ElecKart Market Mix Modeling

Business Understanding & Objective – ElectKart Market Mix Modelling

Background :

ElecKart is an e-commerce firm specialising in electronic products. To enhance their revenues they have done significant investment in their marketing efforts, like promotions over last one year. They are about to create a marketing budget for the next year which includes spending on commercials, online campaigns, and pricing & promotion strategies. They want to reallocate their budget optimally across different marketing levers to improve the revenue response using Market Mix modelling.

Objective :

To develop a market mix model for 3 product sub-categories - Camera accessory, Gaming accessory and Home Audio to observe the actual impact of different marketing levers over sale of last one year (July 2015 -June 2016) and provide an approach .

UpGrad

Data Understanding : Business has provided below 6 dataset for analysis.

Media Investment : This file contains different marketing spend at monthly level.

Column Name	Significance
Year	Year
Month	Month
Total Investment	Monthly Total ad spend in CR
TV	Monthly Total TV ad spend in CR
Digital	Monthly Digital ad spend in CR
Sponsorship	Monthly Sponsorship spend in CR
Content Marketing	Monthly Content marketing spend in CR
Online marketing	Monthly Offline marketing spend in CR
Affiliates	Monthly Affiliates spend in CR
SEM	Monthly SEM spend in CR
Radio	Monthly Radio spend in CR
Other	Monthly Other spend in CR

Special Sale Calendar:

Column Name	Significance
Year	Year
Sales Calendar	List of holidays in given year

Monthly NPS Score: : It contains month wise customer satisfaction score in percentage.

Column Name	Significance
NPS	Customer satisfaction score
Stock Index	Stock Index in that month

Product List:

Column Name	Significance
Product category	Category name
Frequency	Frequency of the products sold
Percent	Percentage w.r.t to total sales

Continued..

Climate data

we need to analyze climate data if it has any effect on the revenue. We need to created as many features as possible with the available data.

Column Name	Significance
Date/Time	Date/Time
Year	Year
Month	Month
Day	Day
Data Quality	Data Quality
Max Temp (°C)	Max Temp (°C)
Max Temp Flag	Max Temp Flag
Min Temp (°C)	Min Temp (°C)
Min Temp Flag	Min Temp Flag
Mean Temp (°C)	Mean Temp (°C)
Mean Temp Flag	Mean Temp Flag
Heat Deg Days (°C)	Heat Deg Days (°C)
Heat Deg Days Flag	Heat Deg Days Flag
Cool Deg Days (°C)	Cool Deg Days (°C)
Cool Deg Days Flag	Cool Deg Days Flag
Total Rain (mm)	Total Rain (mm)
Total Rain Flag	Total Rain Flag
Total Snow (cm)	Total Snow (cm)
Total Snow Flag	Total Snow Flag
Total Precip (mm)	Total Precip (mm)
Total Precip Flag	Total Precip Flag
Snow on Grnd (cm)	Snow on Grnd (cm)
Snow on Grnd Flag	Snow on Grnd Flag
Dir of Max Gust (10s deg)	Dir of Max Gust (10s deg)
Dir of Max Gust Flag	Dir of Max Gust Flag
Spd of Max Gust (km/h)	Spd of Max Gust (km/h)
Spd of Max Gust Flag	Spd of Max Gust Flag

UpGrad

Data Understanding :

Order Details : It contains daily level order details.

Column Name	Significance
fsn_id	The unique identification of each SKU
order_date	Date on which the order was placed
Year	Year of the order
Month	Month of order
order_id	The unique identification number of each order
order_item_id	Suppose you order 2 different products under the same order, it generates 2 different order Item IDs under the same order ID; orders are tracked by the Order Item ID.
GMV	Gross Merchandise Value or Revenue
Units	Number of units of the specific product sold
Deliverybdays	Dispatch delay from Wearhouse
Deliverycdays	Delivery delay to customer
s1_fact.order_payment_type	How the order was paid – prepaid or cash on delivery
Sla	Number of days it typically takes to deliver the product
cust_id	Unique identification of a customer
Pincode	Zip code
product_analytic_sub_category	Product sub category
product_mrp	Maximum retail price of the product
product_procurement_sla	Time typically taken to procure the product

Data Preparation : Data issues

Missing data and data issues in Order Details dataset:

- Missing values in GMV. It is around 0.3%. Some orders are having GMV value of 0
- Few records where GMV value was higher than MRP value.
- Negative values in deliverybdays and deliverycdays column.
- Negative values in customer id and pin code columns.
- Negative values in product_procurement_sla column.
- Outlier test is checked for all the relevant variables and outliers are detected.
- deliverybdays and deliverycdays have more than 75 % of null data

Missing data and data issues in Climate data dataset:

- Columns which does not have more than one unique values, have been discarded.
- several columns e.g. Max Temp Flag , Mean Temp (°C) etc. have missing values. Data is imputed using bfill to fill next row data into these missing data.
- Mean Temp is considered as overall day temperature.

Data Preparation : Data issues continued..

Missing data and data issues in media investments dataset:

- For some Channels like Radio, Others have missing data. We imputed them as 0 assuming there was no expenditure in that mode

Missing data and data issues in special sale calendar dataset:

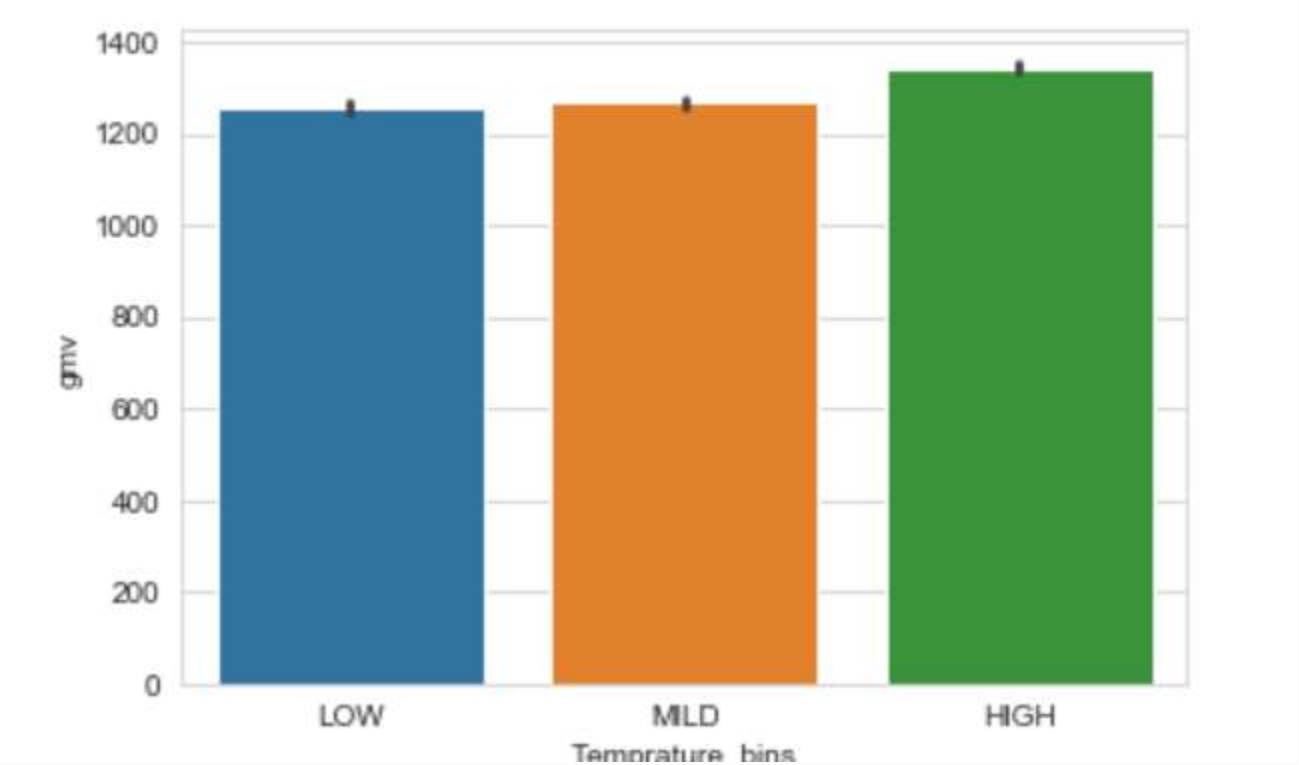
- There were no clear formatted information of holidays. So List is made using that data inside the code.

UpGrad

Data Preparation : Data clean up

- Its is assumed that NPS and stock index will be same for all weeks of that particular month.
- Rows(orders) with missing values for GMV, Customer ID, PIN Code are deleted (Less than 0.5% of entire dataset)
- MRP can't be less than or equal to 0.
- Rows (orders) with 0(MRP) values are deleted, since it is not possible to have product with MRP value as 0.
- deliverybdays and deliverycdays are discarded because of huge null values (79%).
- Negative product_procurement_values are replaced by 0.
- Daily order level data has been aggregated at weekly level for duration between June 2015 to July 2016 for 3 product sub categories - CameraAccessory, Home audio and GamingAccessory.
- Monthly level ad spend has been converted into weekly ad spend.
- Promotional data has been transformed to weekly level which signifies whether that particular week had any promotions, this is derived from Promotion dates given.
- Monthly level NPS score has been considered for each week of the month.
- All different datasets are merged together to form a single master file for carrying out modelling.
- Filtering has been done to create 3 different datasets for CameraAccessory, Home audio and GamingAccessory.

Data Understanding : EDA Analysis



- Days were divided on the basis of climate data after analyzing mean, mode etc..

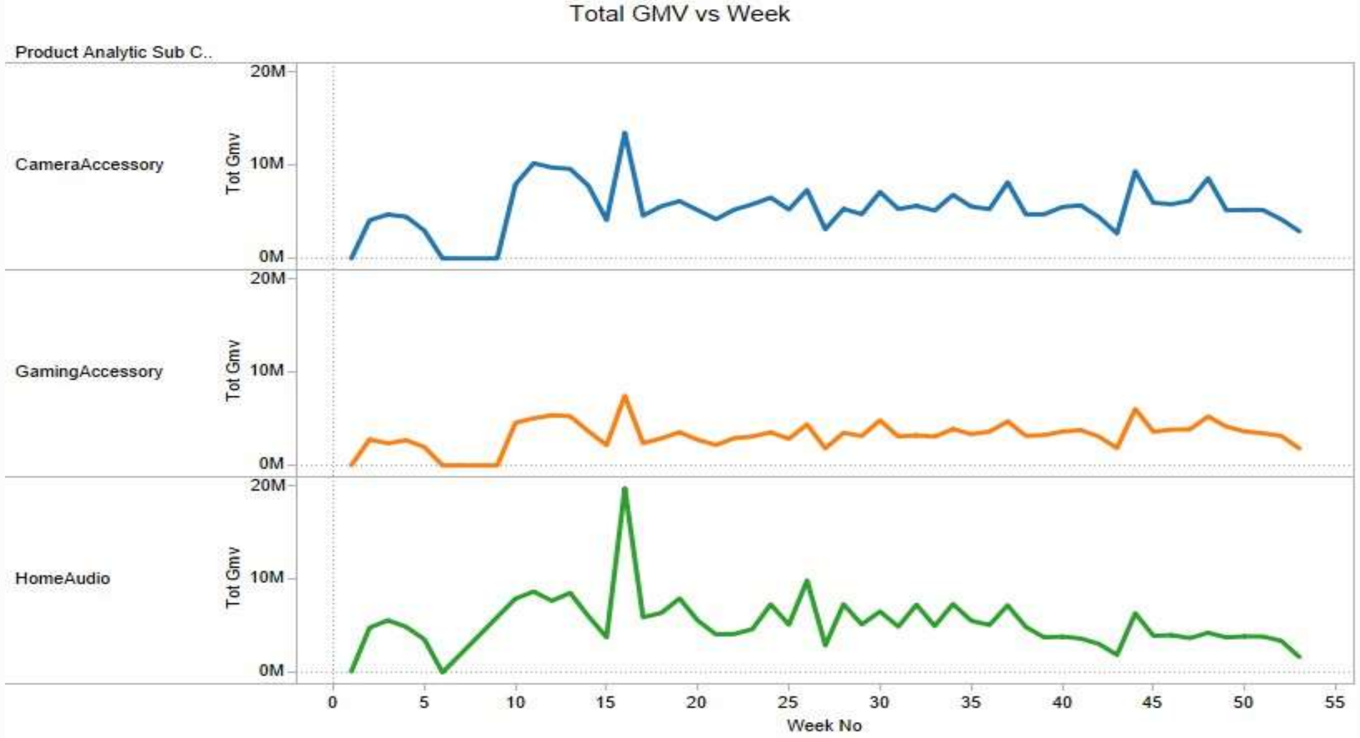
i. LOW temperature days which lies in range [-7,5]

ii. MILD temperature days which lies in range [5,15]

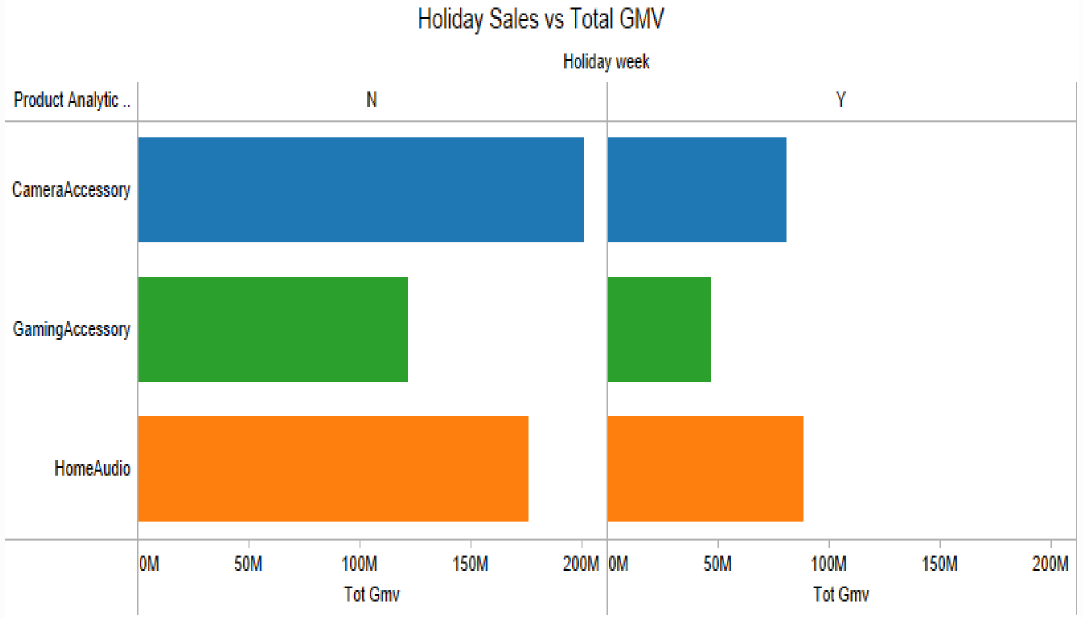
iii. HIGH temperature days which lies in range [15,25]

- We found more revenue generated on Hot days.

Data Understanding : EDA Analysis



The trend of sum of Tot Gmv for Week No broken down by Product Analytic Sub Category. Color shows details about Product Analytic Sub Category. The view is filtered on Product Analytic Sub Category, which keeps CameraAccessory, GamingAccessory and HomeAudio.



Sum of Tot Gmv for each Product Analytic Sub Category broken down by Holiday week. Color shows details about Product Analytic Category. The view is filtered on Product Analytic Sub Category, which keeps CameraAccessory, GamingAccessory and HomeAudio.

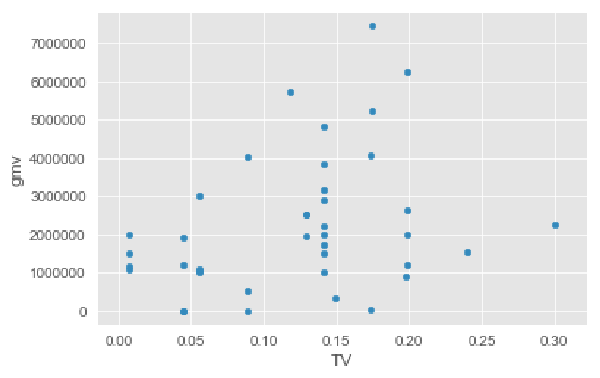
- We can observe that there is sharp spike during the weeks between 10 to 15.
- There are also subtle spikes during the promotional periods across the year.

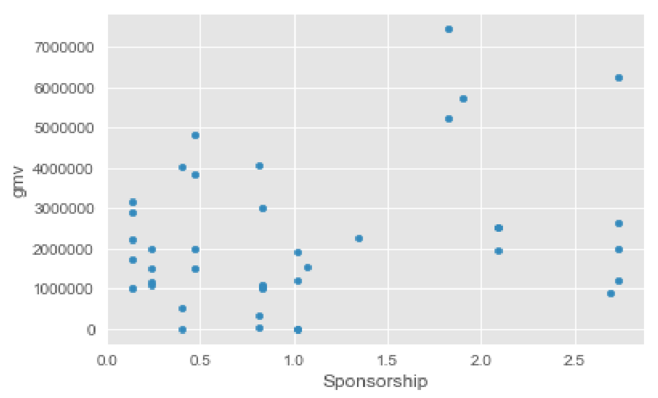
- The ratio of holiday weeks to Non holiday weeks is 1:5.
- Hence, the amount of sales during the Holiday week is far more than the normal week. This suggests holiday weeks are more profitable.

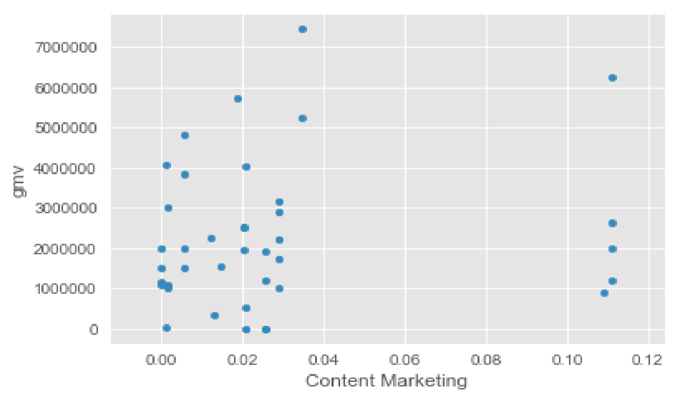
UpGrad

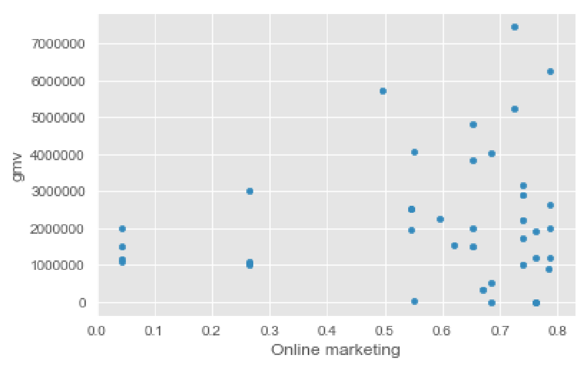
EDA Analysis: Some Visualisations to see the impact of different marketing spends on gmv

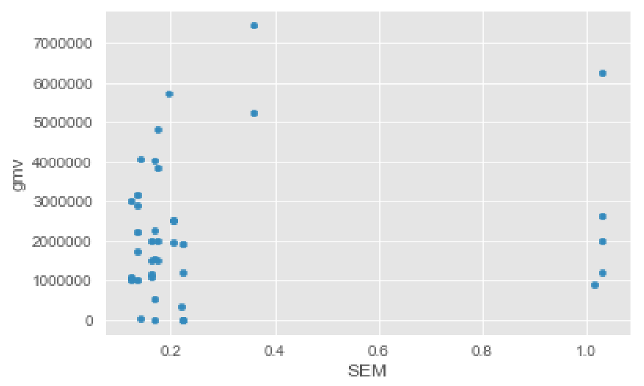
EDA: Identifying out important variables, Category: Game Accessories

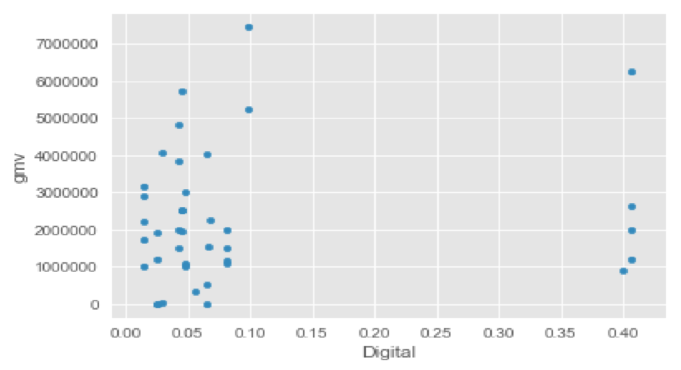








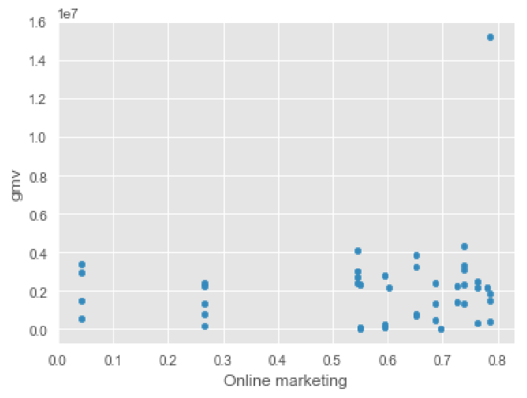


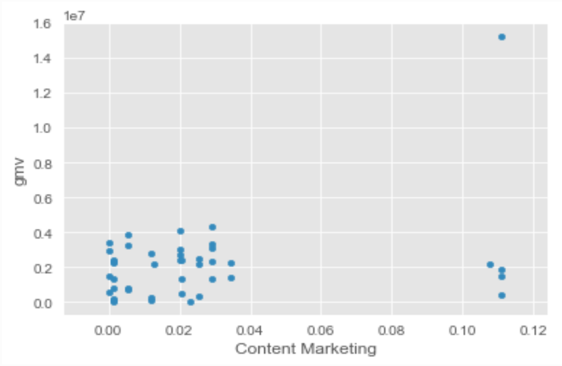


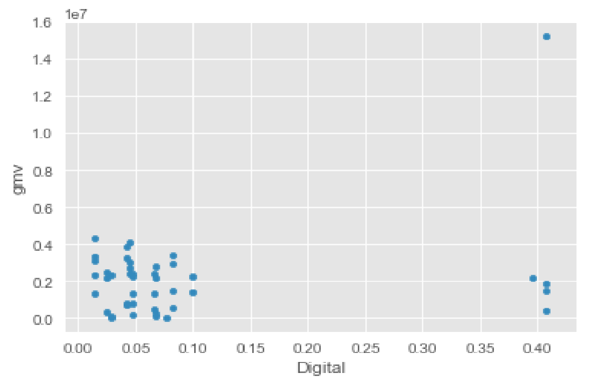
Mostly there are two types of behaviour with ad spend. For few media categories (e.g. TV and Sponsorship) revenue is getting maximized at specific ad spend. Beyond that, with increasing of ad spend, sale is increasing at much lower rate. For few media categories, initially revenue is increasing with ad-spend, then it is almost constant but after a certain ad spend revenue started increasing at much higher rate (e.g. Digital , Content)

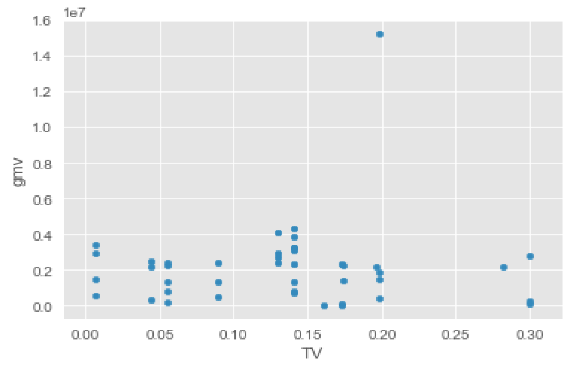
EDA Analysis: Some Visualisations to see the impact of different marketing spends on gmv

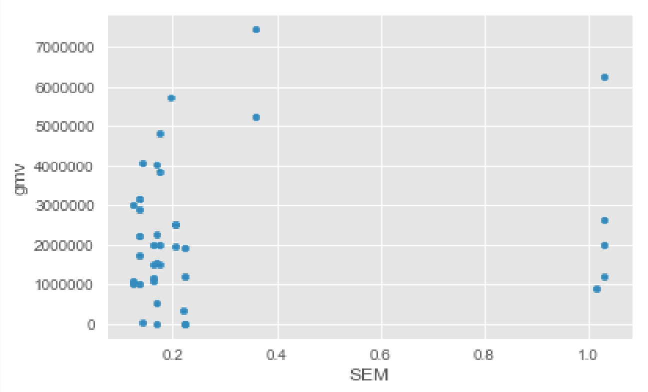
EDA: Identifying out important variables, Category :Camera Accessory

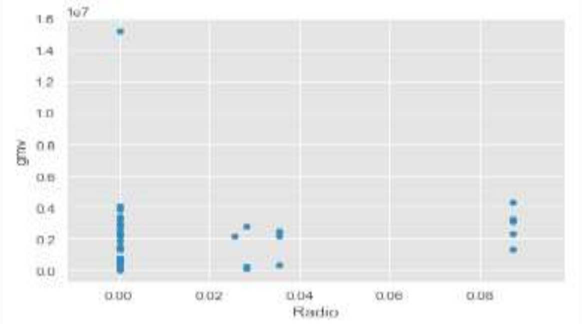








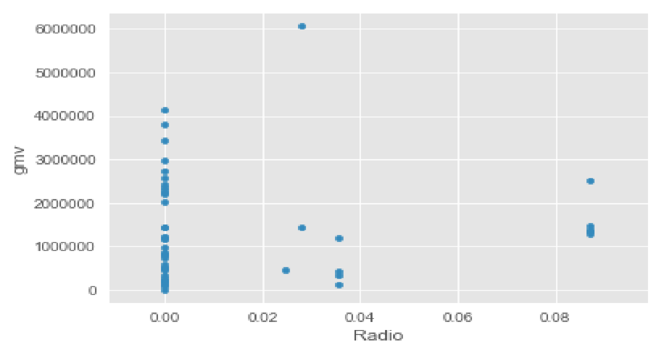


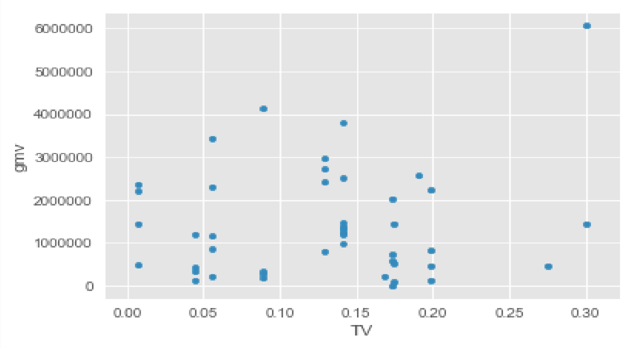


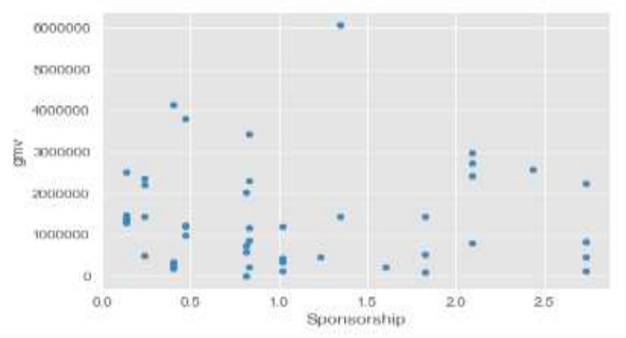
Mostly there are two types of behaviour with ad spend. For few media categories (e.g. TV and Sponsorship) revenue is getting maximized at specific ad spend. Beyond that, with increasing of ad spend, sale is increasing at much lower rate. For few media categories, initially revenue is increasing with ad-spend, then it is almost constant but after a certain ad spend revenue started increasing at much higher rate (e.g. Digital , Content)

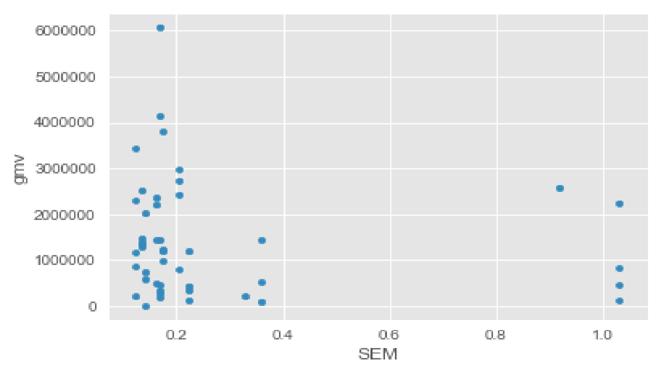
EDA Analysis: Some Visualisations to see the impact of different marketing spends on gmv

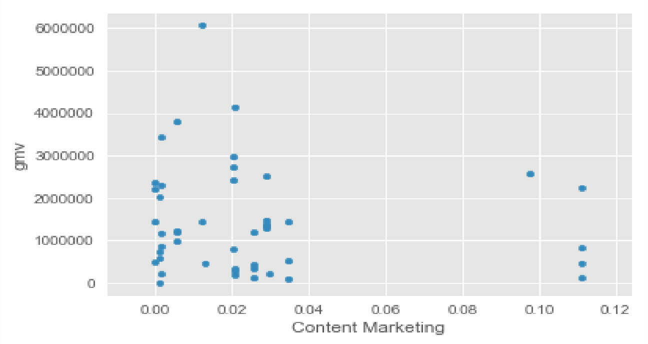
EDA: Identifying out important variables, Category : HomeAudio

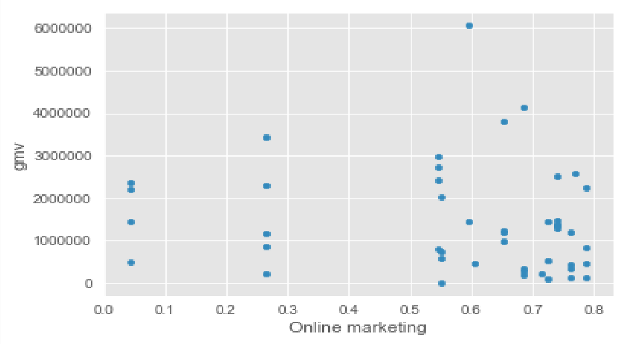












UpGrad

Camera Accessories – Outcome of 4 models

Model	Significant Variables	Adjusted R Square on Train data set	RMSE
Simple Linear Model	CameraTripod GamingGun GamingMouse GamingMousePad	0.918	1488603.89
Multiplicative Model	GamingMousePad sale_price_per_unit GamingMouse SEM GamingAccessoryKit	0.836	29025.61 (log scale)
Koyck Model	GamingGun GamingMouse GamingMousePad	0.907	1168519.29
Distributed Lag Model	BoomBox +CameraRemoteControl+ GamingGun+JoystickGamingWheel+VoiceRecodrder	0.903	4919576.35

Simple linear model is chosen as best model out of 4 models based on high R square and low RMSE value

Game Accessories – Outcome of 4 models

Model	Significant Variables	Adjusted R Square on Train data set	RMSE
Simple Linear Model	BoomBox CameraBatteryGrip CameraBattery	0.227	21659706.79
Multiplicative Model	GamingMousePad CameraBattery product_mrp s1_fact.order_payment_type_Prepaid	0.914	4.80(log scale)
Koyck Model	CameraBattery CameraBatteryGrip	0.142	5337351.49
Distributed Lag Model	CameraHousing GamingAccessoryKit	0.201	1337351.36

Simple Multiplicative model is chosen as best model out of 4 models based on high R square and low RMSE value

UpGrad

Home Audio – Outcome of 4 models

Model	Significant Variables	Adjusted R Square on Train data set	RMSE
Simple Linear Model	GamingHeadset Filter GamePad CameraAccessory CameraBattery	0.484	1219260.20
Multiplicative Model	Lens GamingMousePad product_mrp	0.588	818.93(log scale)
Koyck Model	Binoculars GamingHeadset GamingMemoryCard	0.402	1576222.98
Distributed Lag Model	CameraHousing GamingAccessoryKit	0.532	2830235.55

Simple Multiplicative model is chosen as best model out of 4 models based on high R square and low RMSE value

UpGrad

Recommendations Based on Model Output

Camera Accessories -

Company should promote **Below** product as it has very positive effect on Revenue

CameraTripod
GamingGun
GamingMouse
GamingMousePad

.

Game Accessories -

1- Company should promote **Below** product as it has very positive effect on Revenue

GamingMousePad
CameraBattery

2- People who purchase game accessories are very much MRP price sensitive. It should be reduced.

3- As oppose to Cash on delivery, most game accessories brought in prepaid manner has +ve impact on sale. So There should be less COD option to be given to customer.

.

UpGrad

Recommendations Based on Model Output

Home Audio -

1- Company should promote **Below** product as it has very positive effect on Revenue

Lens
GamingMousePad

2- People who purchase Home accessories are very much MRP price sensitive. This section needs more **Ad Stock.**

.

UpGrad