

Ecommerce Capstone Project

SUBMISSION

1. Mruthyunjaya Kumar
2. Dharmanandana Reddy
3. MuthuKumaran SM
4. Deepak Kumar

Ecommerce Capstone Project - Problem Statement

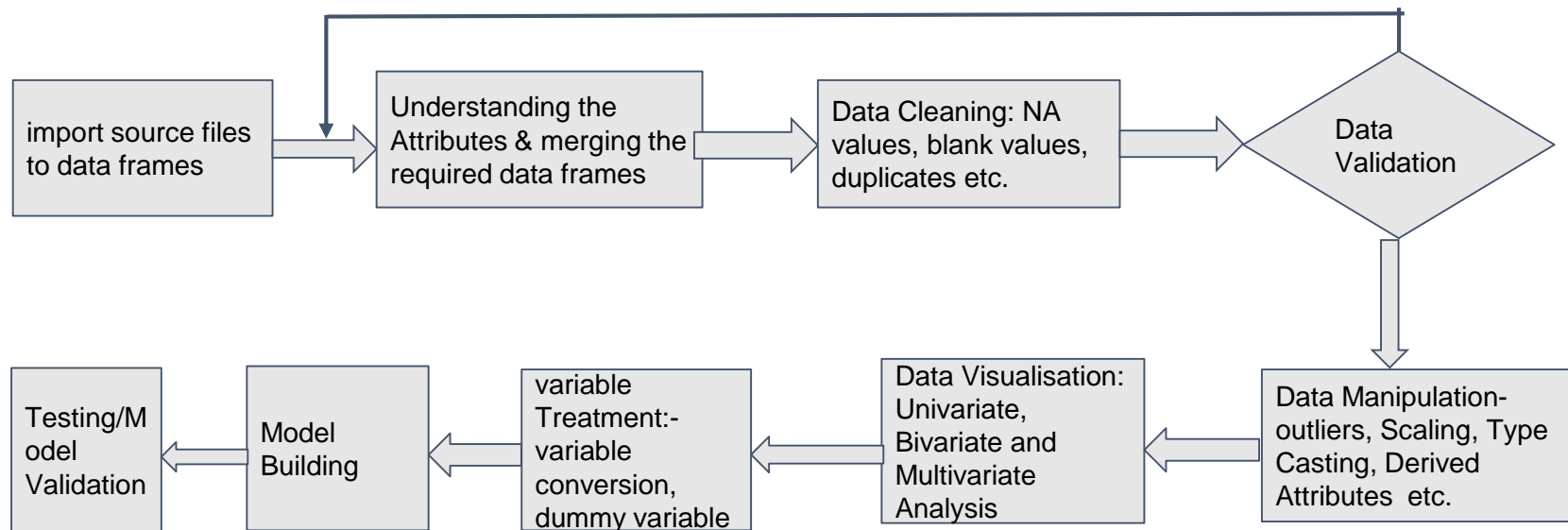
Eleckart is an e-commerce firm. Last year they spent money in marketing and they offered promotions as well but the CFO feels that marketing and promotions was not sufficiently impactful.

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



Data Understanding

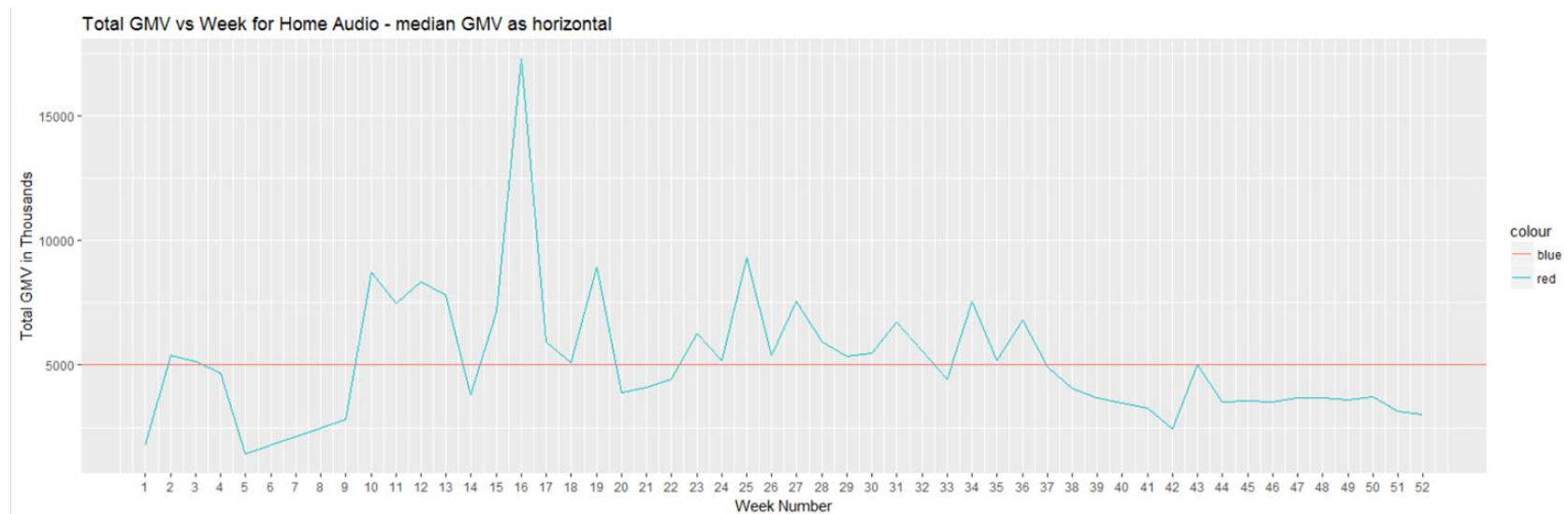
- We have got below datasets -
 - Consume Electronics - This file has order details of each product from June 2015 to July 2016.
 - Product List - This file contains details of product and frequency it was sold.
 - Media Investment - It provides details of money spent on different marketing channels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

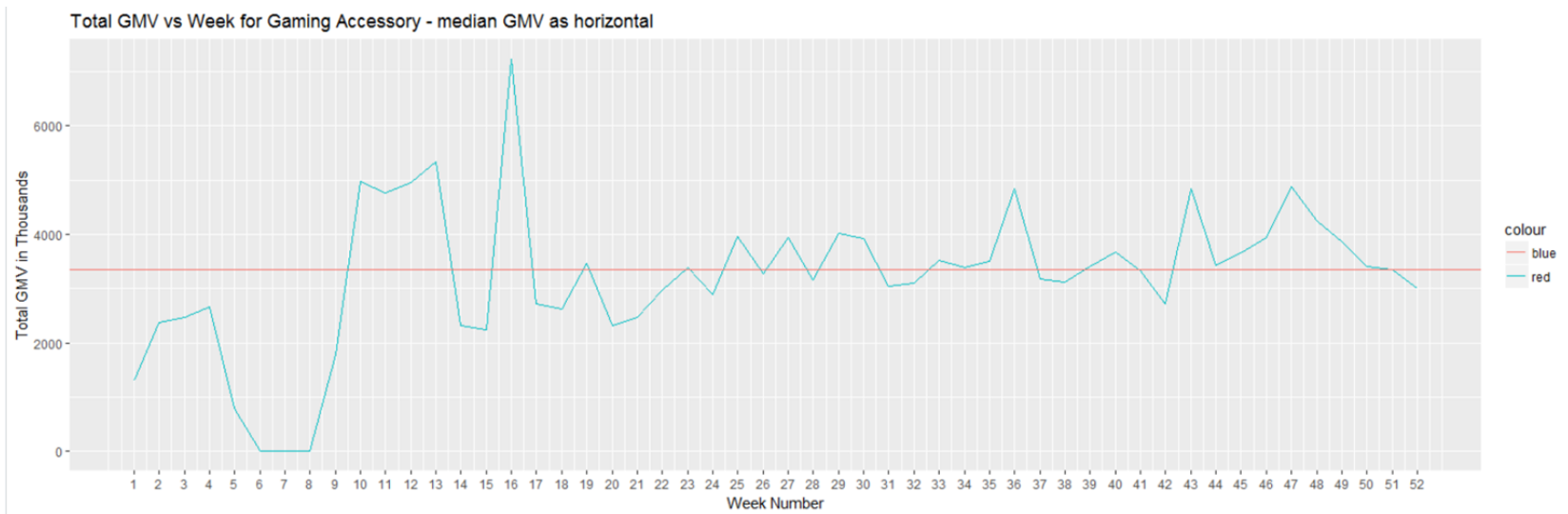
Understanding the trend of sale (Home Audio)

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.



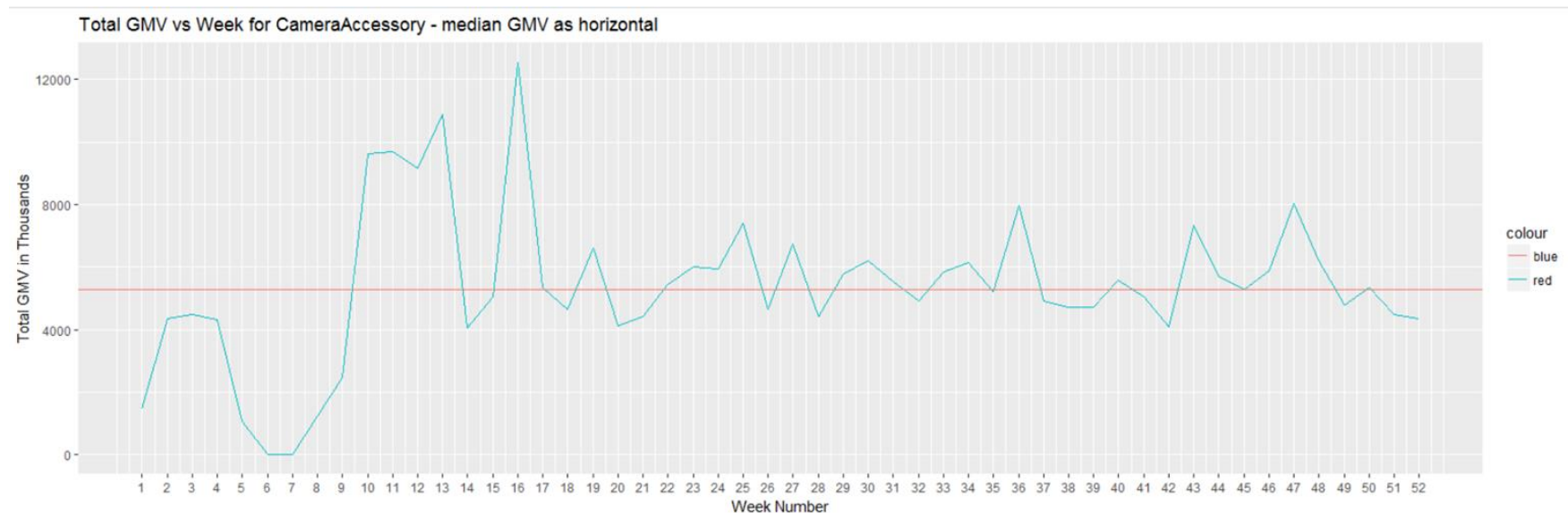
Understanding the trend of sale (Gaming Accessories)

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.



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.

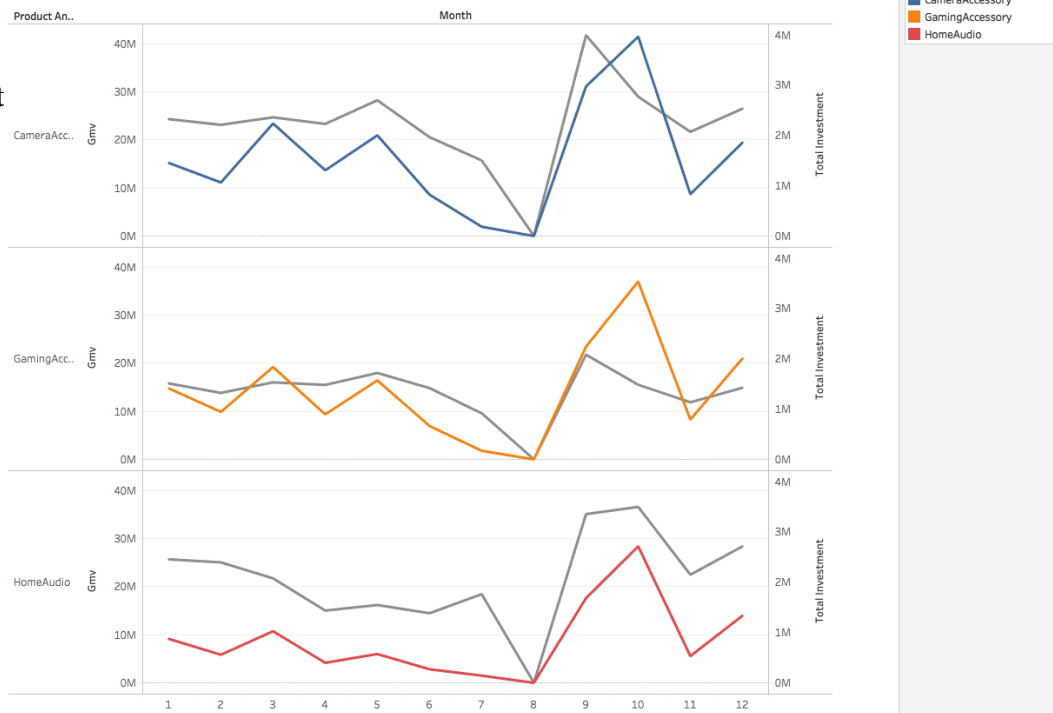


****Month 1-7 is considered as Jan 2016 - Jun 2016**

****Month 8-12 is considered as July 2015 - Dec 2015**

- For Gaming Accessories total marketing investment Is more effective than the Camera Accessories And Home Audio between the month 1 to 6.
- During month 9 to 11 total investment in home Audio has more impact

GMV_vs_Investment_Monthly

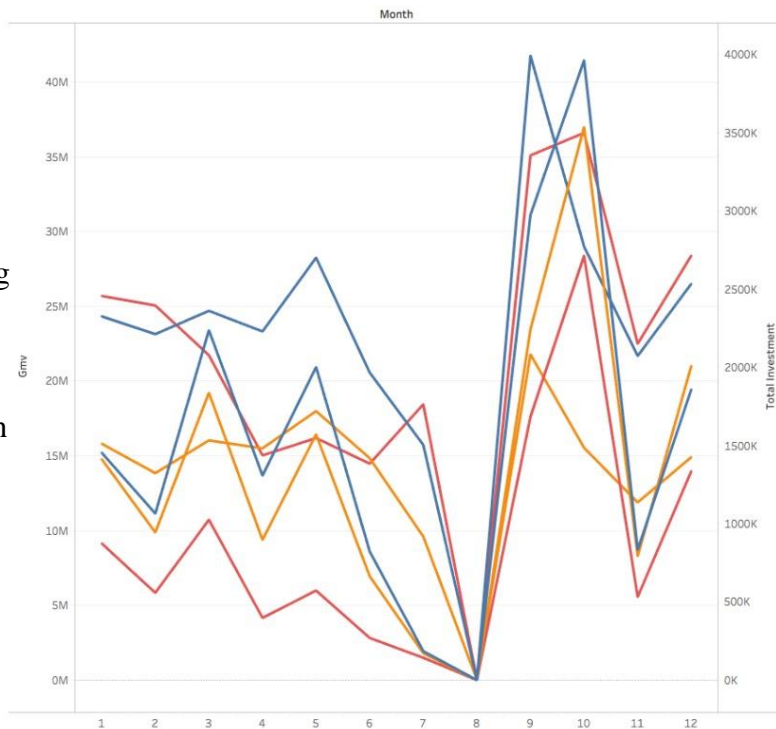


Month 1-7 is considered as Jan 2016 - Jun 2016

Month 8-12 is considered as July 2015 - Dec 2015

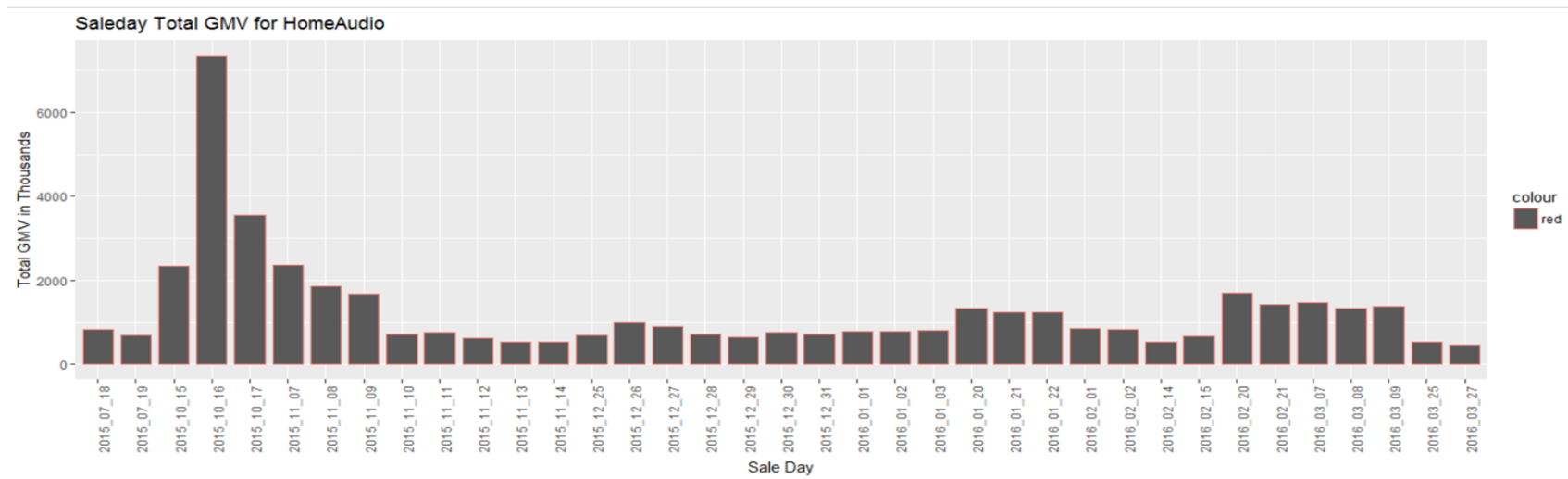
- Home Audio does not have Much impact from the total Marketing investment. We can see that the sales is Decreasing from month 2 to 3 But the investment is increasing In this time period.
- For Camera Accessory also We can see same trend between Month 2 to 4. Total marketing Expense has increased but sale Is constant. There is no impact On sale.

GMV_vs_Investment_Monthly



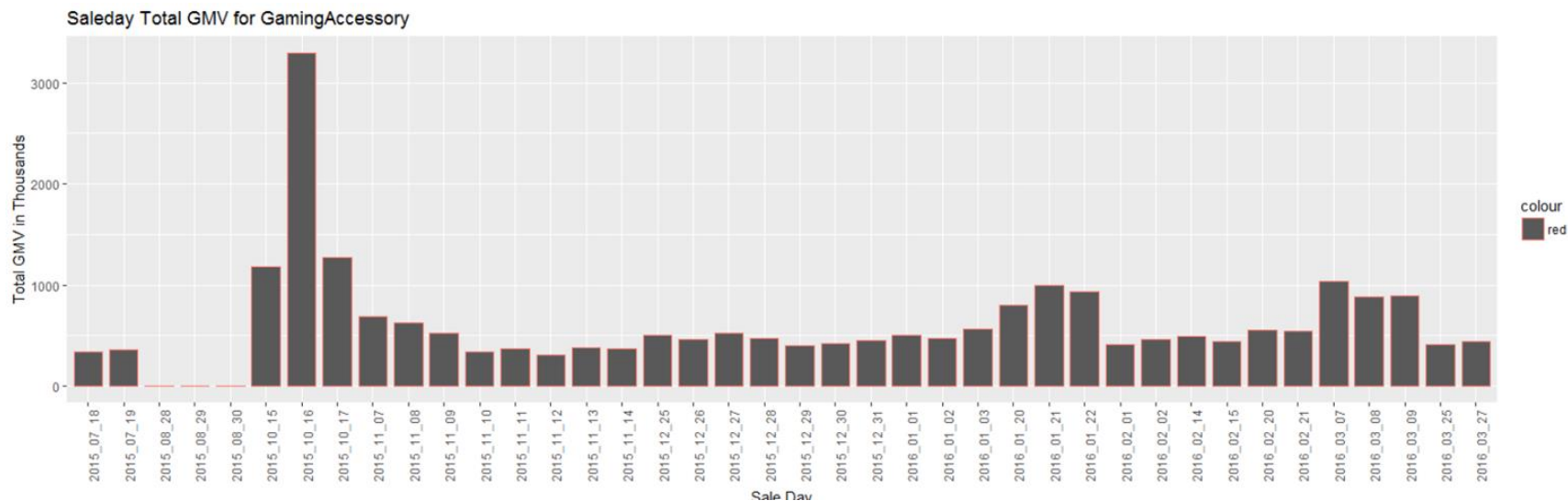
For Home Audio, We can say that whenever promotion is given during festival (Dussehra, Christmas etc.) then the revenue is very high but if the promotions are given frequently on other days then it is not much impactful. When promotion was given on 15/10/2015 then revenue is very high. Similarly on 20/1/2016, 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.



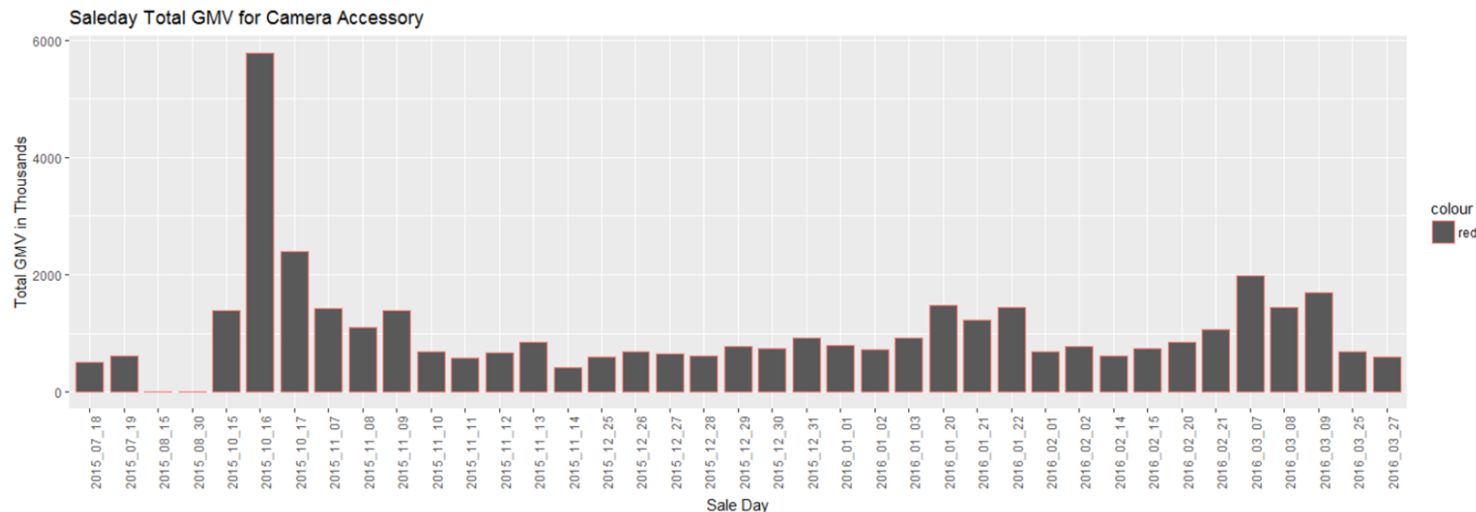
For Gaming Accessory, We can say that the revenue is very high when the promotion is given during Dussehra, republic Day, BSD but on other days it is not much impactful.

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

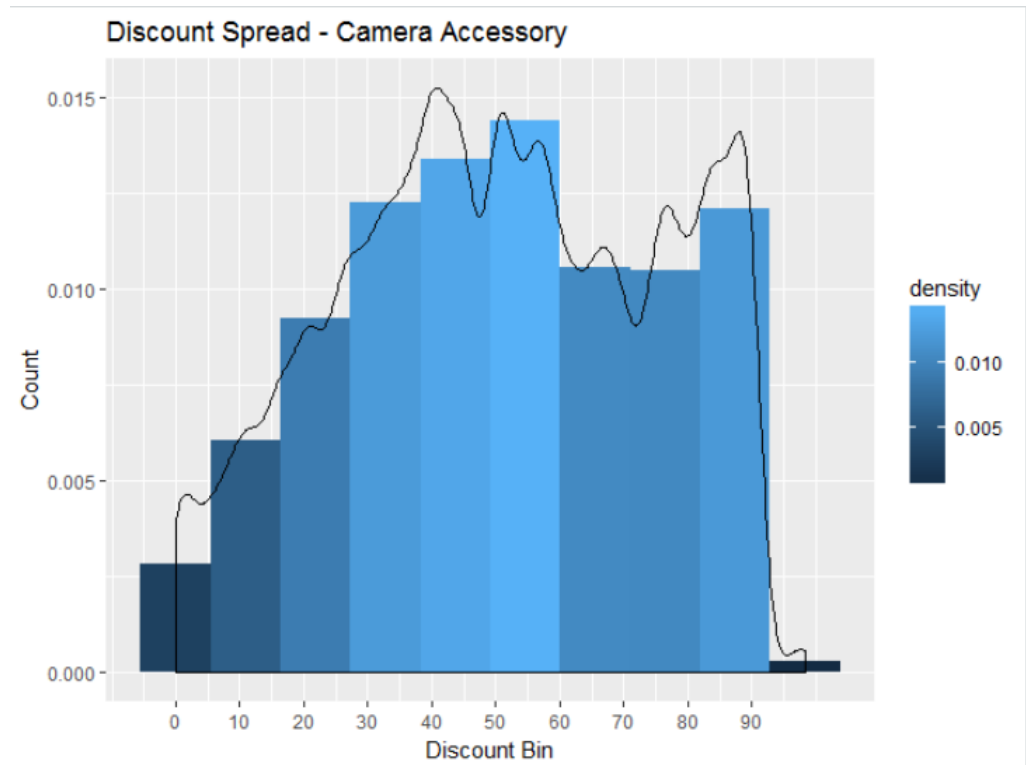


For Camera Accessory, We can see that the revenue is very high when the promotion is given during Dussehra, republic Day, BSD but on other days it is not much impactful.

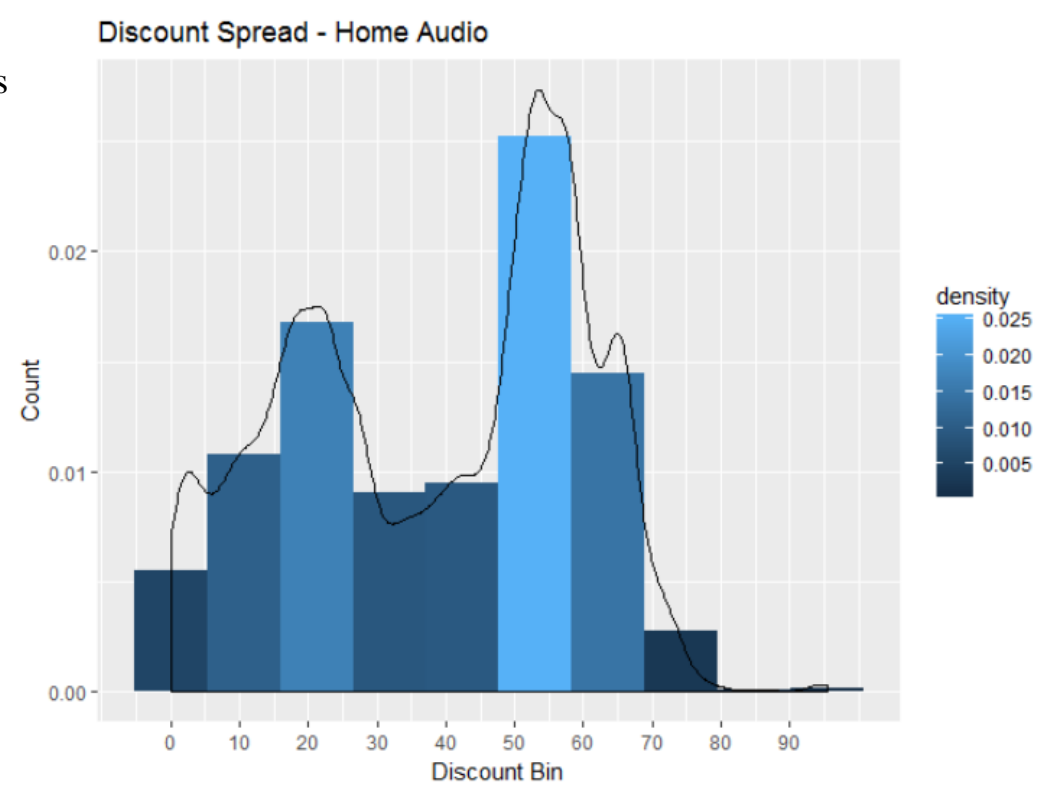
So the promotions given during festival will have very high impact on the revenue.



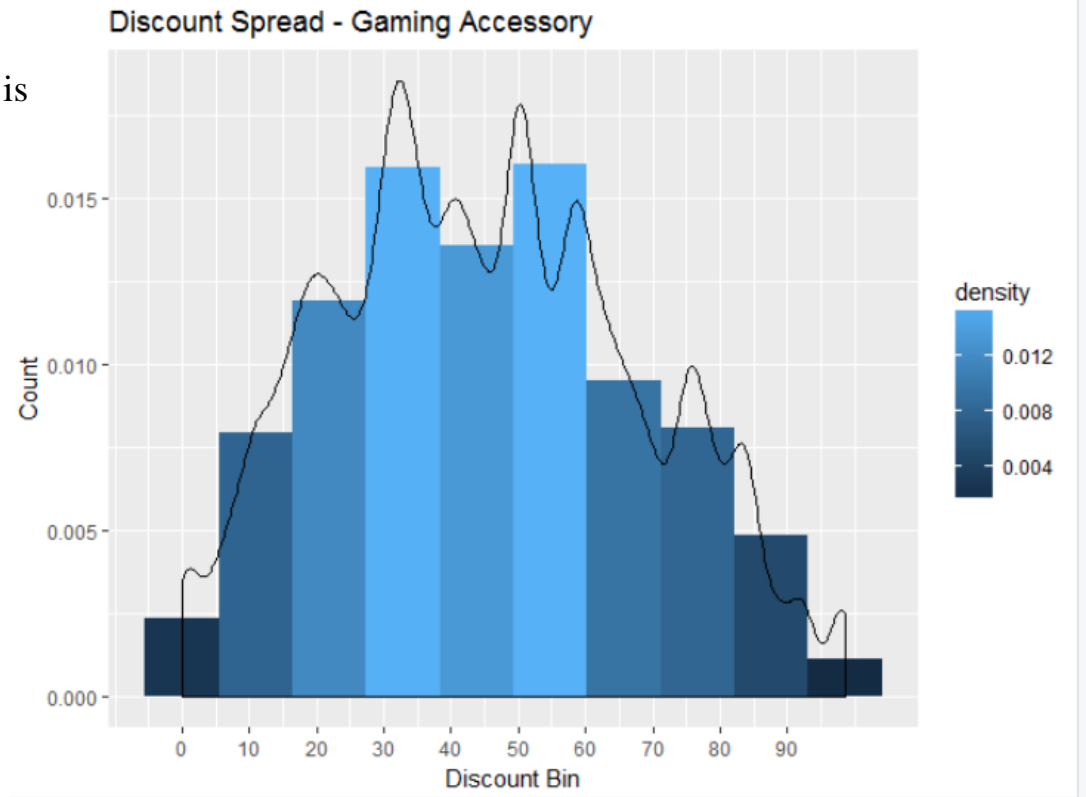
Looking at the plot we can say that
When discount increases the count of sale is
increasing for the “Camera Accessory”.



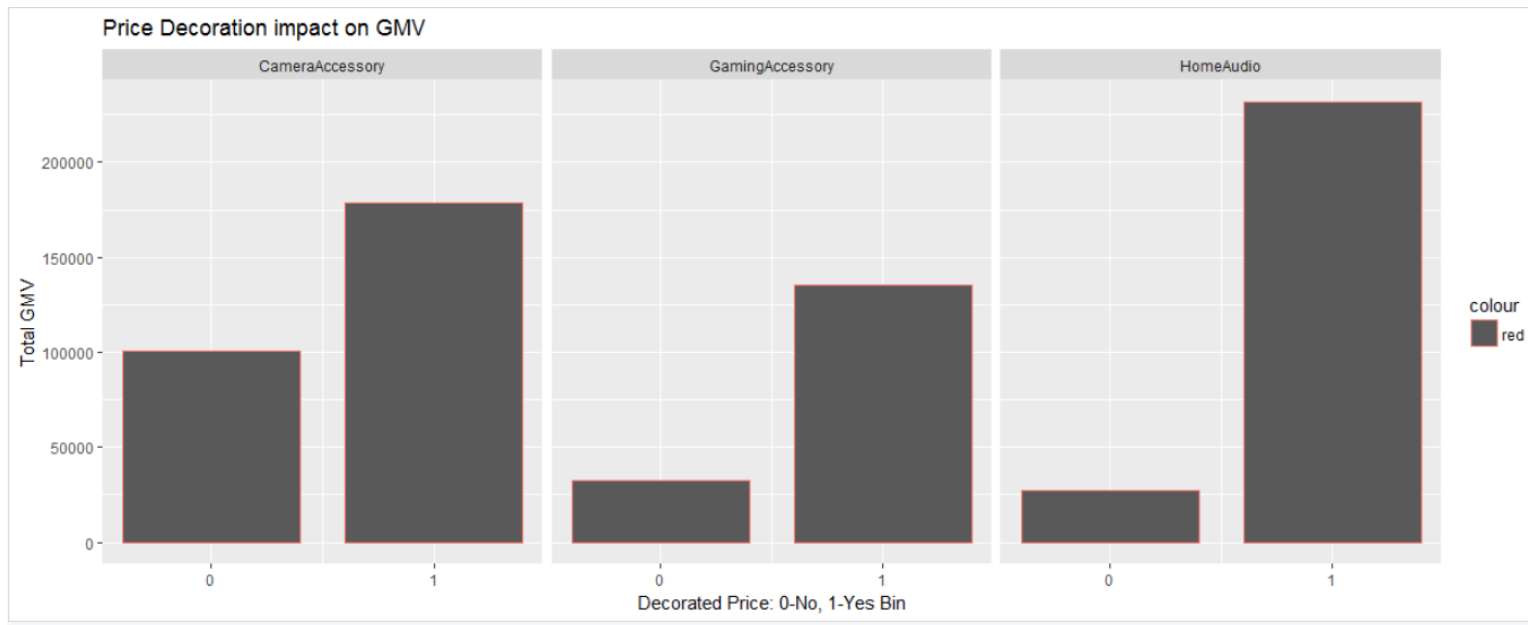
Looking at the plot we can say that
When discount increases the count of sale is
increasing for the “Home Audio”.



Looking at the plot we can say that
When discount increases the count of sale is increasing for the “Gaming Accessory”



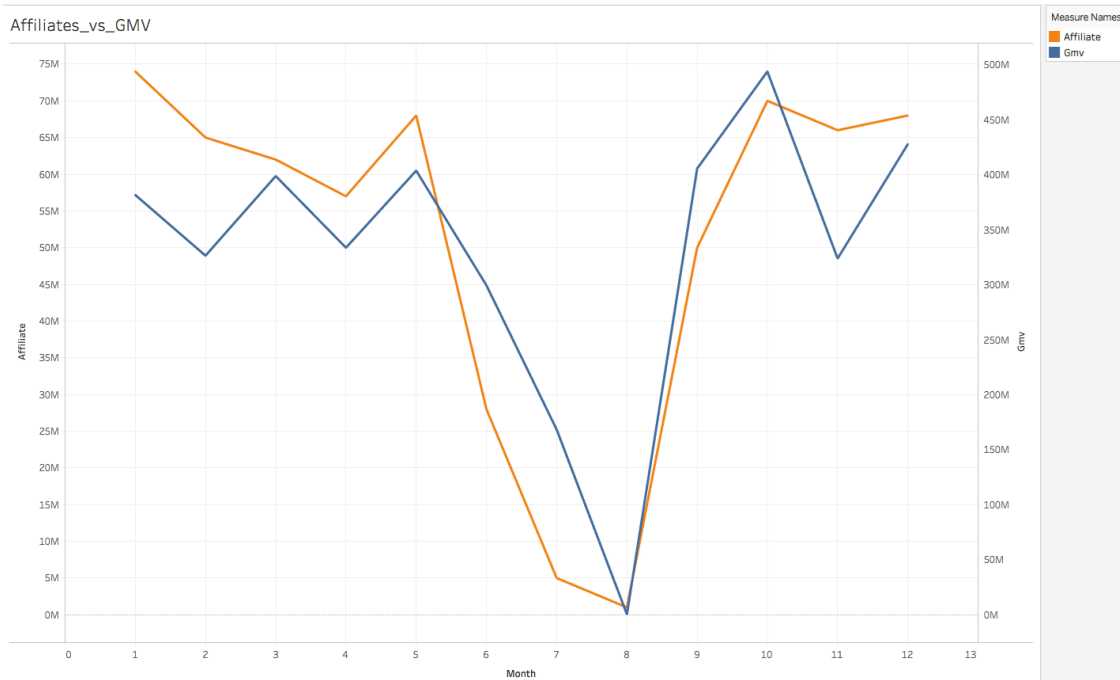
We can see that with the decorated price the revenue is higher for all the 3 categories. So price decoration has high impact on the revenue generated.



***Month 1-7 is considered as Jan 2016 - Jun 2016*

***Month 8-12 is considered as July 2015 - Dec 2015*

- Affiliates spend and revenue Has similar trend
- Affiliate has direct impact on the Sale
- We can move some expense from Month 1,2,3 to 6,7,8 because revenue Is decreasing with decrease in affiliate

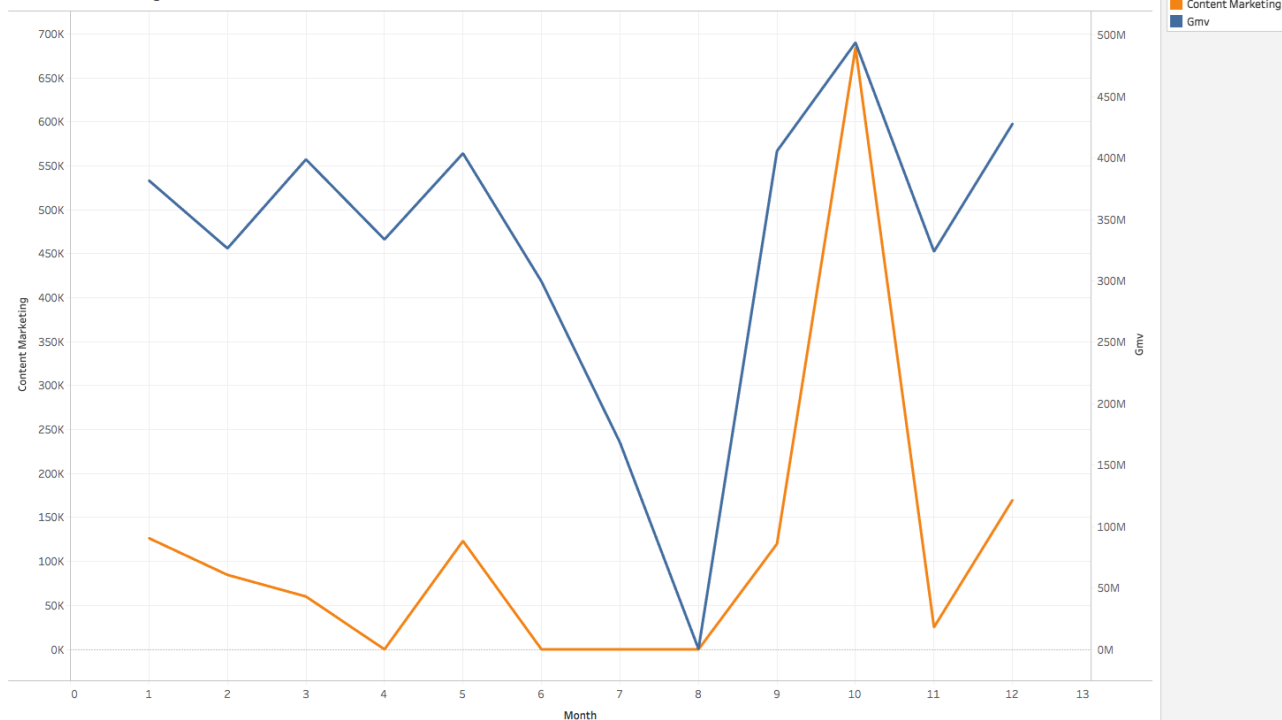


***Month 1-7 is considered as Jan 2016 - Jun 2016*

***Month 8-12 is considered as July 2015 - Dec 2015*

- Aug, sept, Nov (8,9,10) are the months Which has highest sale
- With the increase in content Marketing spend the revenue Has also increased.
- When the content marketing Value decreased or is 0 then Sale has declined mostly
- I think we should put more fund For the content Marketing

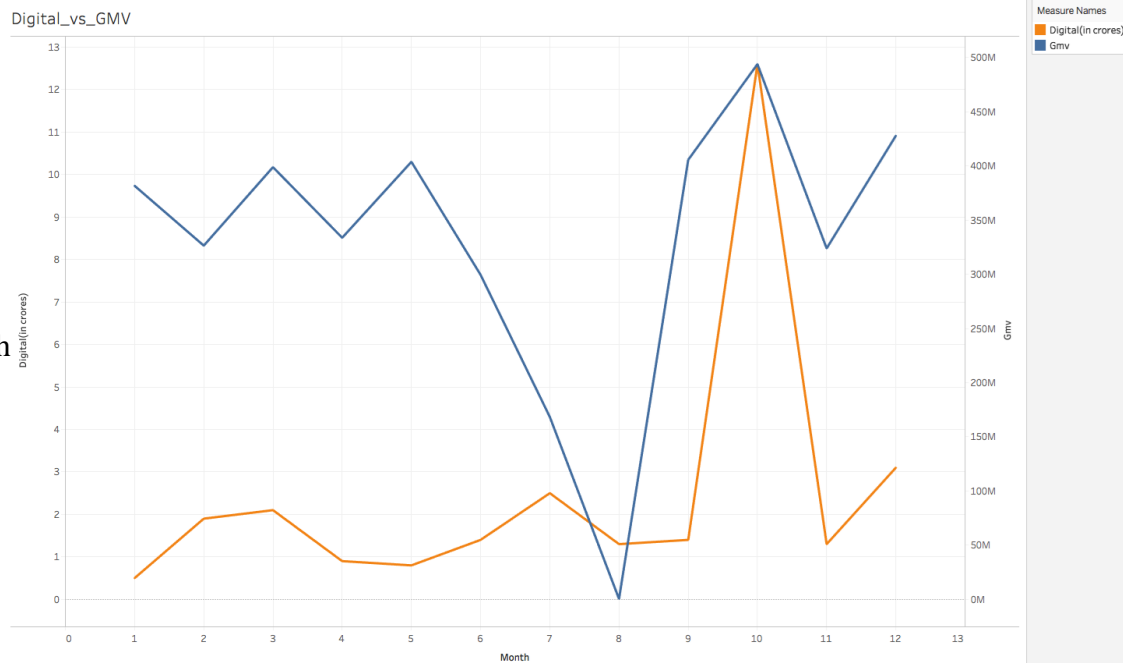
ContentMarketing_vs_GMV



***Month 1-7 is considered as Jan 2016 - Jun 2016*

***Month 8-12 is considered as July 2015 - Dec 2015*

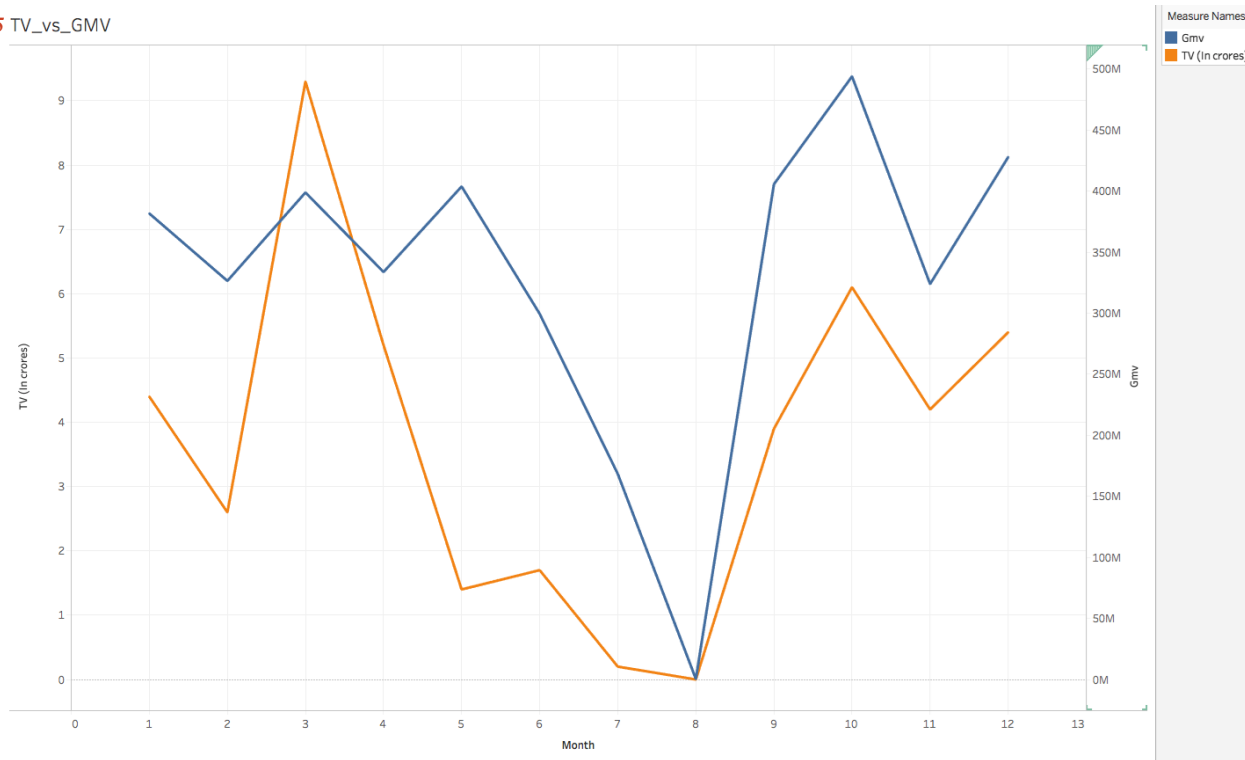
- We can see that during 9th to 11th month its impact is good. With increase in digital expense revenue is increased. But these are festival Months
- Looking at the months from 4 to 7 We can say that digital expense is not Useful here. Revenue is constantly Decreasing here.
- We can stop spending during these month



***Month 1-7 is considered as Jan 2016 - Jun 2016*

***Month 8-12 is considered as July 2015 - Dec 2015* TV_vs_GMV

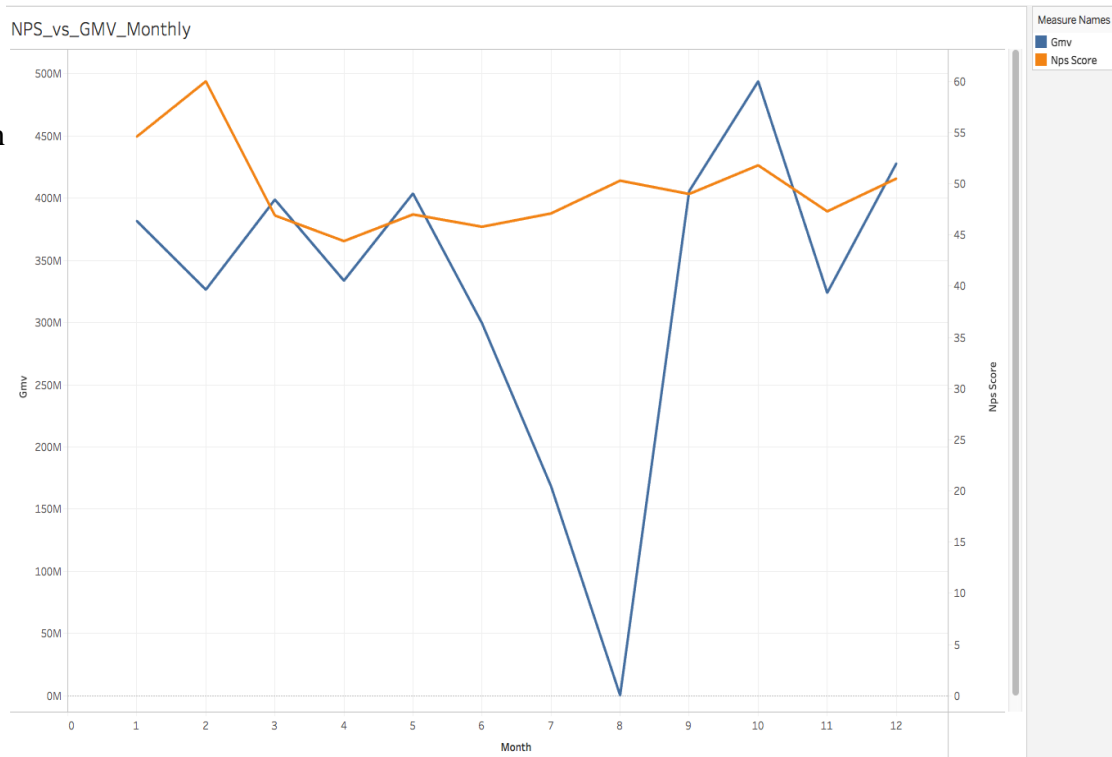
- We can clearly say that TV has direct impact on the Revenue.
- Both the lines TV spend and Revenue generated has similar trend
- During the month 3,4,5,9,10, if We increase the expense on TV Then we may get more Revenue response. Because in This period revenue is Constantly decreasing.
- We can reduce the expense From month 3 because we are Spending too much in this month.



***Month 1-7 is considered as Jan 2016 - Jun 2016*

***Month 8-12 is considered as July 2015 - Dec 2015*

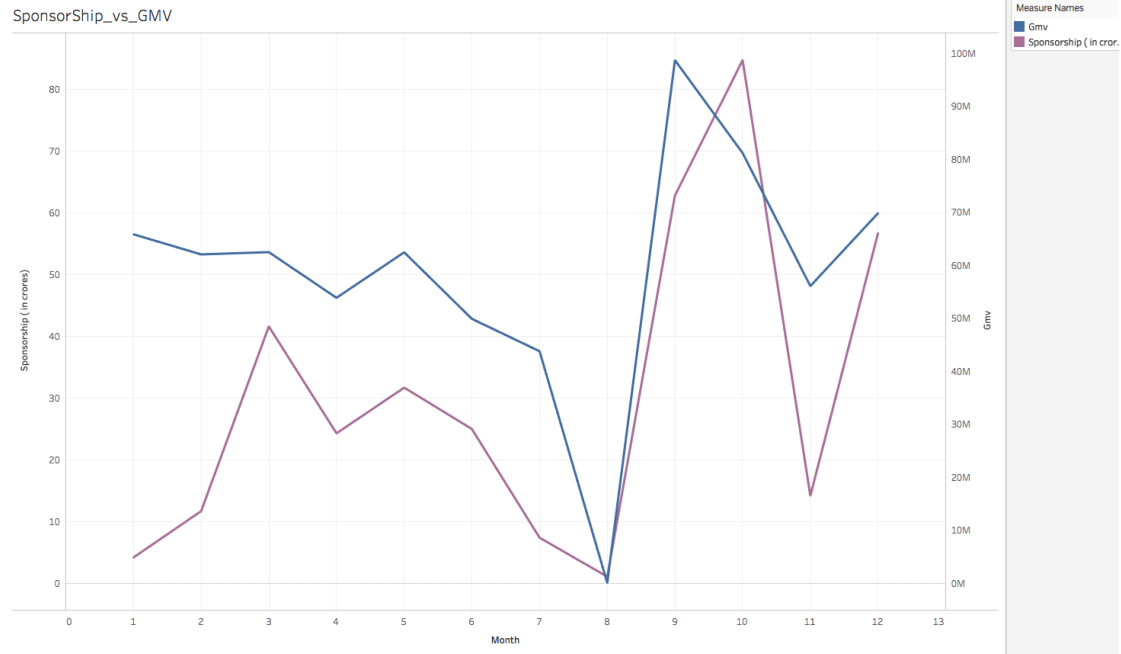
- With slight decrease in NPS score Sales is decreasing rapidly between month 4 and 5. Similarly between 10 and 12. So decrease in NPS score is not good
- But with high increase in NPS score does Not have any impact in sales between Month 1 and 2. Again with high value of NPS score Between month 5 to 9 does not put any Increase in sales.



***Month 1-7 is considered as Jan 2016 - Jun 2016*

***Month 8-12 is considered as July 2015 - Dec 2015*

- Looking at the trend of Sponsorship and GMV lines
We can say that Sponsorship does not have much impact on sales from Month 1 to month 3 and from Month 9-10
- We can decrease the spend on Sponsorship during these months



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

=> We have planned to come up with one final model based on business understanding and accuracy , specificity and sensitivity

Model Summary for Camera Accessory

```
summary(ecommmodel_7)
```

Call:

```
lm(formula = gmv ~ units + sla + product_mrp + discount + i..nps_score,
    data = ecommtrain)
```

Residuals:

Min	1Q Median	3Q	Max
-27333 -202	46	270 38937	

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	67.9946370	49.0634604	1.386	0.166
units	469.9418375	9.1670667	51.264	< 0.0000000000000002 ***
sla	-6.4693353	0.9307241	-6.951	0.000000000000364 ***
product_mrp	0.6649399	0.0005639	1179.137	< 0.0000000000000002 ***
discount	-2380.0723189	10.2633929	-231.899	< 0.0000000000000002 ***
i..nps_score	8.6293435	0.9779172	8.824	< 0.0000000000000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1005 on 161489 degrees of freedom

Model Summary for Camera Accessory

Multiple R-squared: 0.9032, Adjusted R-squared: 0.9032

F-statistic: 3.014e+05 on 5 and 161489 DF, p-value: < 0.000000000000000022

> # Look at VIF of the model again to see the values

> vif(ecommmodel_7)

units	sla	product_mrp	discount	i..nps_score
1.001373	1.007658	1.011407	1.012305	1.001116

> # predicting the results in test dataset

> Predict_1 <- predict(ecommmodel_7,ecommtest[,-2])

> ecommtest\$test_gmv <- Predict_1

> View(ecommtest)

> # Now, we need to test the r square between actual and predicted sales.

> r <- cor(ecommtest\$gmv,ecommtest\$test_gmv)

> r

[1] 0.9516709

> rsquared <- cor(ecommtest\$gmv,ecommtest\$test_gmv)^2

> rsquared

[1] 0.9056776

Coefficients:

Residual standard error: 1005 on 161489 degrees of freedom

Multiple R-squared: 0.9032, Adjusted R-squared: 0.9032

The final model having the following significant variables

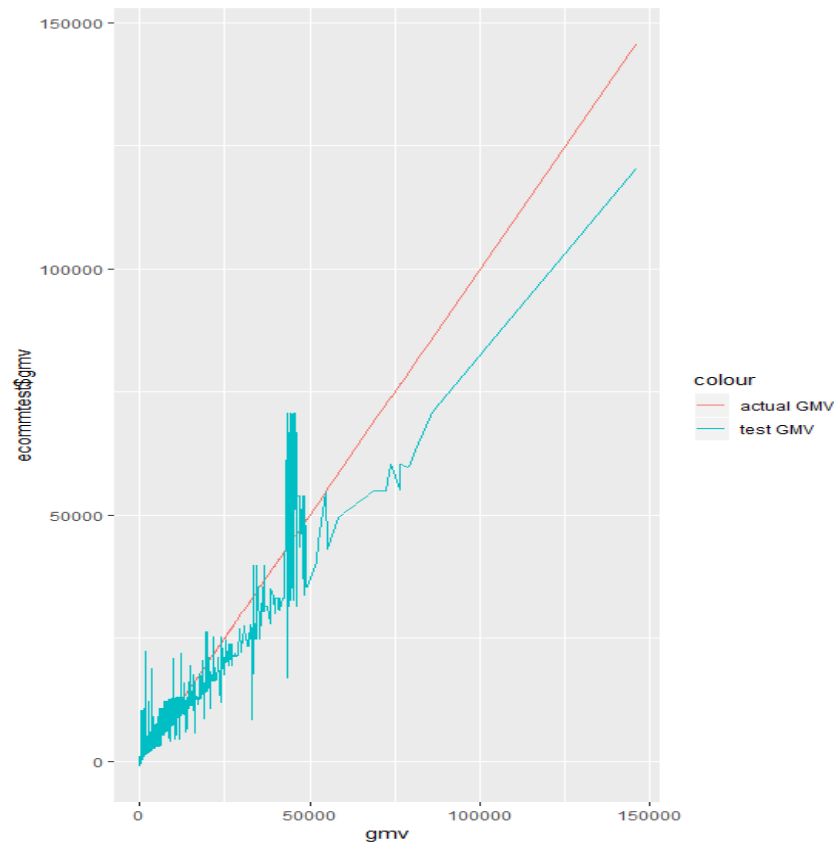
```
> vif(ecommmmodel_7)
```

units	sla	product_mrp	discount	ĩ..nps_score
1.001373	1.007658	1.011407	1.012305	1.001116

Predicting the results in test dataset

```
> r ⇒ 0.9516709
```

```
> rsquared ⇒ 0.9056776
```



Coefficients:

Residual standard error: 182.1 on 83287 degrees of freedom

Multiple R-squared: 0.9911, Adjusted R-squared: 0.9911

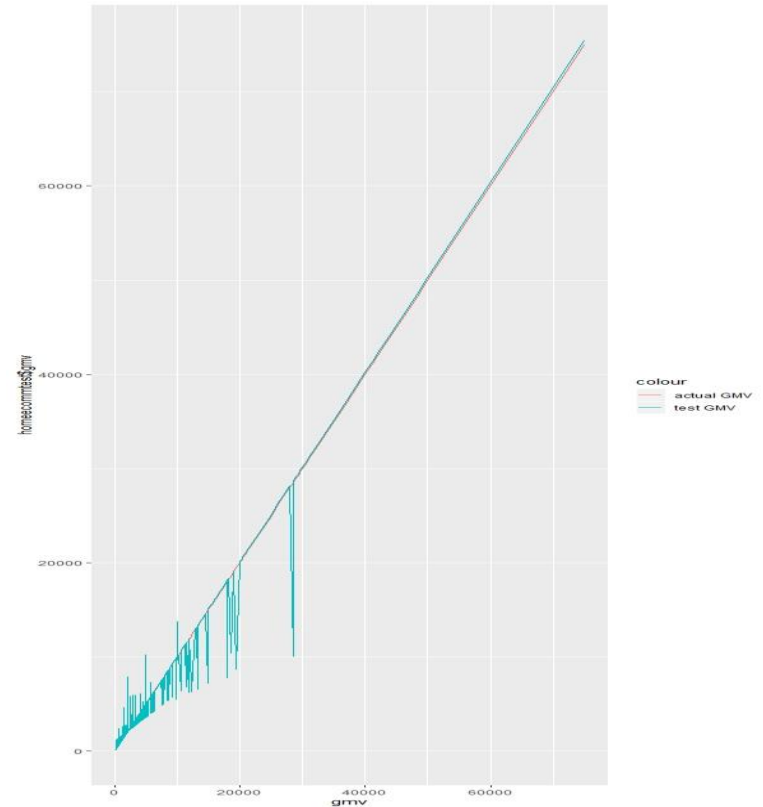
The final model having the following significant variables

units	product_procurement_sla	weeknumber	list_price	Saleday	i..nps_score
1.000945	1.021278	1.032284	1.045222	1.037030	1.054276

Predicting the results in test dataset

> r \Rightarrow 0.9952667

> rsquared \Rightarrow 0.9905558



Coefficients:

Residual standard error: 214.8 on 137880 degrees of freedom

Multiple R-squared: 0.9687, Adjusted R-squared: 0.9687

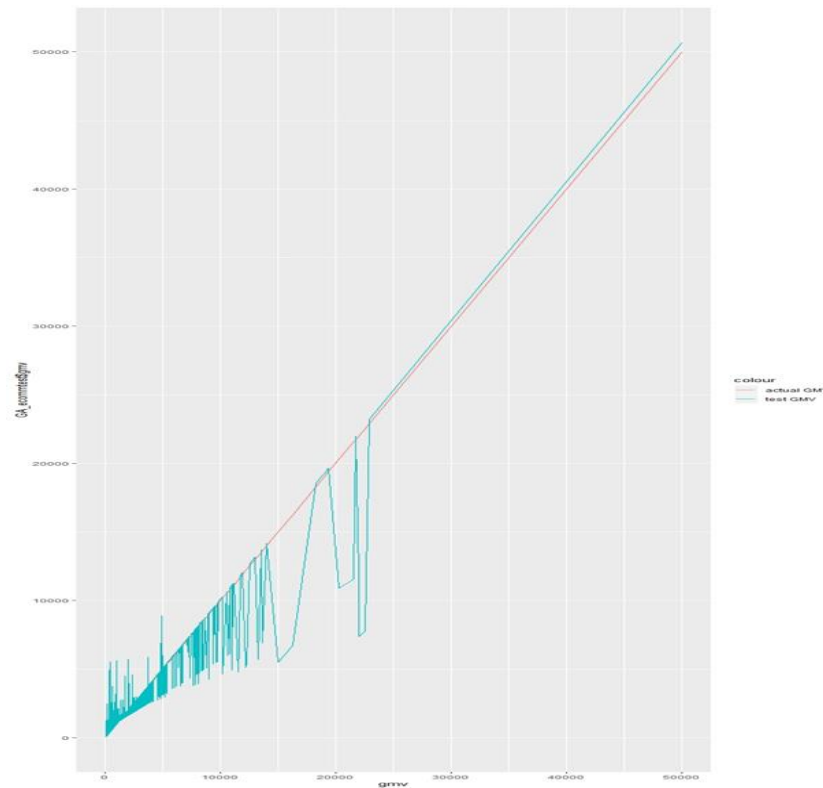
The final model have the following significant variables

units	product_procurement_sla	weeknumber	list_price	discount	Percent
1.001684	1.087474	1.019231	1.127625	1.241926	1.049653

Predicting the results in test dataset

> r \Rightarrow 0.9863499

> rsquared \Rightarrow 0.9728862



Following KPIs have been derived by the insights from EDA

1. GMV: Total value of merchandise sold through a customer-to-customer exchange site
2. Promotional discount : encourage purchases on ecommerce sites
3. Product MRP: marginal value product
4. sla: contract between service provider and the end user
5. NPS : measures the willingness of shoppers to recommend your company to someone
6. Ad stock: advertising on consumer purchase behavior
7. Promotion type : measures effectiveness of promotion type over sales
8. holiday week : measures effectiveness of Holiday week over sales
9. delivery status: early, on time, delayed
10. Media Investment variables – Affiliate, sponsorship, content marketing, Radio

=> Following KPIs were explored with the insights from EDA

1. website traffic (purchased/not purchased): measuring the effectiveness of various marketing channels.
2. Average Order Value: Average amount customers spend when they purchase from website - $\text{Revenue} / \text{Number of Orders}$
3. customer lifetime value: average amount spent by each customer over their lifetime
4. Product Affinity - This tells us that which products are purchased together
5. Customer Retention Rate: evaluate how ecommerce business retains customers
6. Number of sales
7. conversion rate: Sales totals generated by channel such as digital marketing etc.,
8. Return On Ad Spend : measures the efficacy of a digital advertising campaign – $(\text{Revenue} / \text{Media Spend}) * 100\%$
9. Cost Per Action: Monitors the cost to drive subscriptions or transactions - $\text{Media Spend} / \text{Transactions}$

Thank You