Capstone Project: Market Mix Modeling

Group Case Study – Ecommerce

By:

Krutika Shahu

Prashant Chaturvedi

Business objective:

To create a **market mix model** for ElecKart (an e-commerce firm based out of Ontario, Canada) for 3 product sub-categories - Camera Accessory, Gaming Accessory and Home Audio, to **observe the actual impact** of **various marketing variables** over one year (July 2015 to June 2016) and **recommend the optimal budget allocation** for different marketing levers for the next year.

The objective is further classified into the following sub-goals:

- Performance driver analysis
- Impact analysis on marketing ROI
- Optimizing marketing spends

Data Understanding:

The data provided for analysis of budget optimization was between July 2015 to June 2016, following are the Files provided**:**

- Main Consumer file with order details at a daily basis
- Media Investment file with amount invested in each advertising medium for the past year
- Sale Calendar file showing dates from past year when there was a promotional offer
- NPS file showing net promotion score and company stock value for last year
- Weather file having detail weather reports from last year in the state of Ontario, Canada

Exploratory Data Analysis

Handling Incorrect values:

- Imputed "\N" value in deliverybdays & deliverycdays columns by 0
- Treating incorrect GMV values (where gmv > product_mrp * units) by imputing the faulty MRP values with GMV/units
- Dropped Negative values for product_procurement_sla, deliverybdays & deliverycdays columns
- Dropped large values(0.3%) for product_procurement_sla columns
- Dropped Columns with Single Unique Value
- Dropped some of the ‘Id’ Columns which are unimportant for the analysis
- After converting all column values to lower case, we see that there are around 99283 (6.33%) rows that are duplicates, hence dropped them
- There were no Null values in the dataframe, however there where **Whitespaces** hence we converted these whitespaces to NaNs and the dropped them
- There were Outliers in the variable such as 'SLA', 'deliverybdays', 'deliverybdays', 'gmv', 'product_mrp', 'list_price’ , so to further not loose any information, we CAPPED the values above 99 percentile to the value corresponding to 99 percentile, rather than dropping it.
- Sorted data between July, 2015 – June, 2016 and dropped all other data out of this date.
- Done binary encoding for categorical variable with 2 levels and Created dummy variables (One Hot Encoding) for categorical variable with multiple levels.
- As per problem statement, Created 3 separate dataframes for 3 product subcategories - camera accessory, home audio and gaming accessory
- Roll Up daily Order Data to Weekly Level by aggregating the numeric variables based on Week#

Feature Engineering

Week#: Generating Week# column from the order date

List Price: List Price = GMV * Units

Payday Week: If Payday falls within the week, then payday week = 1, else 0

Holiday Week: If Holiday falls within the week, then payday week = 1, else 0

Product Type - Luxury / Mass-market: If GMV value is greater than 80 percentile, then luxury, else mass-market

Discount%: Discount% = 100*(product_mrp – list price) / product_mrp

SMA#: 3 & 5-weeks Simple Moving Average for all Advertising media channels, NPS and Stock Index

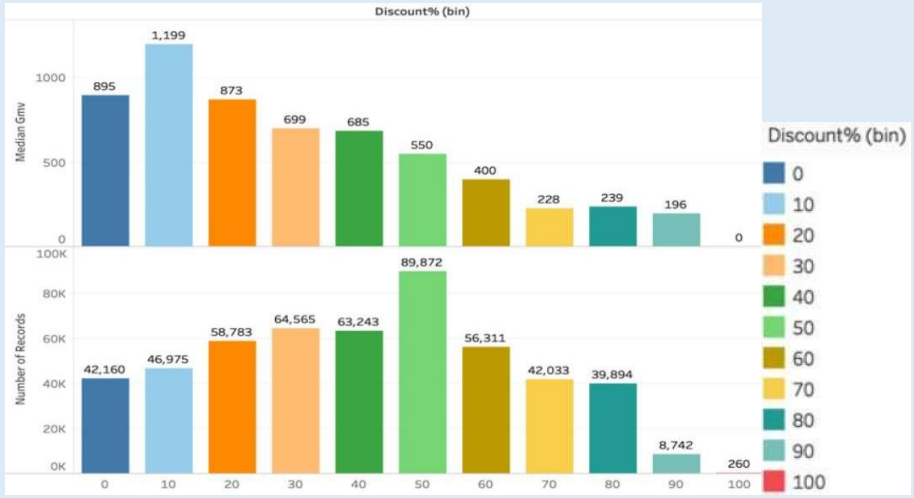
EMA#: 8-weeks Exponential Moving Average for all Advertising media channels

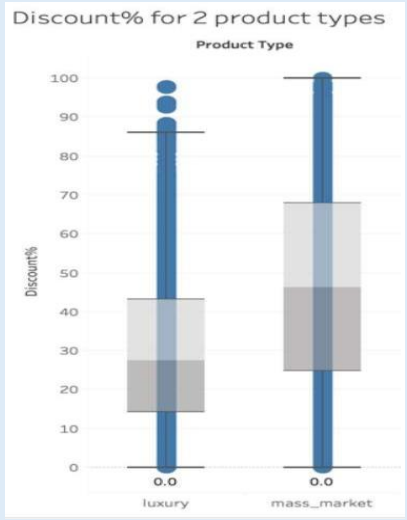
Lag Variables: Lag variables(lag by 1, 2 & 3 days) for all KPIs were taken for Distributive Lag Models

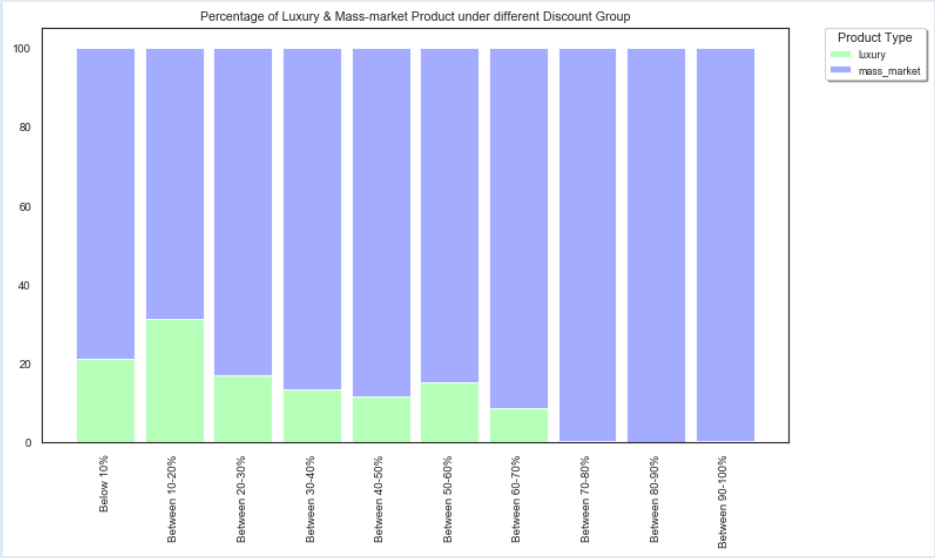
Adstock Values: Calculating Ad Stock values for all Advertising media(assuming ad stock rate as 60%)

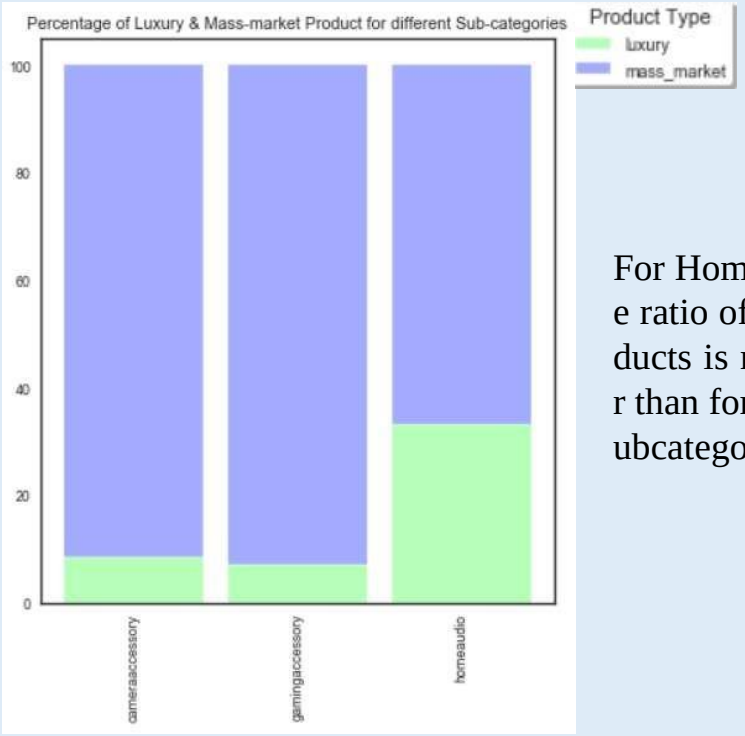
Visulization of data

- Median income is highest if the average discount rate is between 10% and 20%. But beyond that average income is slowly beginning to fall.
- On the other hand, the sales show a steady increase with Discount percentage increase until it peaks at 50-60 percent after which it starts falling again.
- Maximum number of luxury products have been given a 10-20 per cent discount
- It shows that the income falls at higher discount although the revenues are strong, which means a loss to the client. An average 10-20 per cent discount is the company's most profitable.
- The median percentage of discounts offered for luxury items is less than for Mass Market Products. This is a recognized phenomenon among luxury products or luxury brands to offer reduced or no rebates
- The median percentage of discounts given for luxury items is less than for Mass Market Items. This is a common phenomenon among luxury goods or luxury brands to offer small discounts, keeping their items exclusive.

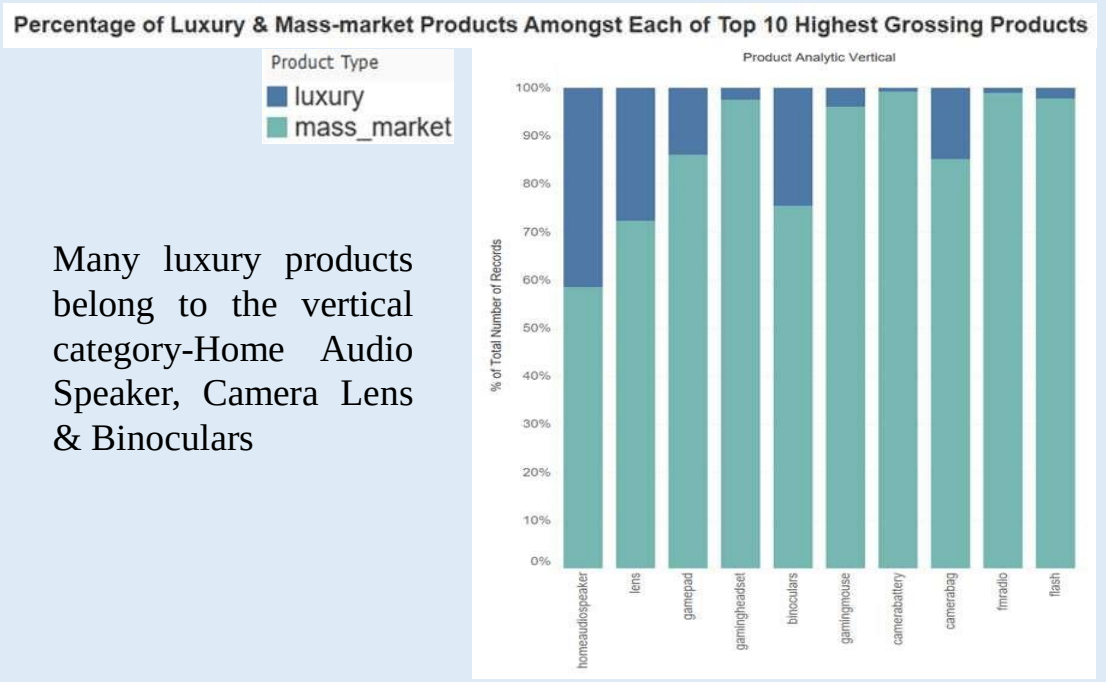




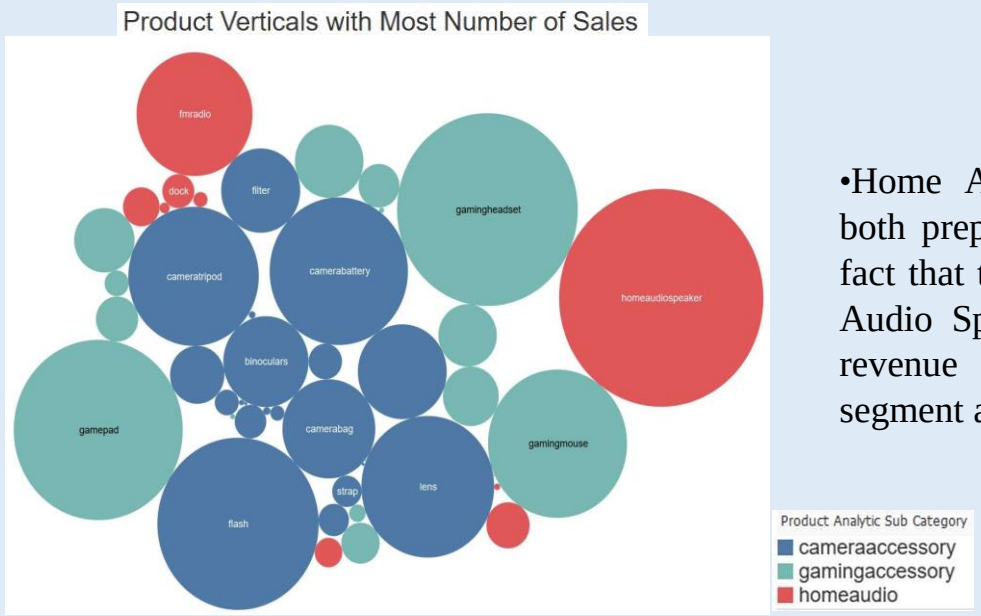




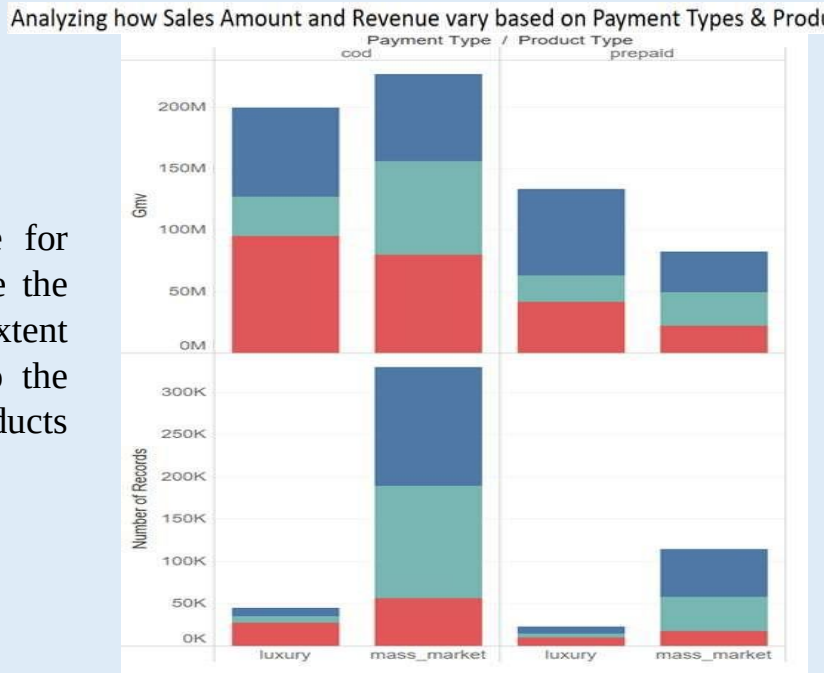
For Home Audio, th e ratio of luxury pro ducts is much highe r than for the other s ubcategories.



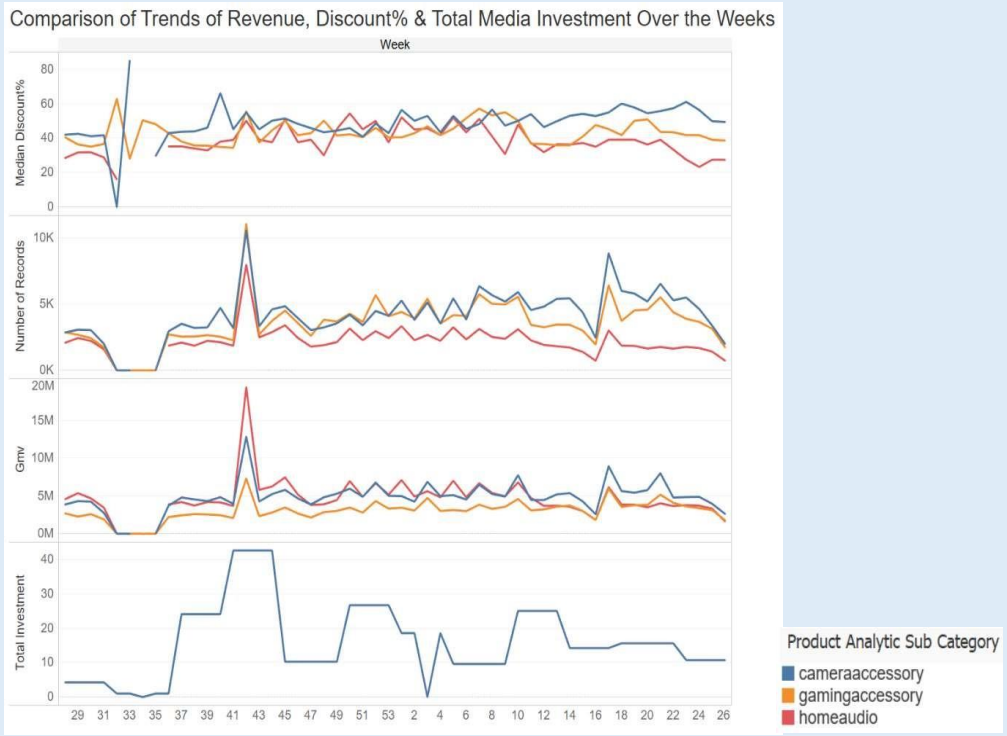
Many luxury products belong to the vertical category-Home Audio Speaker, Camera Lens & Binoculars

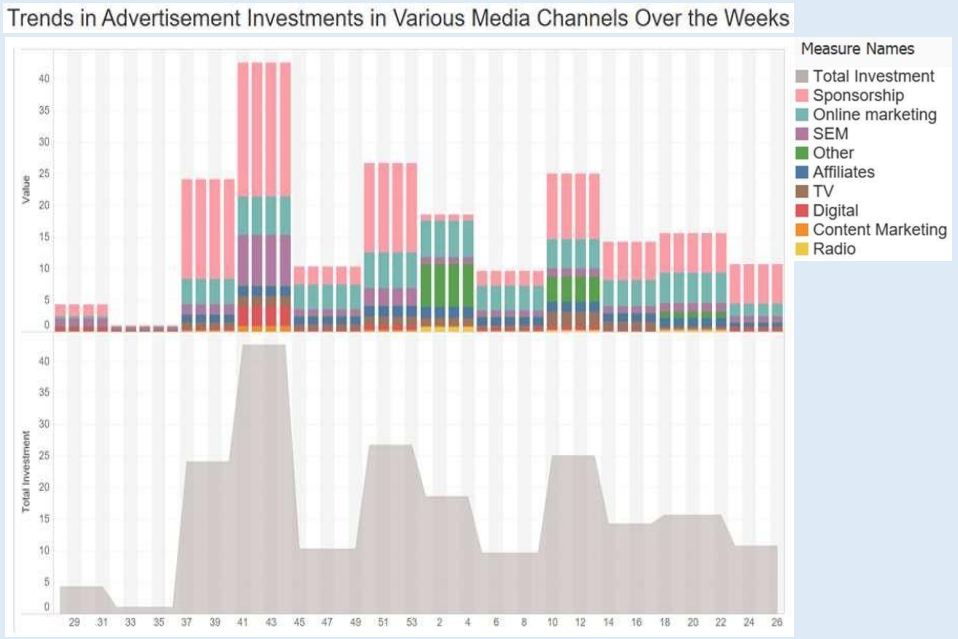


•Home Audio generates more revenue for both prepaid and COD products despite the fact that they are marketed to a lesser extent Audio Speaker primarily contributes to the revenue generated by the COD products segment and generates more revenue.



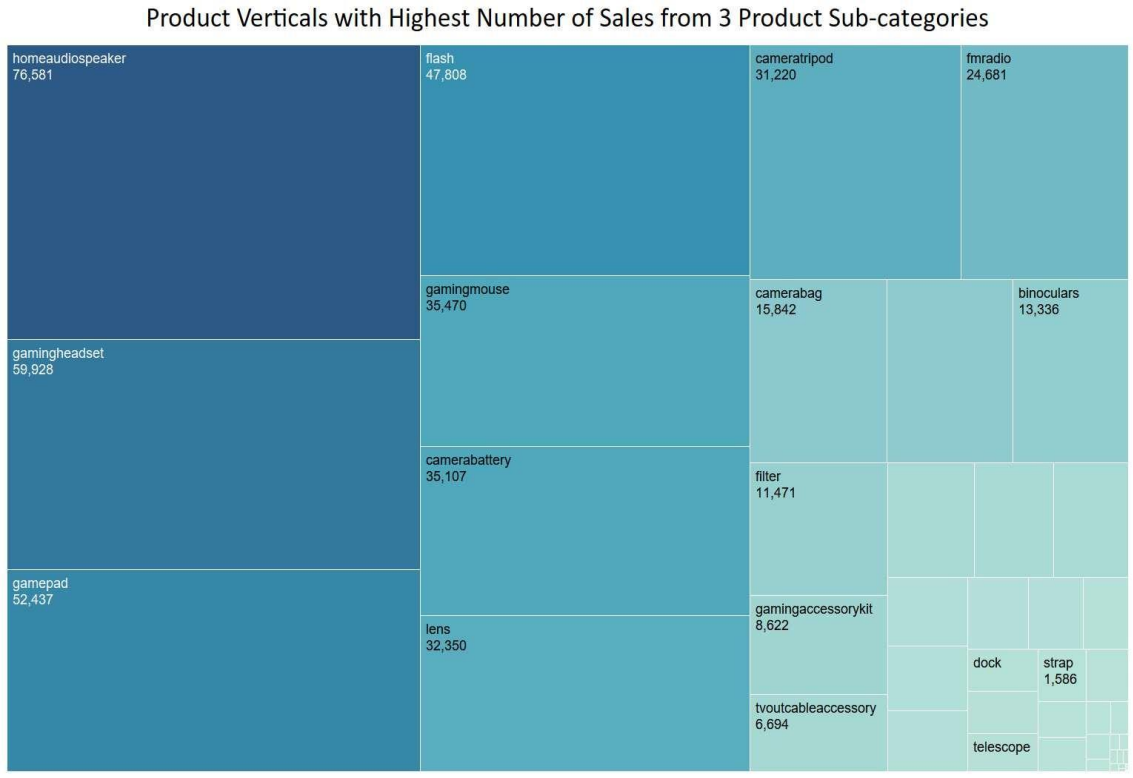
All the graphs show a steep rise for week # 42 (during ' Thanksgiving ').



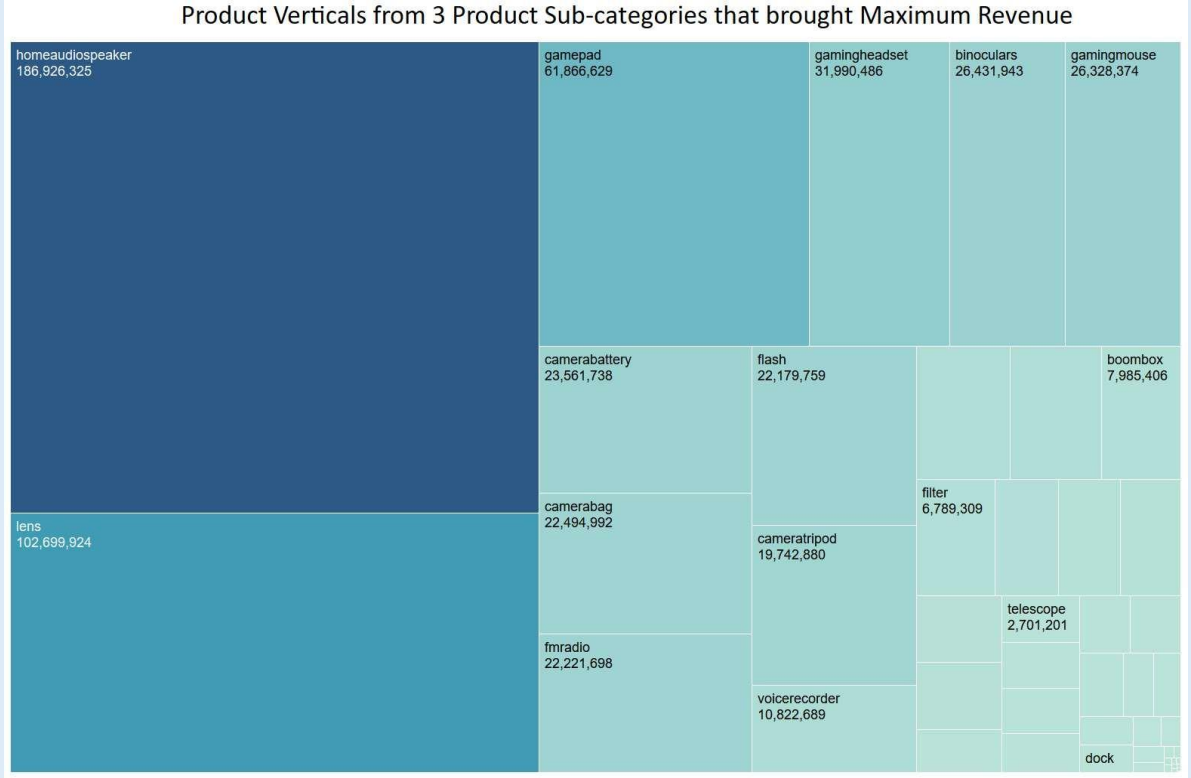


- Revenue generated for the weeks 32-35(August) was the lowest of all 3 subcategories of goods. This can be interpreted as a direct relationship to the minimum total investment level in Ads. Also, discount was lowest for all items except camera accessories. In order to bring in higher sales, discount percent was increased to post this fall in revenue. In the case of gaming, this rise in Discount rate was most noted accessories.
- The total discount level given for home audio items is generally lower relative to the other subcategories of products.
- Most of the Ad Investment was made in sponsorships over the past year, followed by Online Marketing & Search Engine Marketing (especially in Thanksgiving).
- The revenue from other products, barring home audio products, was seen to be constant for the next 3 weeks after which, the revenue began to pick up.

Within the Home Audio category, Home Audio Speaker had the most no sales led by Gaming Headset & Gamepad under Gaming Accessory.



Home Audio Speaker has earned the highest revenue under the Home Audio category, led by Camera Lens under Camera Accessory & Gamepad under Gaming Accessories.



Model Selection

The table below includes descriptions of all the developed versions, their accuracy scores and the top 5 KPIs returned by them:

Product Sub-category	Linear Regression Model	Cross Validation	R2 Score	MSE Score	Top 5 KPIs
cameraaccessory	Additive	No	0.83	0.17	product_vertical_lens, product_vertical_camerabattery, product_vertical_camerabag, product_vertical_camerahousing, Online marketing
		Yes	-0.8	1.08	
	Multiplicative	No	0.84	0.36	**product_vertical_lens, product_vertical_camerabattery, is_mass_market, product_vertical_camerabatterycharger, TV**
		Yes	**0.91**	**0.09**	
	Koyck	No	0.84	0.16	product_vertical_lens, product_vertical_camerabag, product_vertical_camerahousing, product_vertical_camerabattery, Online marketing
		Yes	0.27	0.73	
	Distributive Lag Model (Additive)	No	0.87	0.12	product_vertical_lens, product_vertical_filter, product_vertical_camerabag, product_vertical_cameraremotecontrol, is_mass_market
		Yes	0.82	0.17	
	Distributive Lag Model (Multiplicaitive)	No	0.77	0.5	is_mass_market, product_vertical_lens, product_vertical_cameraaccessory, product_vertical_camerabattery, product_vertical_cameratripod
		Yes	0.82	0.18	
gamingaccessory	Additive	No	0.93	0.05	product_vertical_gamepad, product_vertical_gamingheadset, is_mass_market, product_vertical_gamingaccessorykit, product_vertical_gamingmouse
		Yes	0.51	0.49	
	Multiplicative	No	0.94	0.09	**product_vertical_gamingheadset, is_mass_market, product_vertical_gamingmouse, product_vertical_gamepad, Online marketing_SMA_3**
		Yes	**0.94**	**0.06**	
	Koyck	No	0.93	0.05	product_vertical_gamepad, product_vertical_gamingheadset, is_mass_market, product_vertical_gamingaccessorykit, product_vertical_gamingmouse
		Yes	0.49	0.51	
	Distributive Lag Model (Additive)	No	0.87	0.1	product_vertical_gamepad, product_vertical_gamingaccessorykit, is_mass_market, product_vertical_motioncontroller, product_vertical_gamingkeyboard
		Yes	0.92	0.08	
	Distributive Lag Model (Multiplicaitive)	No	0.93	0.11	product_vertical_gamepad, product_vertical_gamingmouse, is_mass_market, product_vertical_gamingkeyboard, is_cod
		Yes	0.89	0.11	
homeaudio	Additive	No	0.96	0.09	product_vertical_homeaudiospeaker, is_mass_market, Digital_SMA_3, product_vertical_fmradio, is_cod
		Yes	0.73	0.27	
	Multiplicative	No	-0.63	0.34	**product_vertical_homeaudiospeaker, is_mass_market, product_vertical_fmradio, Radio_Ad_Stock, Sponsorship**
		Yes	**0.86**	**0.14**	
	Koyck	No	0.96	0.09	product_vertical_homeaudiospeaker, is_mass_market, is_cod, NPS, Mean Temp
		Yes	0.7	0.3	
	Distributive Lag Model (Additive)	No	0.42	1.39	product_vertical_homeaudiospeaker, product_vertical_karaokeplayer, is_mass_market, is_cod, product_vertical_fmradio
		Yes	0.53	0.47	
	Distributive Lag Model (Multiplicaitive)	No	-0.23	0.26	product_vertical_homeaudiospeaker, is_mass_market, product_vertical_fmradio, is_cod, product_vertical_voicerecorder
		Yes	0.57	0.43	

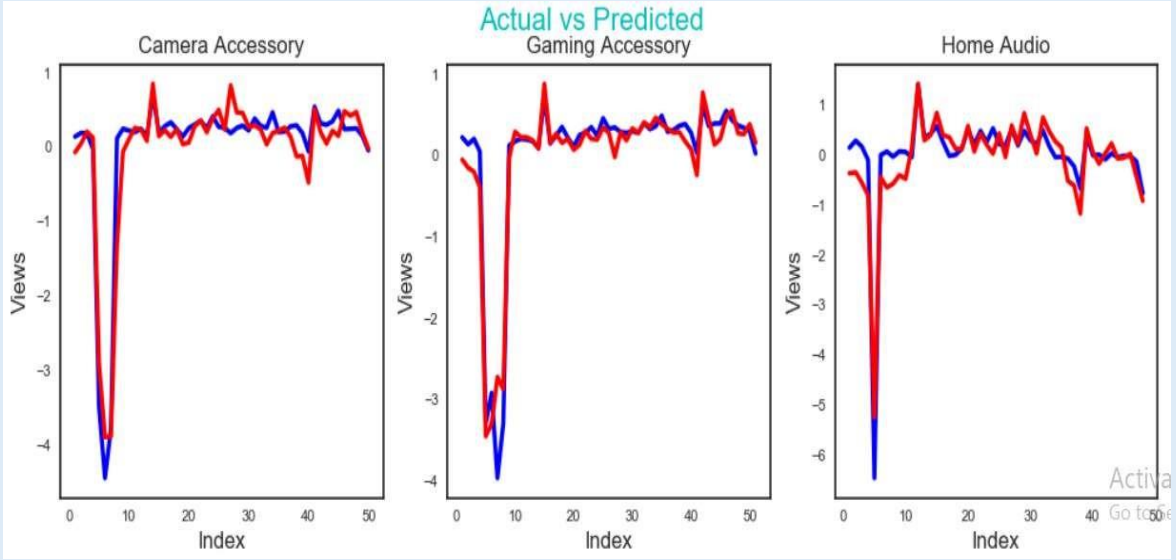
- The model's selection criteria are based on the precision parameters - R2 score & MSE score and the market significance of the model's selected important attributes.
- We also tried to select models with cross validation because, although those without, sometimes give us good scores, due to limited dataset, they are not very reliable and generalizable.
- By referring to the dashboard model, we are finalizing the following models for the 3 product subcategories listed- Camera Accessory, Gaming Accessory & Home Audio:

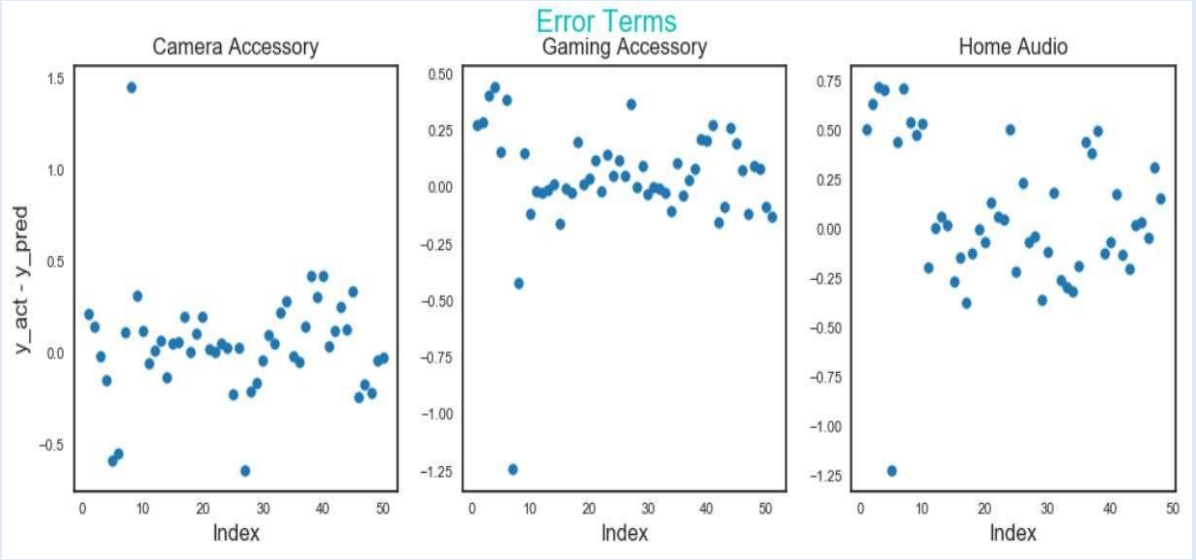
Product Sub-category	Linear Regression Model	R-square on Test Dataset	Mean Square Error	Top 5 KPIs
cameraaccessory	Multiplicative with CV	0.91	0.09	product_vertical_lens (**0.181**)
				product_vertical_camerabattery (**0.160**)
				is_mass_market (**0.149**)
				product_vertical_camerabatterycharger (**0.121**)
				TV (**0.105**)
gamingaccessory	Multiplicative with CV	0.94	0.06	product_vertical_gamingheadset (**0.250**)
				is_mass_market (**0.234**)
				product_vertical_gamingmouse (**0.224**)
				product_vertical_gamepad (**0.211**)
				Online marketing_SMA_3 (**0.157**)
cameraaccessory	Multiplicative with CV	0.86	0.14	product_vertical_homeaudiospeaker (**0.469**)
				is_mass_market (**0.289**)
				product_vertical_fmradio (**0.224**)
				Radio_Ad_Stock (**0.147**)
				Sponsorship (**0.121**)

- We note that all 3 models chosen for the 3 subcategories are Multiplicative ones.
- This finding shows us that for all of the 3 model there is some conflict between the KPIs.
- These models tell us about the revenue growth versus the KPIs ' interactive growth.

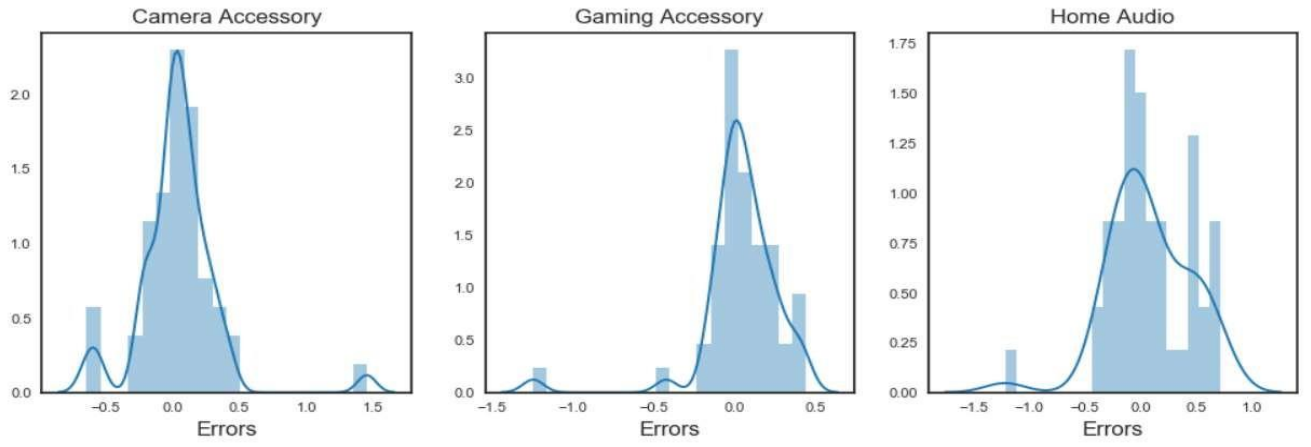
Model Validation

Drawing the Error Terms scatter plot to test the distribution to ensure continuous variation (homoscedasticity) of the error terms.
As the error values change, the variance does not increase or decrease, or obey a trend.





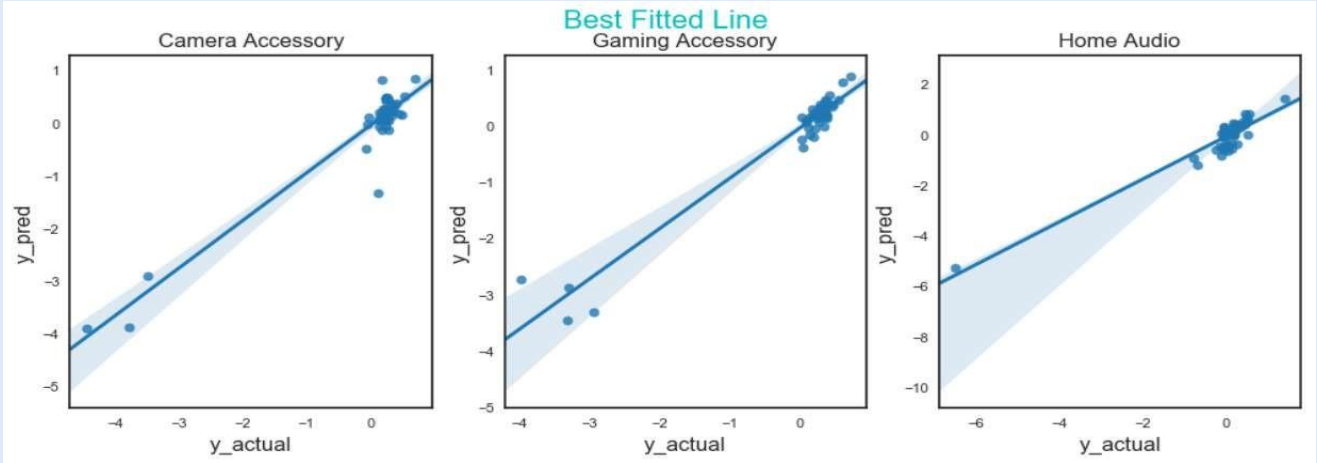
To test the correlation, map the real and expected price values from the dataset.



Plotting of terminology for the error distribution.

The error terms obey a standard 0-mean distribution, with a few outlier values missing.

Plotting a scatter plot from the dataset of current and expected price values to test the distribution and draw through the best suited rows.



Considering the top 5 KPIs from the models for our 3 product subcategories, we can see that the equation of our best fitted lines as follows:

Camera Accessory

- Revenue = 0.0 + (0.181 × product_vertical_lens) + (0.160 × product_vertical_camerabattery) + (0.149 × is_mass_market) + (0.121 × product_vertical_camerabatterycharger) + (0.105 × TV) + …

Gaming Accessory

- Revenue = 0.0 + (0.250 × product_vertical_gamingheadset) + (0.234 × is_mass_market) + (0.224 × product_vertical_gamingmouse) + (0.211 × product_vertical_gamepad) + (0.157 × Online marketing_SMA_3) + …

Home Audio

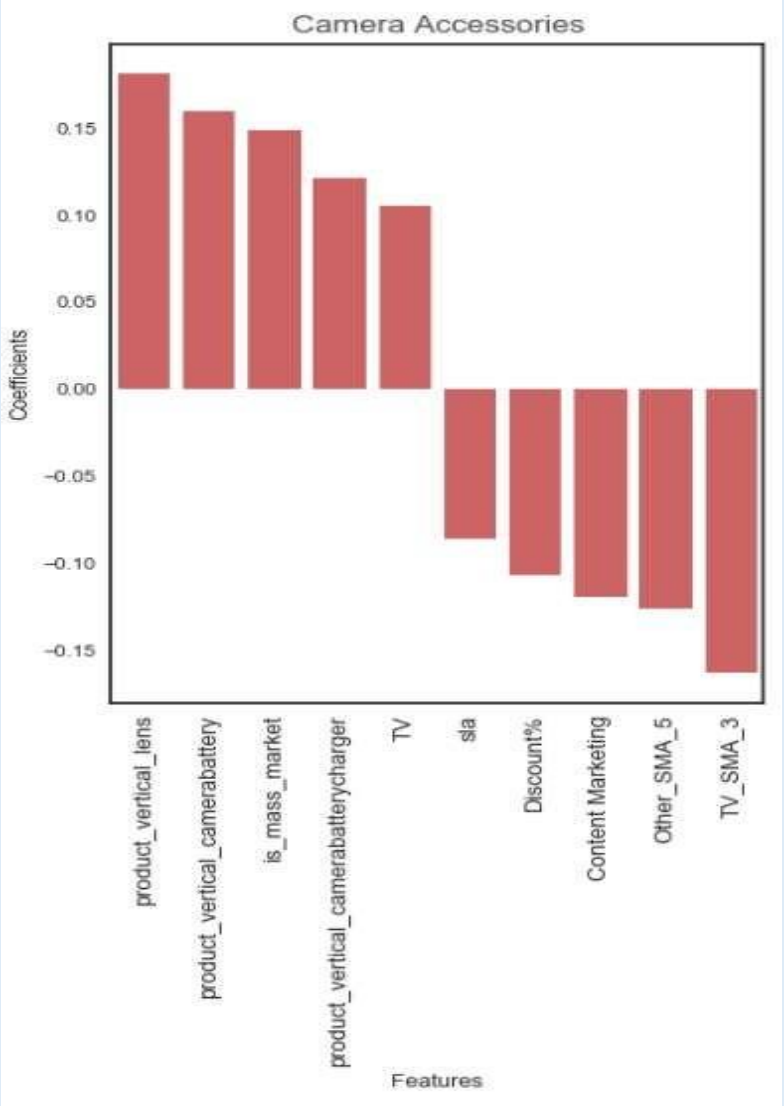
- Revenue = 0.0 + (0.469 × product_vertical_homeaudiospeaker) + (0.289 × is_mass_market) + (0.224 × product_vertical_fmradio) + (0.147 × Radio_Ad_Stock) + (0.121 × Sponsorship) + …

This equation implies how revenue can grow at any of these independent KPIs with a unit growth with all other KPIs held constant.

Recommendations

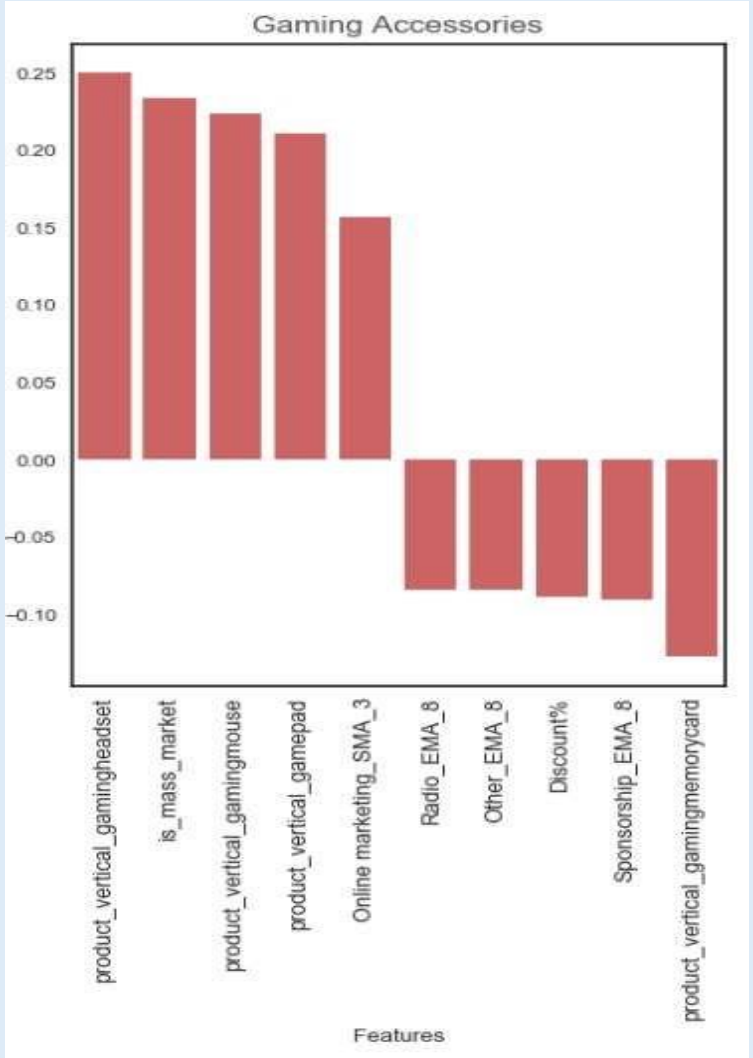
Camera Accessory :

- Business will support ' Lens, ' ' Video Batteries ' & ' Camera Battery Chargers ' because they earn the highest income.
- TV ad investment has a positive impact on sales. One unit of television advertising will boost revenue by 0.105 units. In the other hand, digital marketing spending has negative impacts.
- Compared to the Luxury products, ' Mass Consumer ' products are greater contributors to the increased revenue.
- Higher percentage of generally given discounts for this sub-category works adversely towards rising profits.



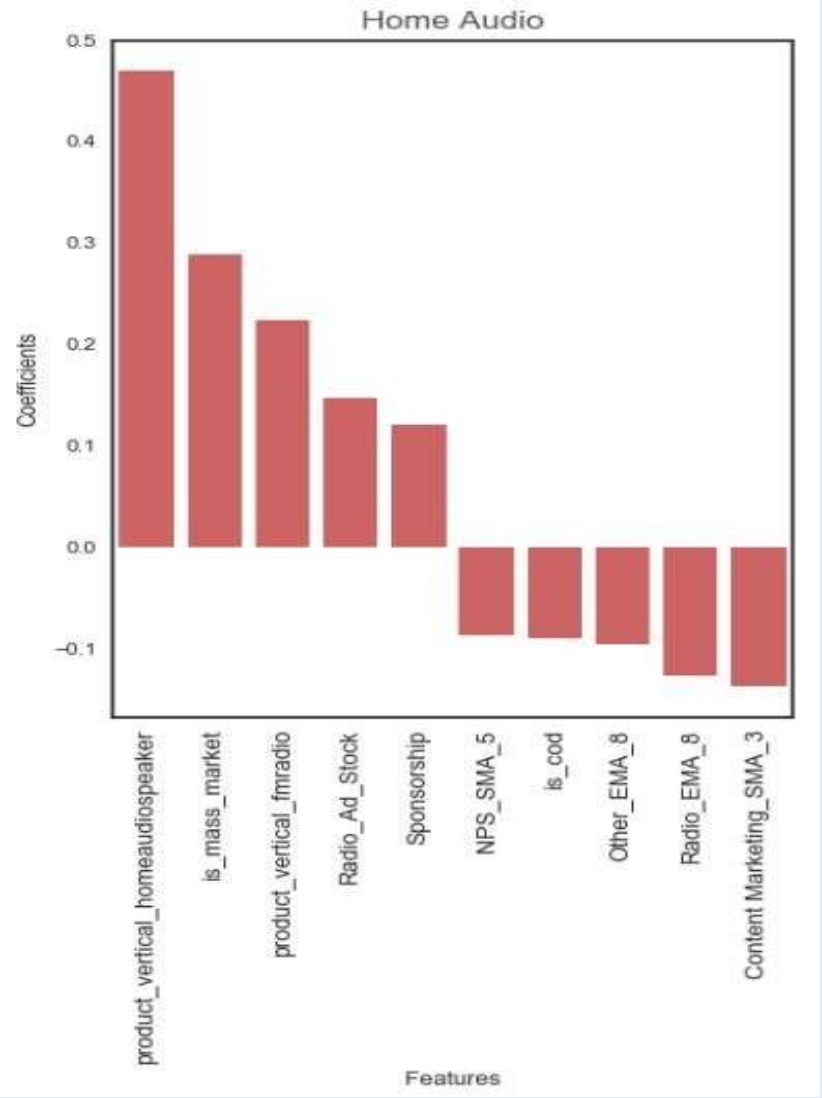
Gaming Accessory :

- Business should be promoting ' Gaming Helmet, ' ' Gaming Mouse ' & ' Gamepad ' as they receive the highest income. Instead, ' Gaming Memory Cards ' leads to loss.
- Digital media, radio & other advertisement spending has a significant cumulative impact on sales. In turn, marketing investment has a negative cumulative effect.
- `Similar to the Luxury products, mass-market products are greater contributors to the increased revenue.
- In general, higher percentage of Discounts offered for this sub-category works adversely against raising sales.



Home Audio :

- Business should be promoting ' Home Audio Speakers ' & ' FM Radios ' as they get the highest sales.
- ' Mass-market ' goods contribute more to the increased revenue relative to the Luxury products.
- Spending on Radio Adstock (carry over impact of Radio Advertisement) continues to significantly boost the sales.
- Spending on Sponsorship ads has a positive impact on sales.On the other hand, content marketing spending has negative impacts.
- COD payments for this sub-category in general are poor at raising sales.



Overall Recommendation:

- Most purchases occur when the discount range ranges from 50-60%. That doesn't necessarily help to boost revenue however. EDA shows that an average 10-20 percent discount is the most lucrative, particularly among luxury items, for the group.
- Generally, most of the Home Audio items sold are luxury items and therefore customers prefer to use COD rather than pay up front.
- More investment is made in Advertising during holiday time (e.g. Thanksgiving), and successful promotional deals were carried out. It usually improves the profits. But just offering promotions on several media channels without sufficient ads for it doesn't help. We have seen that the revenue generated from all 3 drug subcategories was the lowest for the weeks 32-35(August), even though the median discount percentage was raised after the initial drought. In addition, this sales fall can be seen during the given timeframe as a direct relation to the minimum amount of total expenditure in Ads.