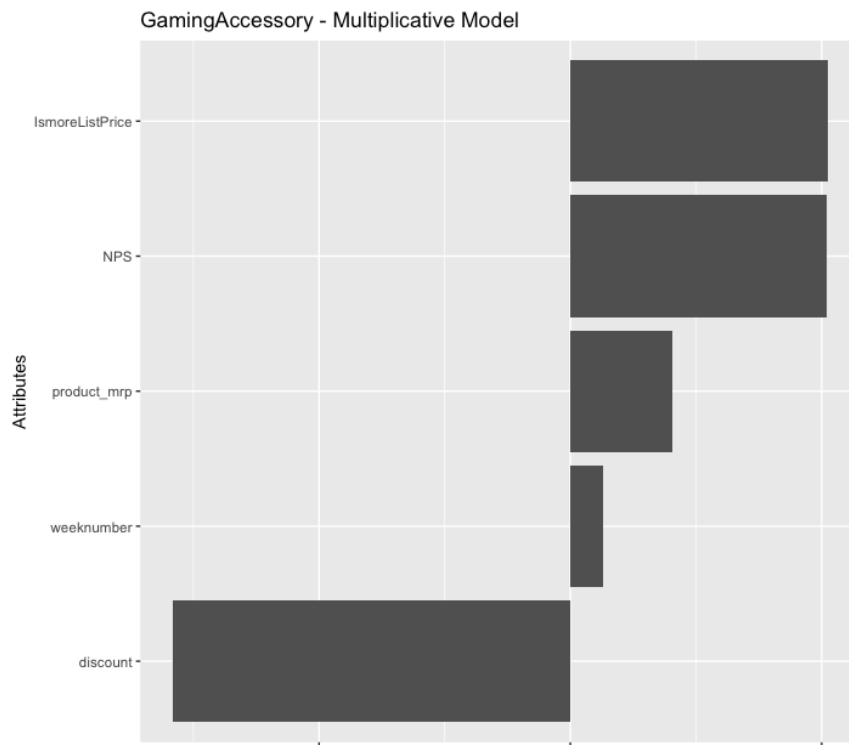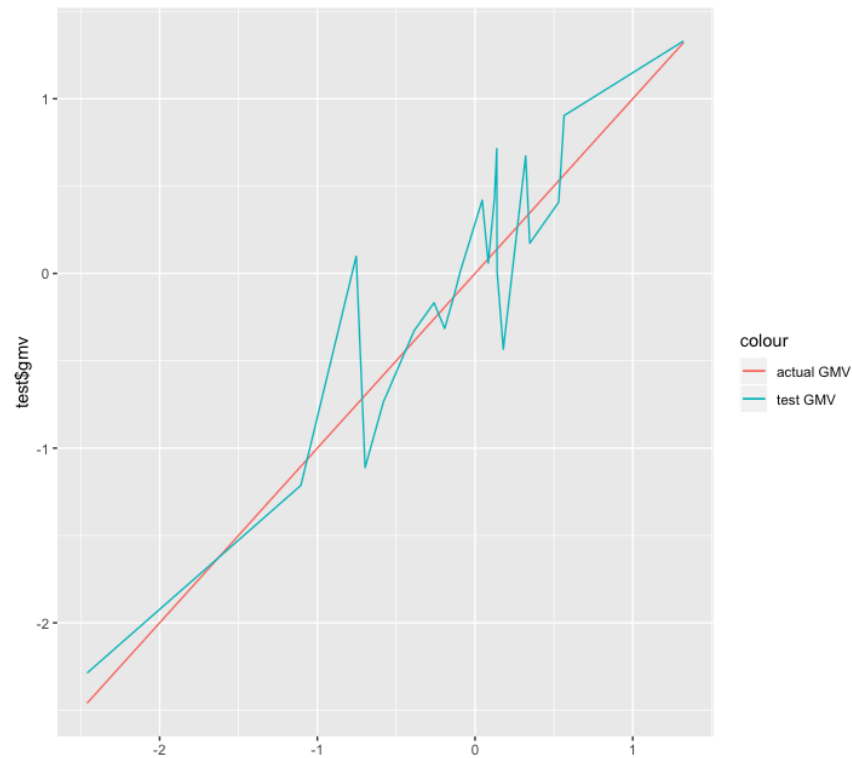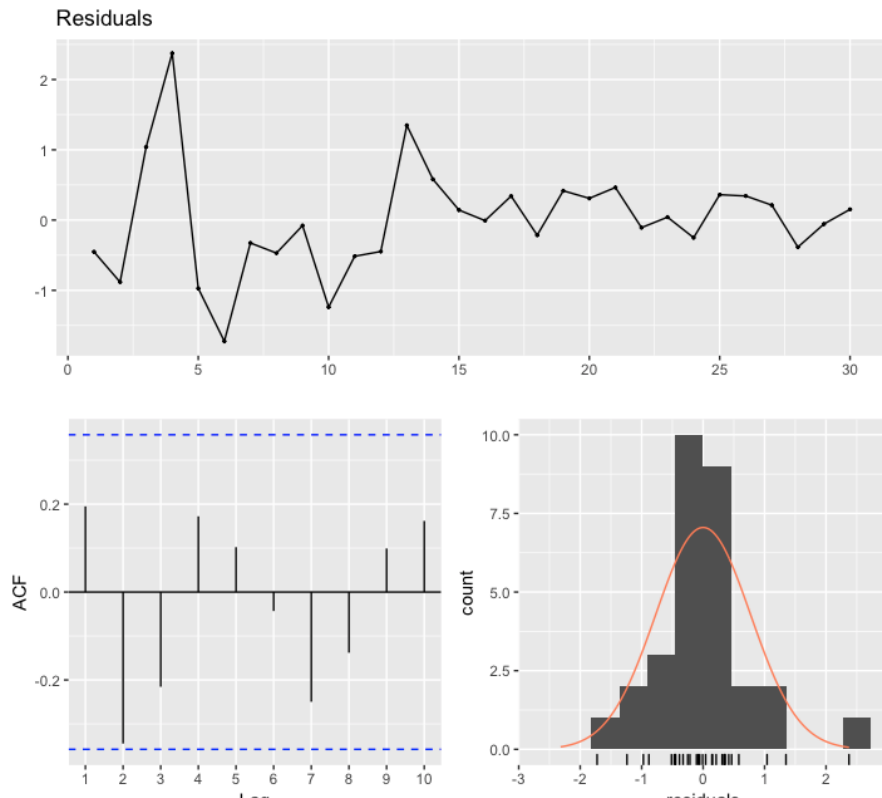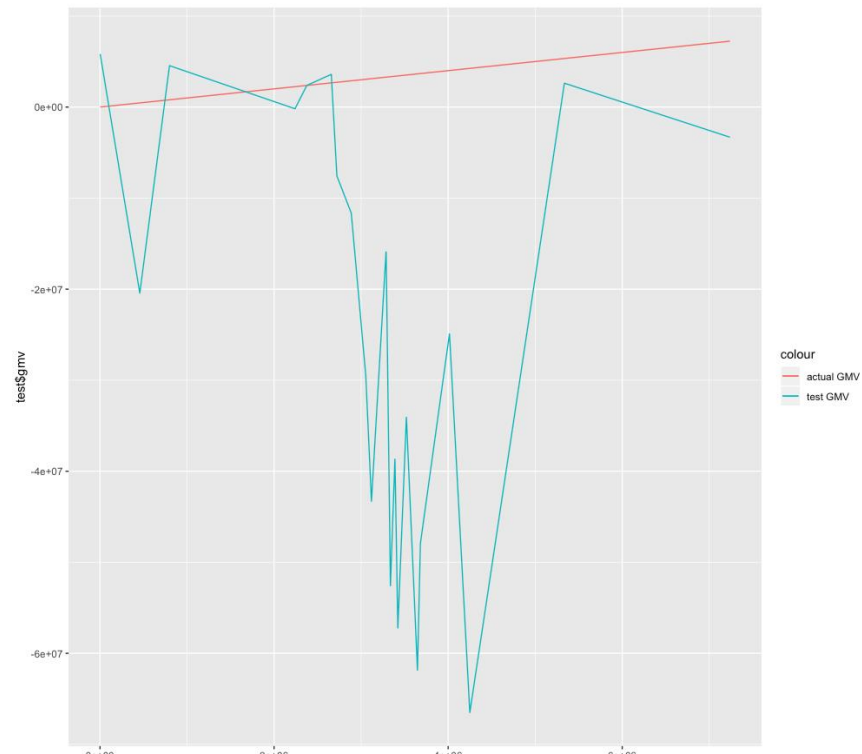# Gaming Accessories – Five models outcome

Model	Significant variables	Multiple R square	Adjusted R square
Linear Model	weeknumber + product_mrp + units + NPS + GamingHeadset + GamingMouse + GamingMousePad + JoystickGamingWheel + TVOutCableAccessory + TV + Sponsorship + Content_Marketing + Online_marketing	0.9798	0.9644
Multiplicative Linear Model	weeknumber + product_mrp + discount + NPS + IsmoreListPrice	0.9736375	0.94797
Distributed Lag Model	JoystickGamingWheel + TVOutCableAccessory	0.7932	0.7393
Auto Regressive DL Model	payment_GamingAccessory	0.6434	0.6023
Kyock Model	GA_data$gmv	0.7982	0.7898

Model 1,2,3 are using multiple independent series for the dependent variable GMV
Model 4,5 are giving best values with the GMV compared against for specific features (1:1)
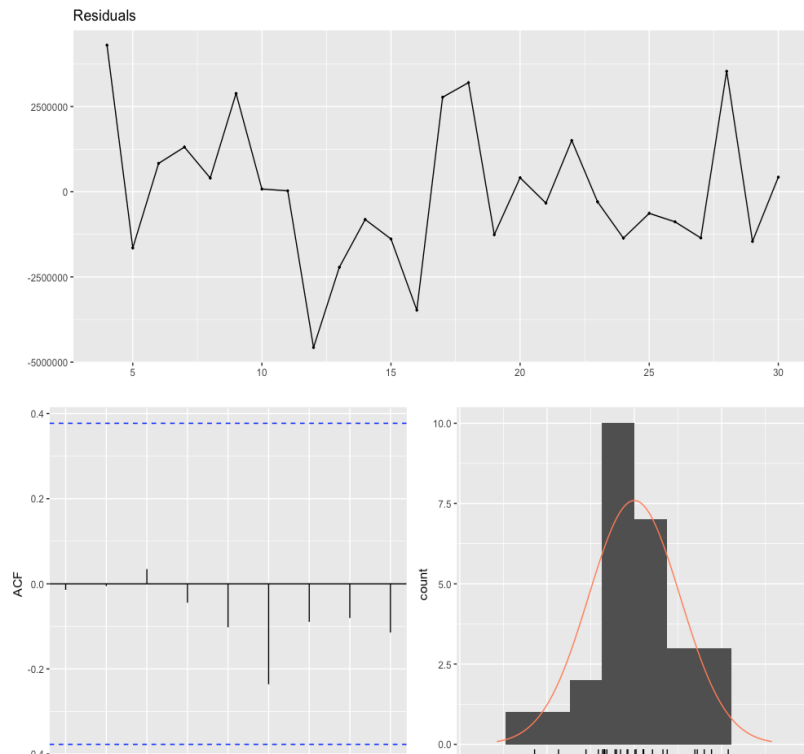
Linear Model: Gaming Accessary
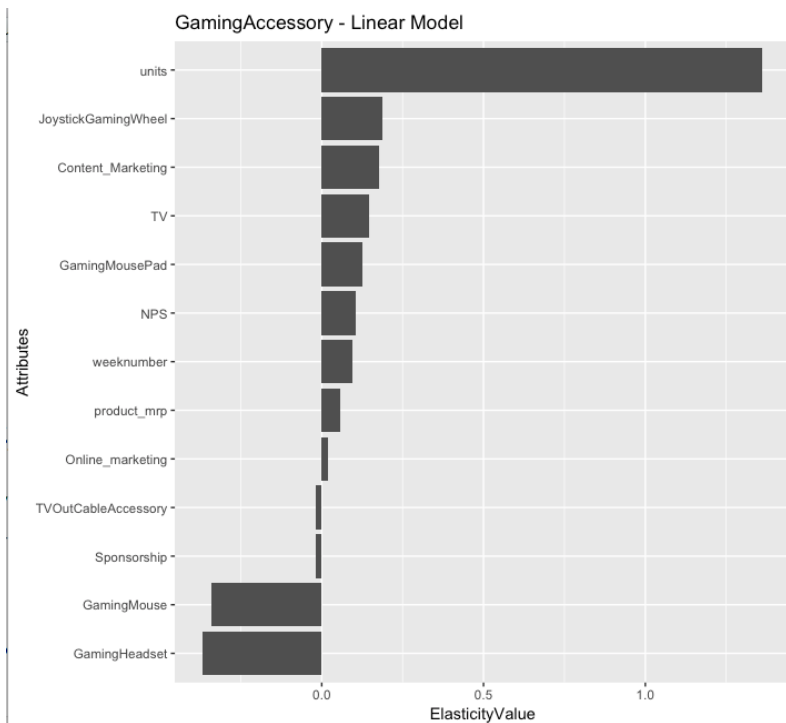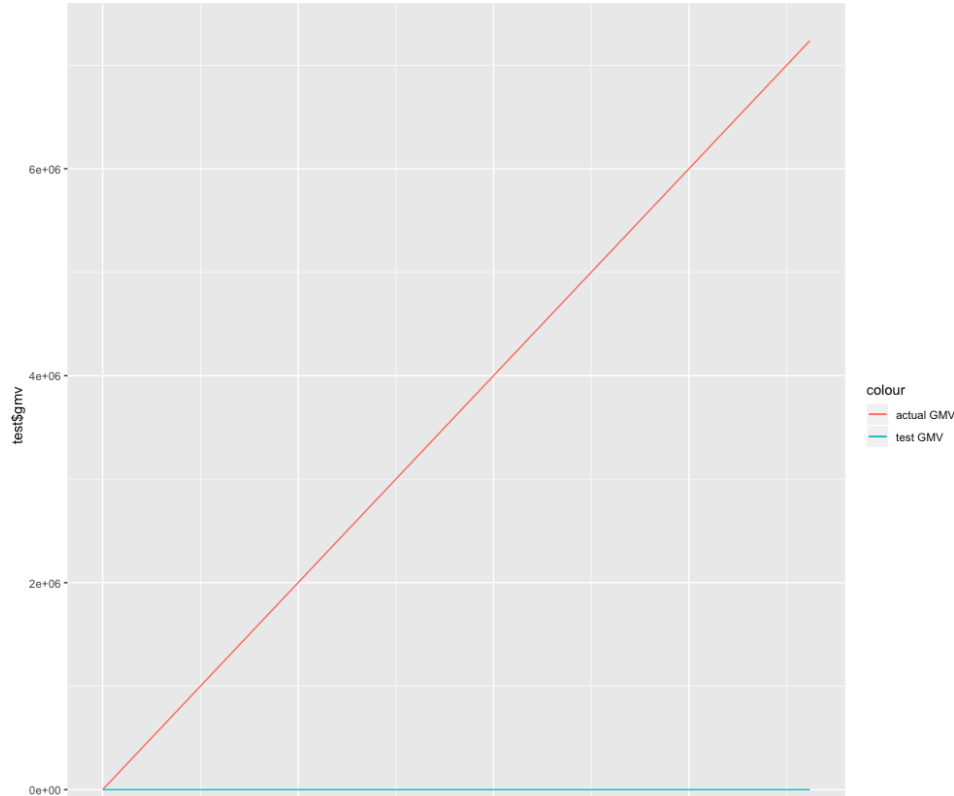
Multiplicative Model : Gaming Accessories
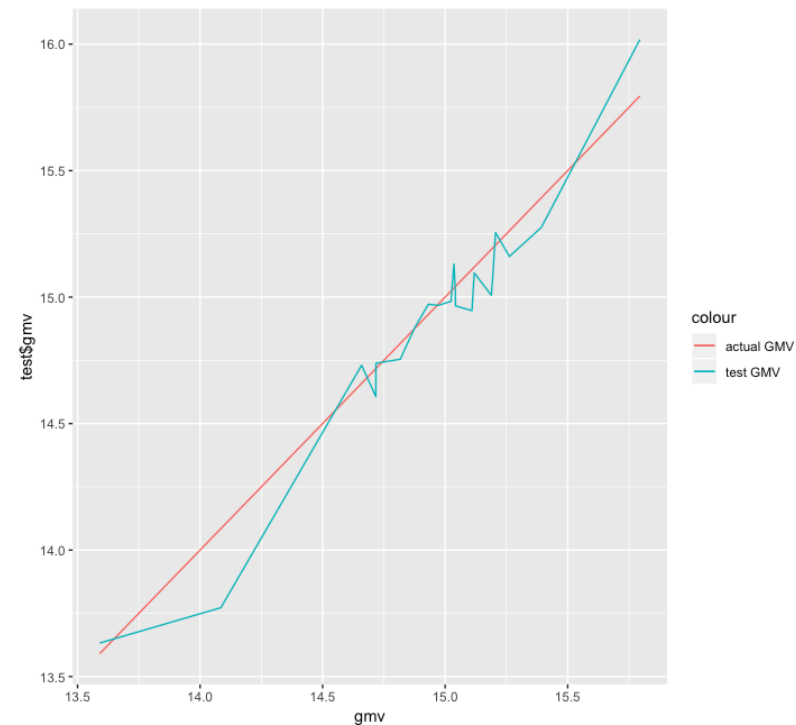
Distributed Lag Model : Gaming Accessories

Auto Regressive DL Model : Gaming Accessories
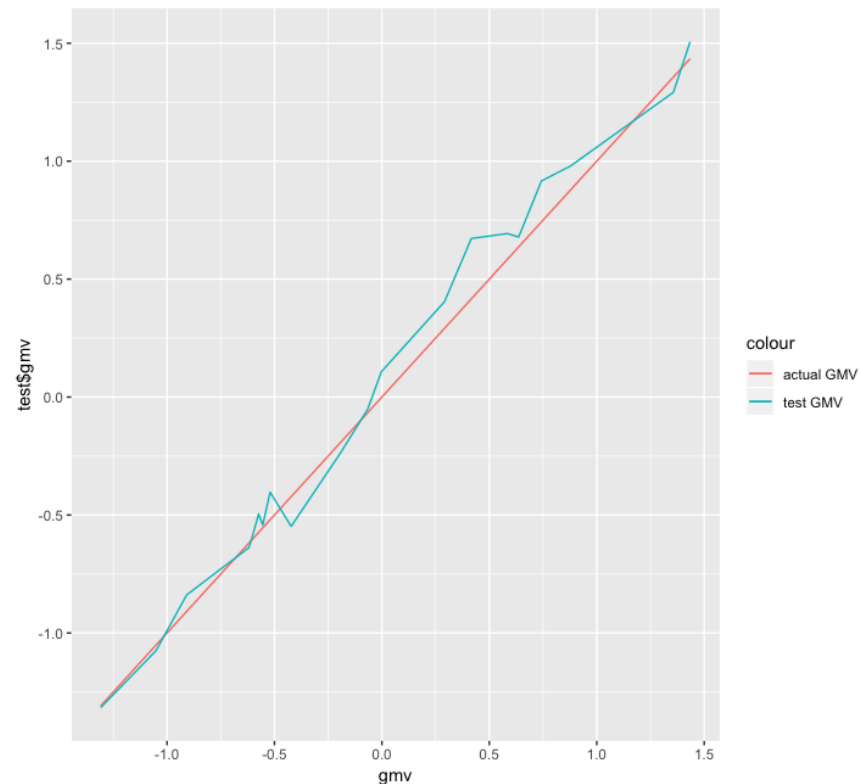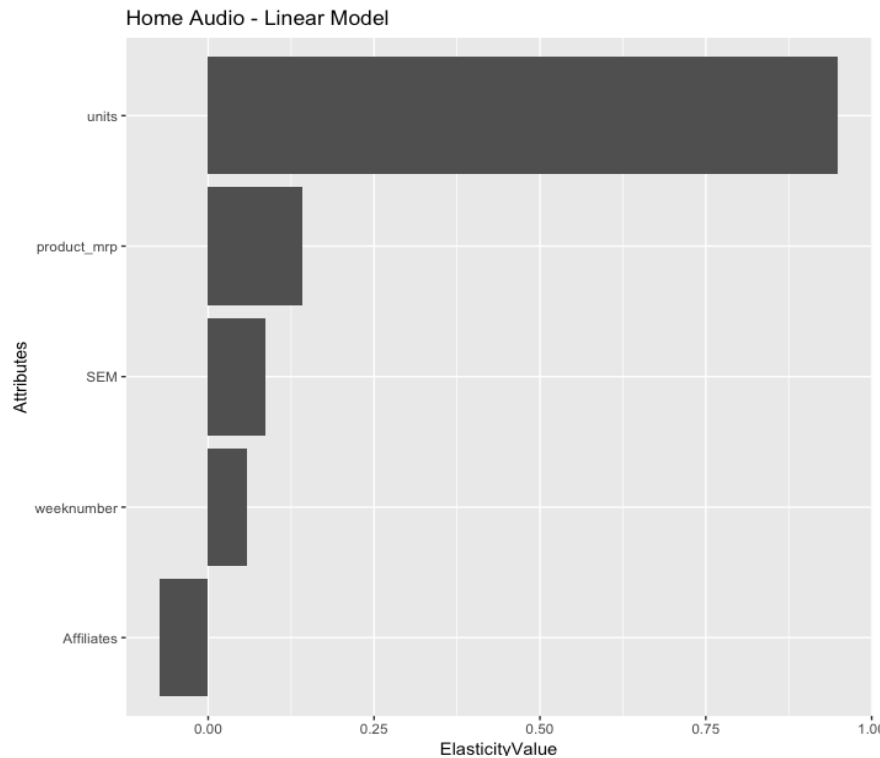
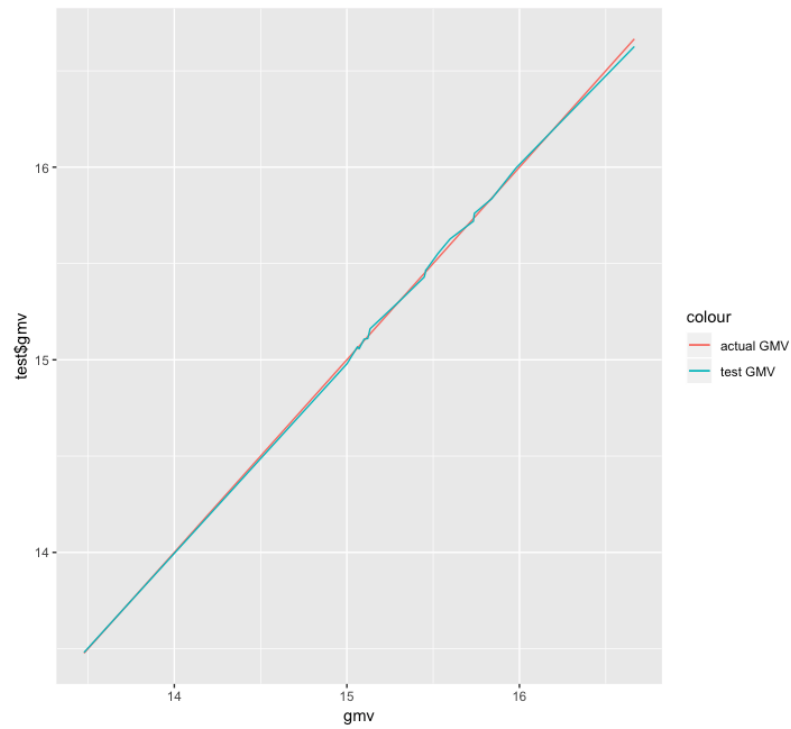# Kyock Model : Gaming Accessories

Home Audio – Five models outcome

Model	Significant variables	Multiple R square	Adjusted R square
Linear Model	weeknumber + product_mrp + units + Affiliates + SEM	0.9941	0.9894
Multiplicative Linear Model	product_mrp + units + discount + Dock + VoiceRecorder	0.9989	0.9987
Distributed Lag Models	weeknumber + product_mrp + units + Affiliates + SEM		
Auto Regressive DL Model	units_homeAudio	0.9809	0.9766
Kyock Model	Home_Audio_Data$units	0.638	0.6223

Model 1,2,3 are using multiple independent series for the dependent variable GMV
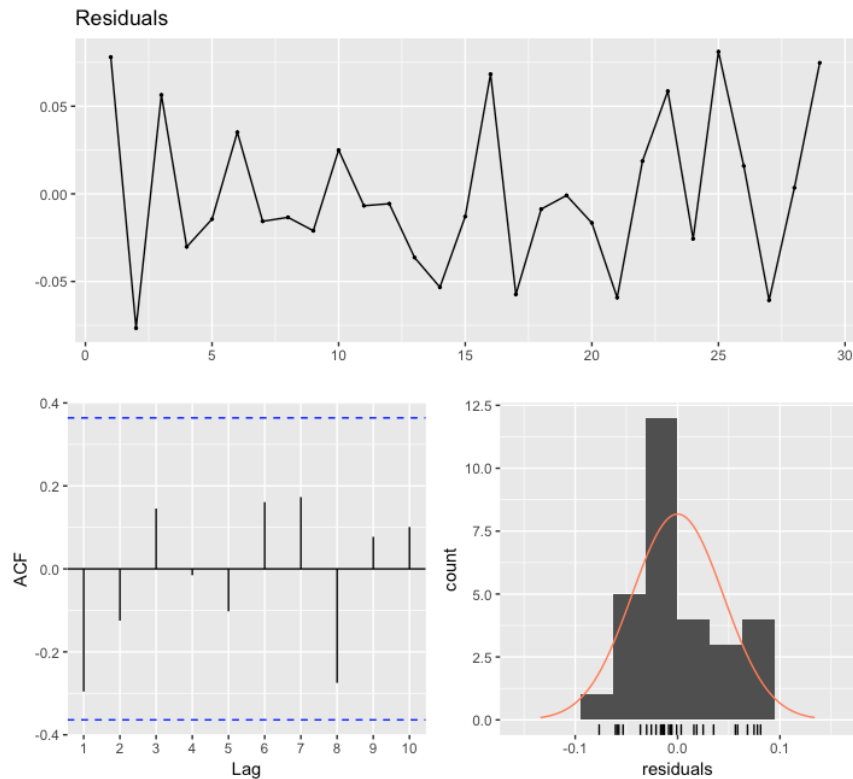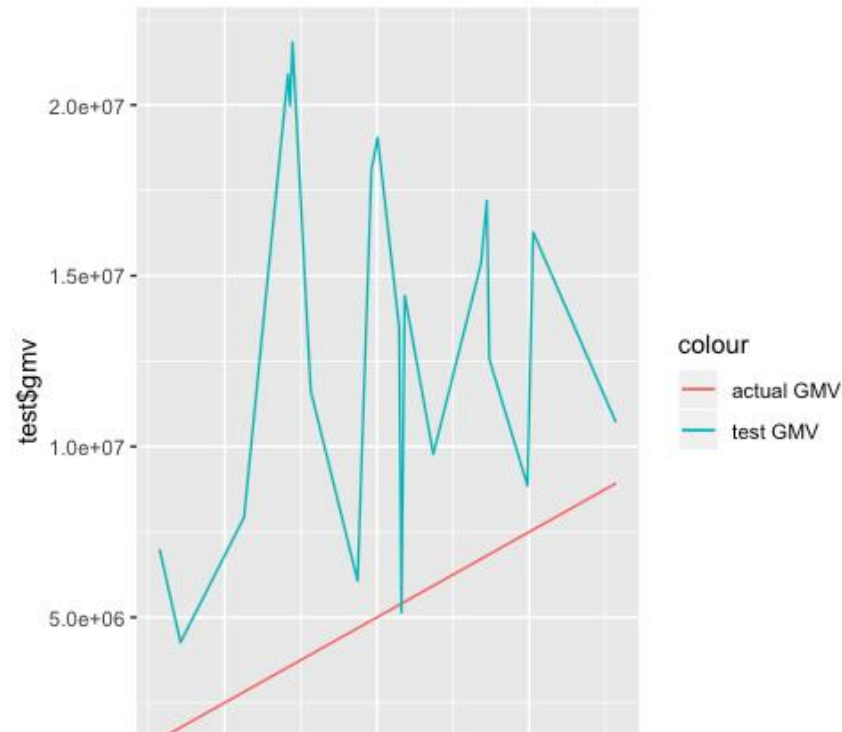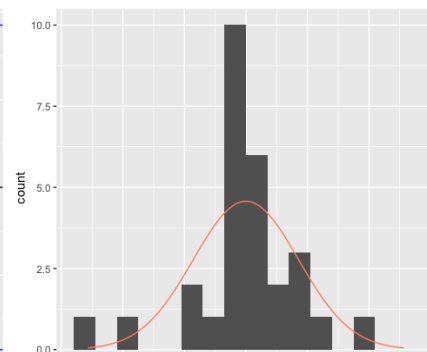Model 4,5 are giving best values with the GMV compared against for specific features (1:1)
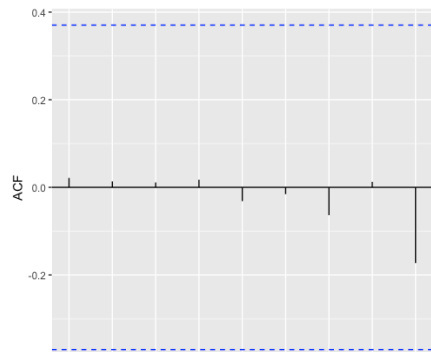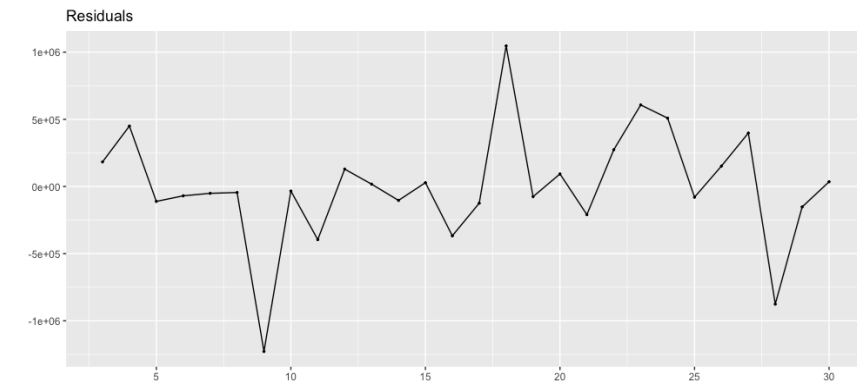
Linear Model : HOME AUDIO
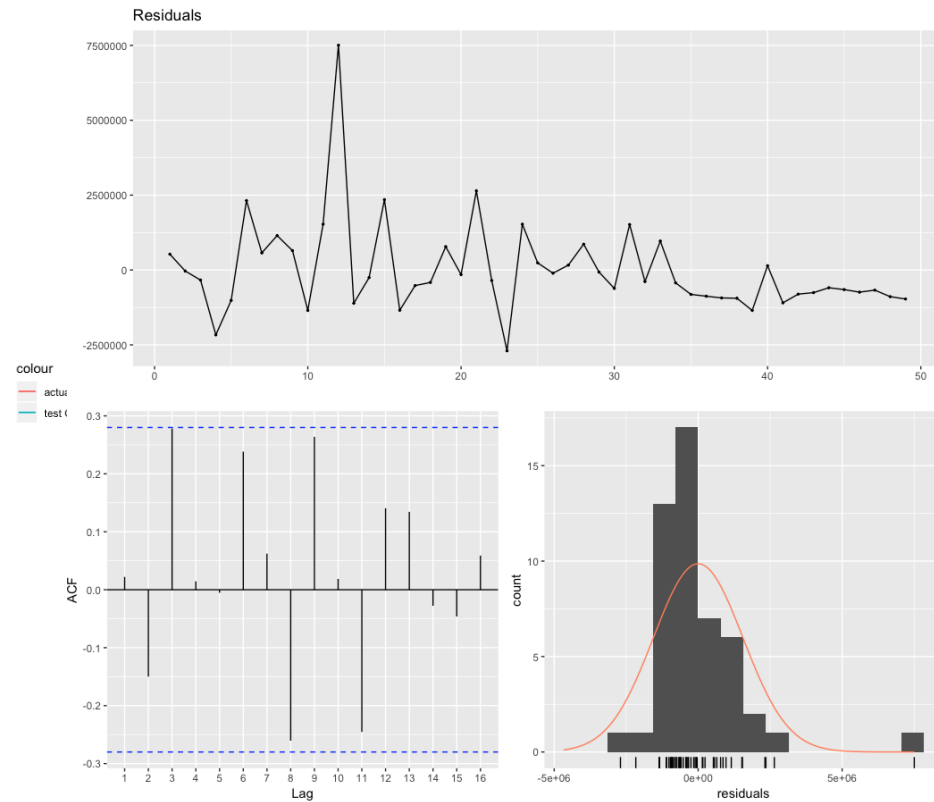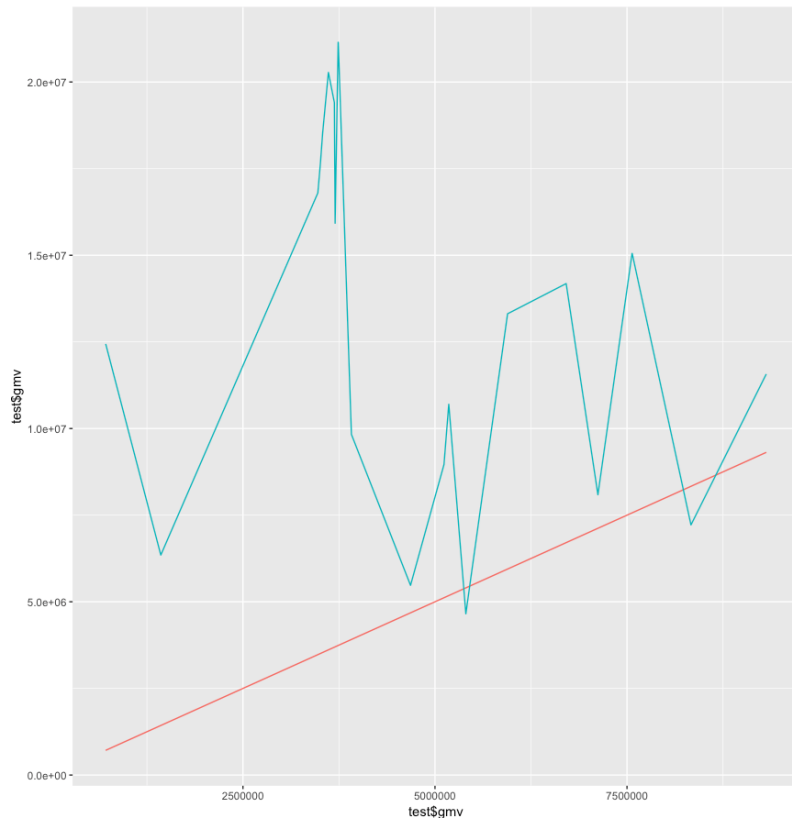
Multiplicative Model : HOME AUDIO
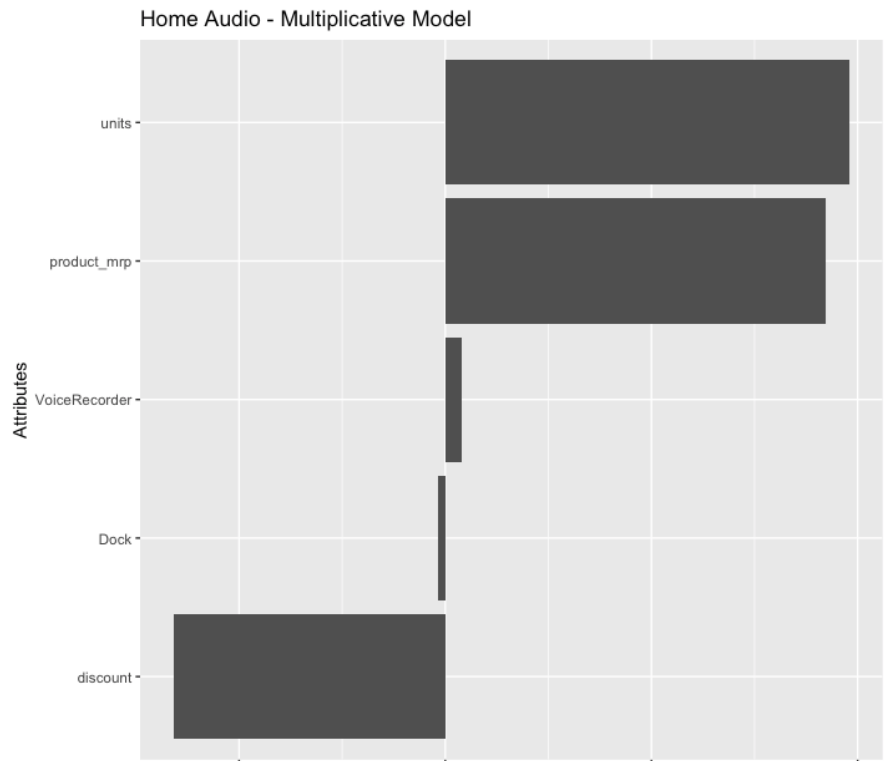
Distributed Lag Model : HOME AUDIO

Auto Regressive DL Model : HOME AUDIO

Kyock Model : Home Audio

Recommendations to CFO / CMO – Camera Accessories

Based on Linear model we conclude TV, Digital and Sponsorship has high impact.

Based on Distributed lag model we conclude that Flash sale is very much effective on gross revenue.

Based on Koyck Model we conclude that Affiliates has good impact on revenue.

Recommendations to CFO / CMO – Game Accessories

Based on Linear model we conclude that TV, Content_marketing , Online_Marketing and Sponsership has high impact on gross revenue.

Based on Multiplicative Model we conclude that NPS is effective.

Based on Koyck Model we conclude that Affiliates has good impact on revenue.

Recommendations to CFO / CMO – Home Audio

Based on linear model we conclude that Affiliates and SEM has very high impact on Home Audio gross revenue.

Distributed lag model tells that Affiliates and SEM is very much effective on gross revenue.

We recommend to the Management to invest on the above identified Marketing Channel for the 3 Categories.