Veer Pal Singh
Data Scientist
March 20, 2022

Table of contents

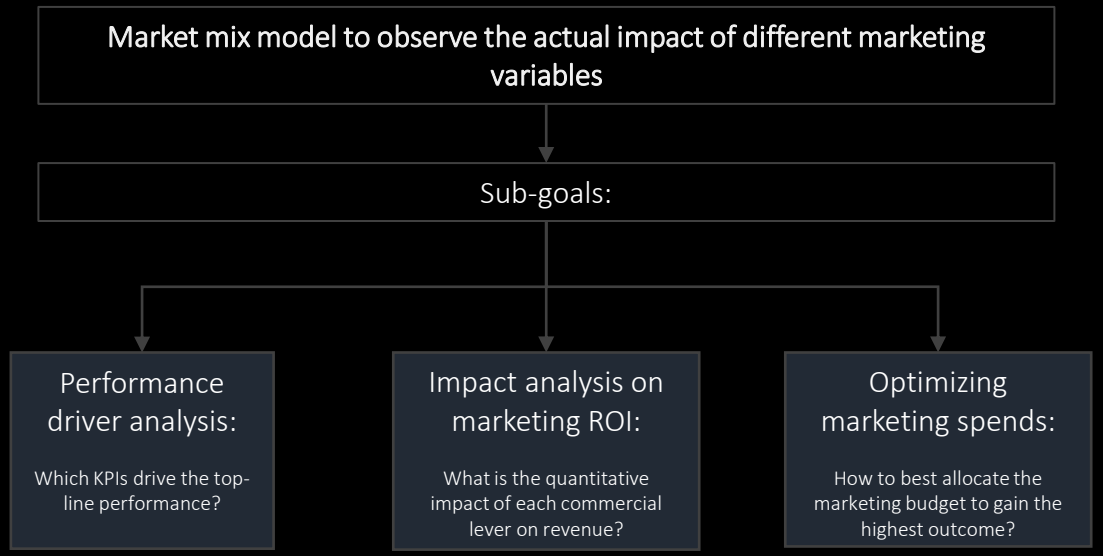
Market Mix Model: Budget Optimization

ElecKart is an e-commerce firm based out of Ontario, Canada specializing in electronic products. Over the last one year, they had spent a significant amount of money on marketing. Occasionally, they had also offered big-ticket promotions (similar to the Big Billion Day). They are about to create a marketing budget for the next year, which includes spending on commercials, online campaigns, and pricing & promotion strategies. The CFO feels that the money spent over the last 12 months on marketing was not sufficiently impactful, and, that they can either cut on the budget or reallocate it optimally across marketing levers to improve the revenue response using Market Mix Model.

Imagine that you are a part of the marketing team working on budget optimization. You need to develop a market mix model to observe the actual impact of different marketing variables over the last year. Using your understanding of the model, you have to recommend the optimal budget allocation for different marketing levers for the next year.

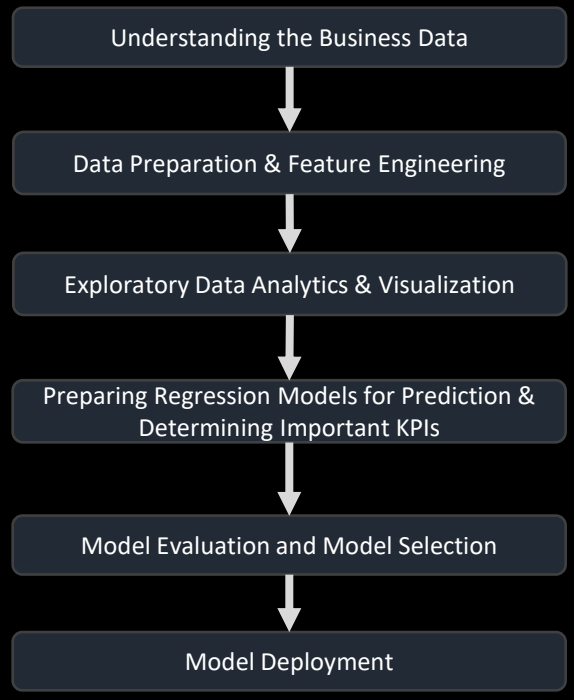
Business Objective

To develop 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.



Methodology

The approach for this project has been designed to follow the CRISP-DM Framework. The various stages of the framework are represented below in a sequential flow:



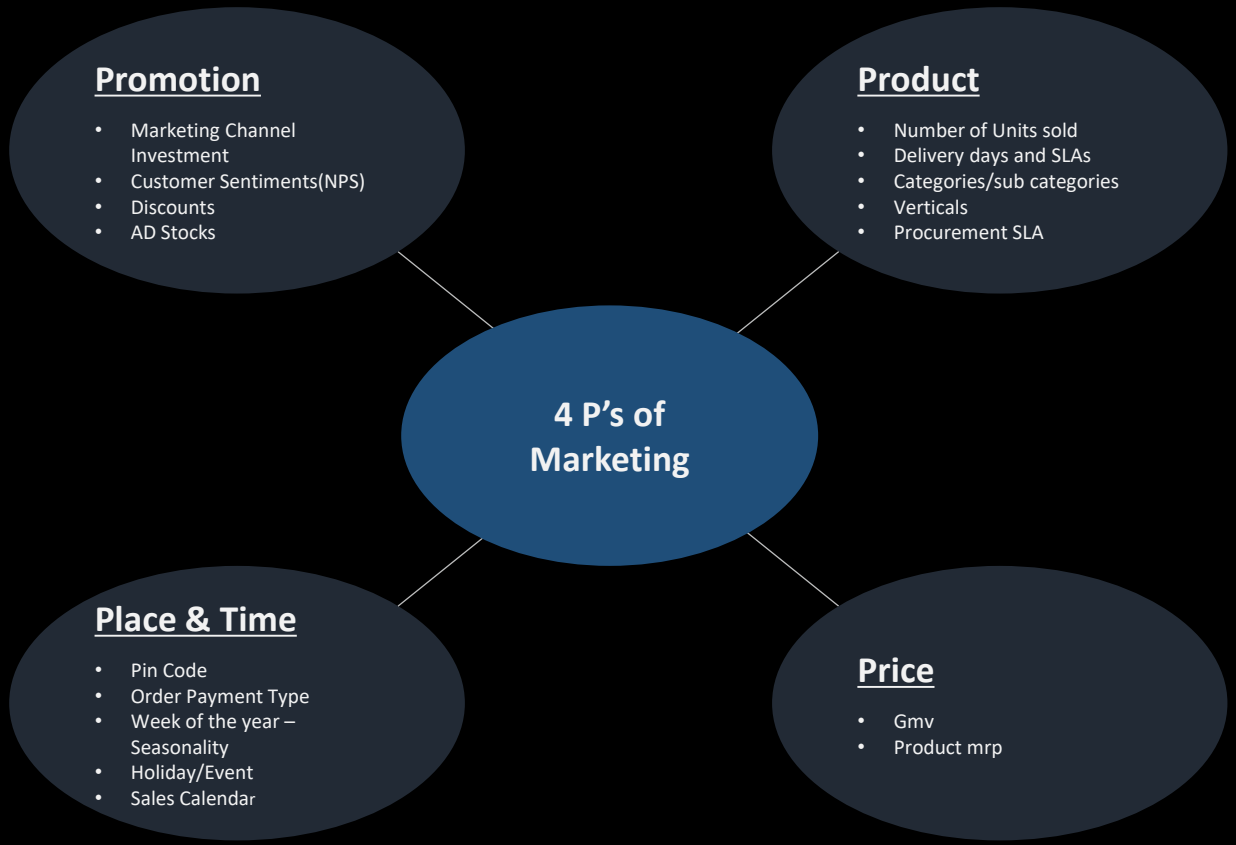
Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Description of Datasets

The data was collected from multiple sources with the following information:

- Main Consumer file with order details had features such as order date, order ID, gross merchandise value, number of units sold for specific products, etc.
- Media Investment file with monthly spend on various advertising channels the amount invested in each advertising medium for the past year
- Sale Calendar file showing dates from the past year when there was a promotional offer
- Monthly NPS score file showing net promotion score and company stock value for last year
- Stock Index of the company on a monthly basis
- Product details with information like product category and vertical (camera accessory, home audio, and gaming accessory)
- Weather file having detailed weather reports from last year in the state of Ontario, Canada

Product, price, promotion, and place form the four Ps of the marketing mix. These are the key factors that are involved in introducing a product or service to the public.



Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Data Preparation and Data Cleaning

Data preparation is the process of preparing raw data so that it is suitable for further processing and analysis. Key steps include collecting, cleaning, and labeling raw data into a form suitable for machine learning (ML) algorithms and then exploring and visualizing the data.

Handling Incorrect values in some columns

- Imputing "\N" value in delivery days &delivery days by 0
- Treating incorrect GMV values (where gmv > product_mrp units) by imputing the faulty MRP values with GMV/units
- Handling Negative values for product_procurement_sla, deliverybdaysdelivery daysays by dropping them
- Handling large values(0.3%) for product_procurement_sla by dropping them

De-Duplication ahead and dropped them of Data

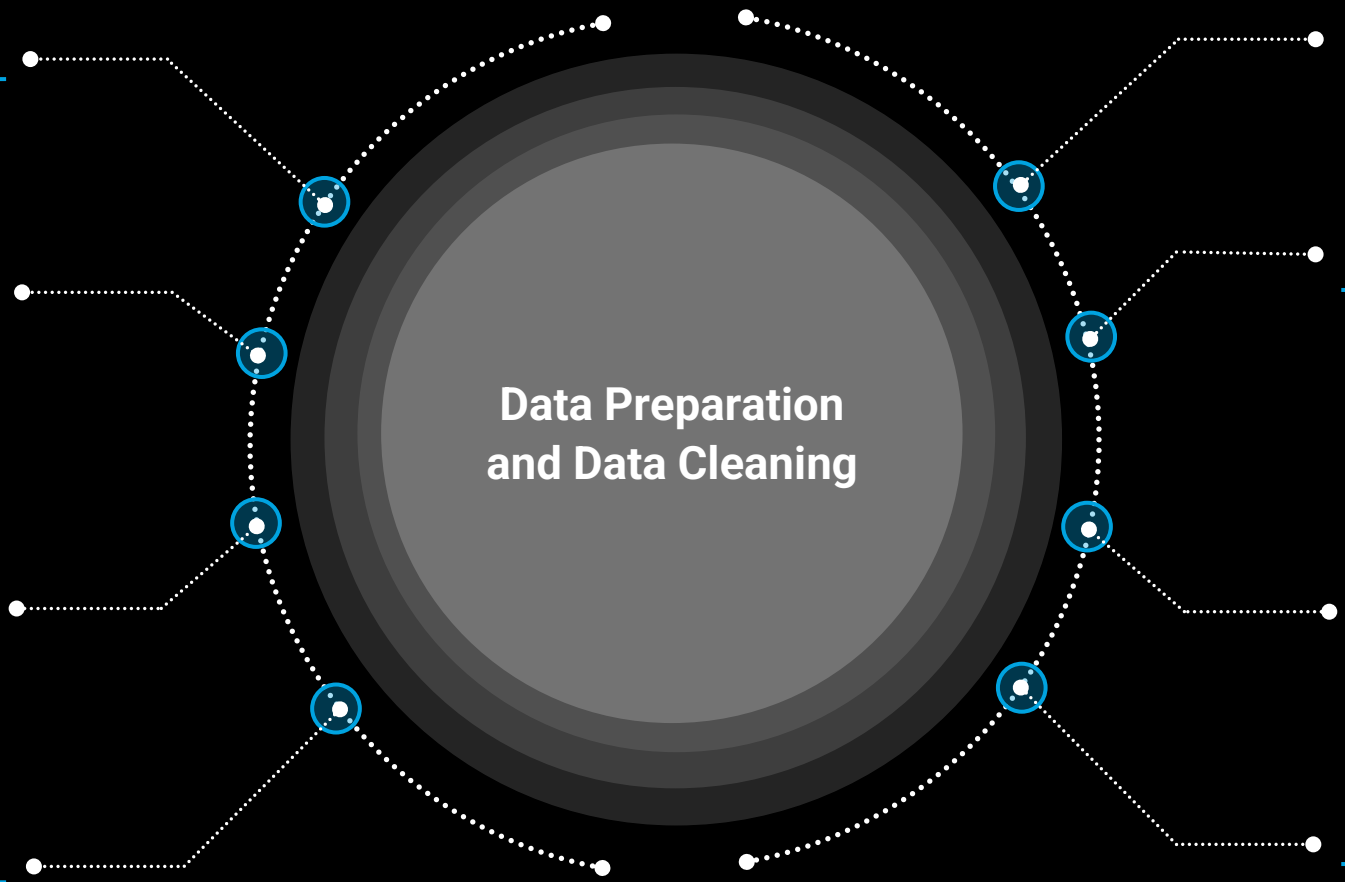
- After converting all column values to lower case, we see that there are around 99283 (6.33%) rows that are duplicates. We went De-Duplication ahead and dropped them

Treating Null values and Whitespaces

- Initially there weren't any NULL values in the dataframe. However, there were quite a few Whitespaces present in some of the columns in the dataframe
- We first converted these whitespaces to NaNs and the dropped these values

Selecting One Year Data

- Selecting1 Year Data from July, 2015 – June, 2016. In the process, 592 records were dropped



Dropping Insignificant columns

- Dropping Columns with Single Unique Value (as it doesn't add any information to the analysis)
- Dropping some of the 'Id' Columns which are insignificant to the analysis

Outlier Treatment

- Since we have already deleted some records on erroneous grounds, in order that we don't lose any further data, we chose not to delete outlier values
- For the variables - 'SLA', 'deliverybdays', 'deliverybdays', 'gmv', 'product_mrp', 'list_price' where outliers are present, we CAPPED the values above 99 percentile to the value corresponding to 99 percentile
- Thus the outliers couldn't affect the predictive model while at the same time there was enough data to build a generalizable

Converting Categorical Attributes to Numerical Form

- Binary encoding for categorical variable with 2 levels
- One Hot Encoding for categorical variable with multiple levels by creating dummy variables

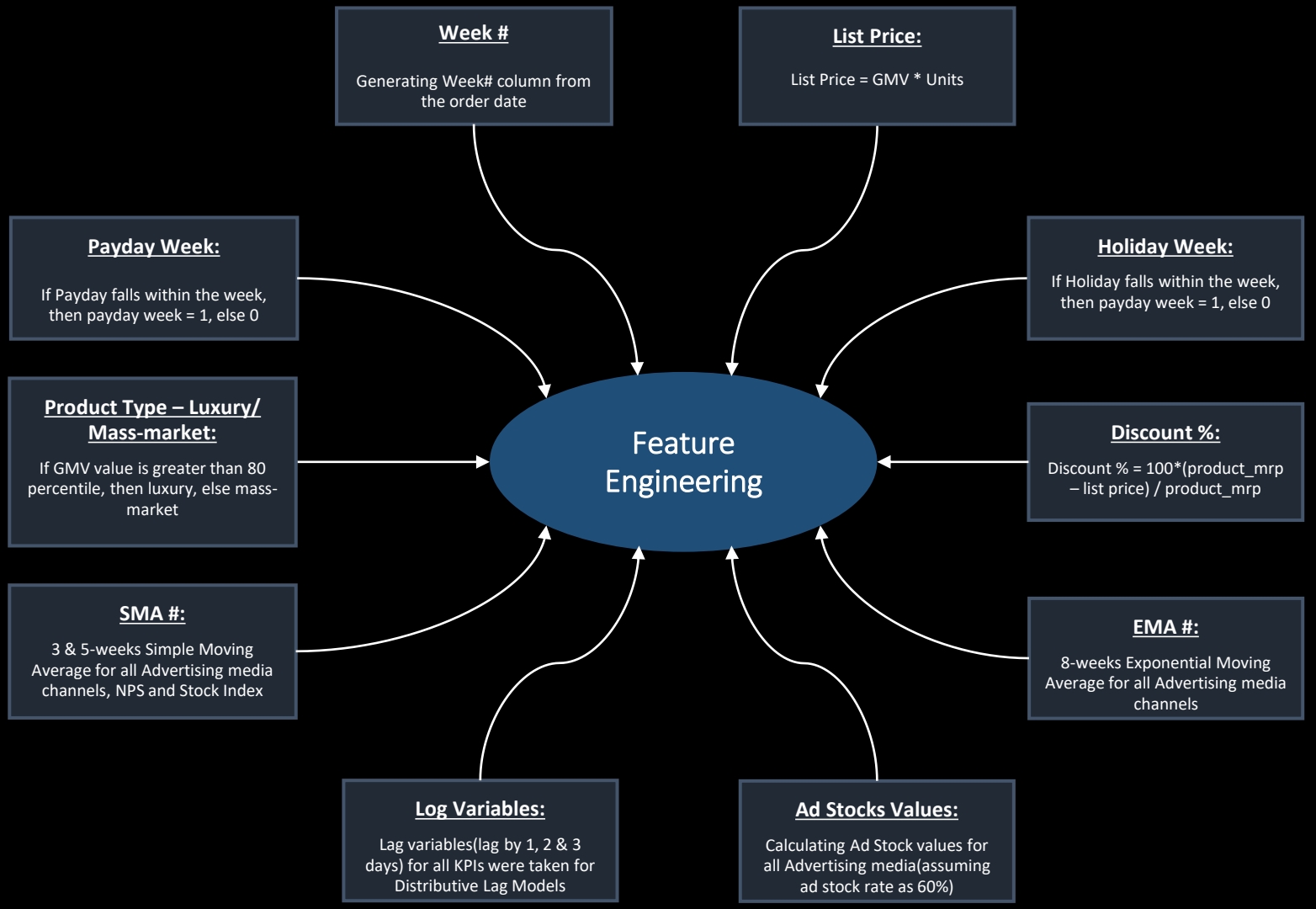
Additional Data Preparation for Model Building

- Merging Order dataset with all other secondary dataframes
- Extracting 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
- Scaling and dividing the master dataframes into train and test datasets for all 3 product subcategories

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

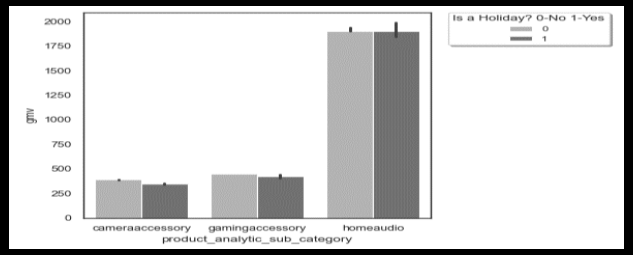
Feature Engineering: Creation of New KPI's

Marketing KPIs are measurable metrics your company can track to gauge performance over channel-specific marketing activities. It proves the Return on Investment (ROI) about the effectiveness of marketing campaigns and strategies so the business can double down on what's working and adjust for what's not.

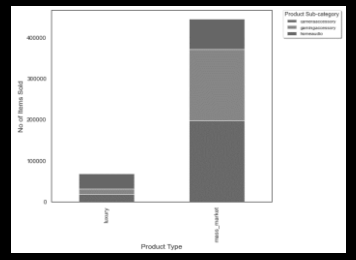


Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

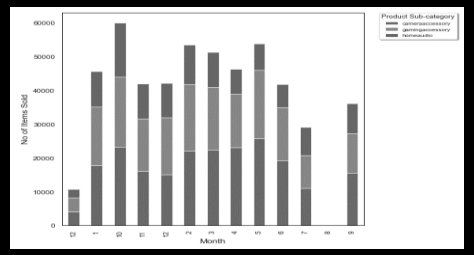
EDA Analysis & Visualization:



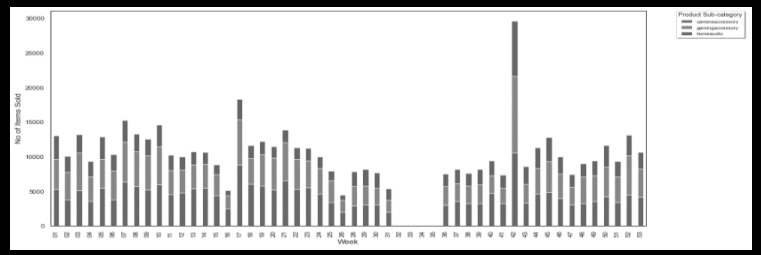
Average Revenue from Holiday/Non-holiday days for the 3 product subcategories



No of items(Luxury/Mass-market) sold per 3 product subcategories



Total items sold per 3 product subcategories per Month



Total items sold per 3 product subcategories per Week

Top 10 Product Verticals which brought the Maximum Revenue for 3 product sub-categories

	product_analytic_sub_category	product_analytic_vertical	gmv	product_in_category
0	homeaudio	homeaudiospeaker	1.873206e+08	homeaudiospeaker in homeaudio
1	cameraaccessory	lens	1.085308e+08	lens in cameraaccessory
2	gamingaccessory	gamepad	6.187440e+07	gamepad in gamingaccessory
3	gamingaccessory	gamingheadset	3.199049e+07	gamingheadset in gamingaccessory
4	cameraaccessory	binoculars	2.658427e+07	binoculars in cameraaccessory
5	gamingaccessory	gamingmouse	2.632837e+07	gamingmouse in gamingaccessory
6	cameraaccessory	camerabattery	2.356174e+07	camerabattery in cameraaccessory
7	cameraaccessory	camerabag	2.249499e+07	camerabag in cameraaccessory
8	cameraaccessory	flash	2.228150e+07	flash in cameraaccessory
9	homeaudio	fmradio	2.222170e+07	fmradio in homeaudio

Home Audio Speaker under Home Audio segment brought the largest revenue followed by Camera Lens under Camera Accessory & Gamepad under Gaming Accessory

Top 10 Product Verticals with most no of sales for 3 product sub-categories

	product_analytic_sub_category	product_analytic_vertical	units	product_in_category
0	homeaudio	homeaudiospeaker	76581	homeaudiospeaker in homeaudio
1	gamingaccessory	gamingheadset	59928	gamingheadset in gamingaccessory
2	gamingaccessory	gamepad	52437	gamepad in gamingaccessory
3	cameraaccessory	flash	47808	flash in cameraaccessory
4	gamingaccessory	gamingmouse	35470	gamingmouse in gamingaccessory
5	cameraaccessory	camerabattery	35107	camerabattery in cameraaccessory
6	cameraaccessory	lens	32350	lens in cameraaccessory
7	cameraaccessory	cameratripod	31220	cameratripod in cameraaccessory
8	homeaudio	fmradio	24681	fmradio in homeaudio
9	cameraaccessory	camerabag	15842	camerabag in cameraaccessory

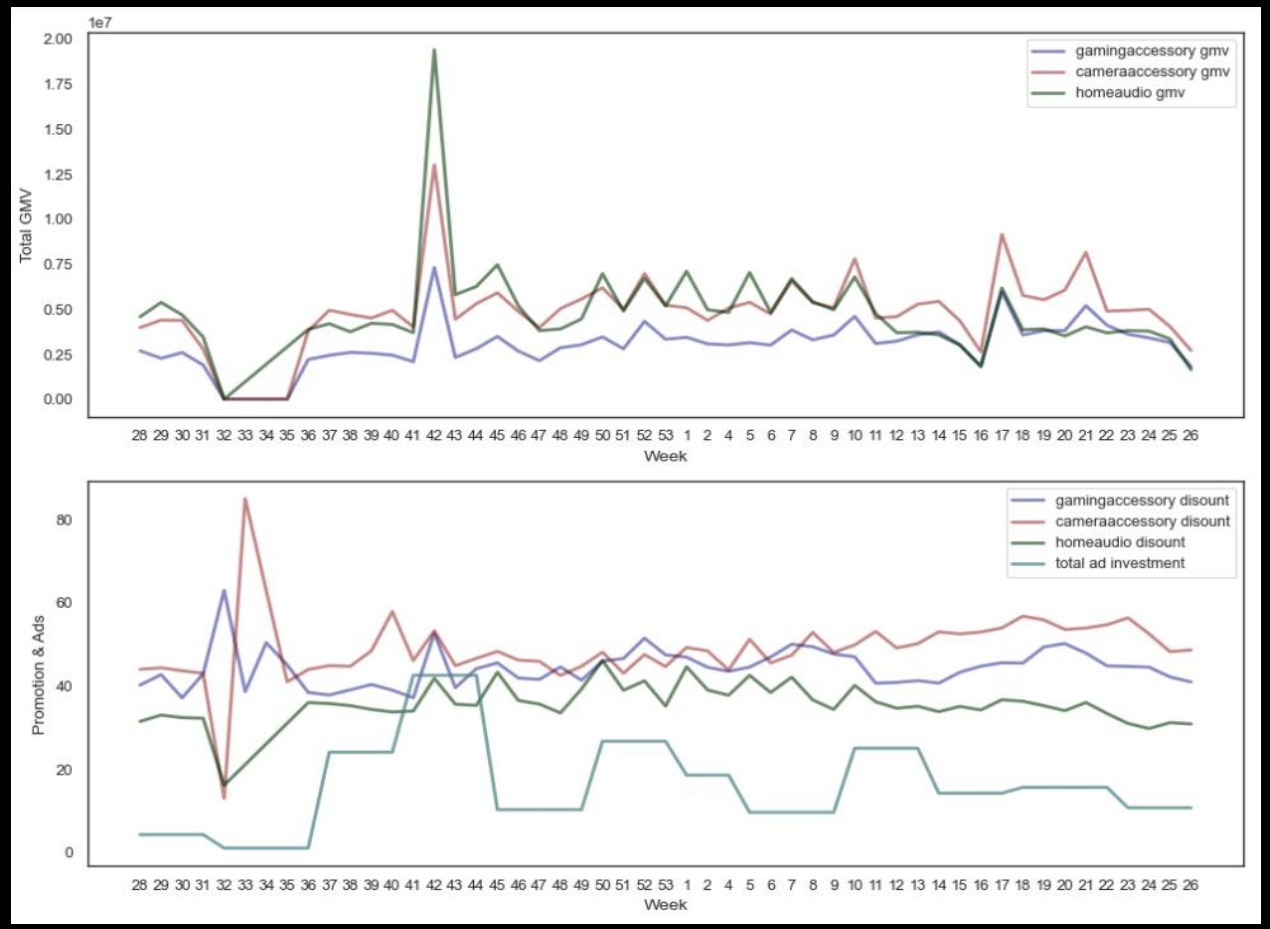
Home Audio Speaker under Home Audio segment had the most no of sales followed by Gaming Headset & Gamepad under Gaming Accessory

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

EDA Analysis & Visualization (contd.):

Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. With modern analytics platforms, these visualizations can go far beyond line graphs.

Comparison of Trends of Revenue Discount% & Total Media Investment over the weeks



The following observations were noted from the above time series plots:

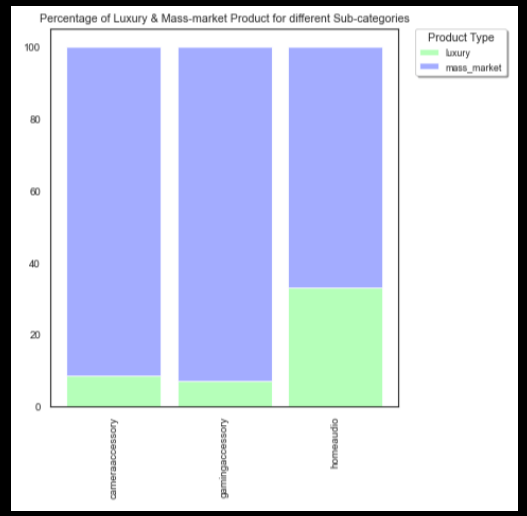
- For the week #42 (during Thanks giving), all the graphs show a steep rise. Revenue increased because of both higher discount% and increased Ad Investment.
- For the week 32(August), Revenue generated was the lowest from all 3 product subcategories. This can be observed as a direct relation to minimum amount of total investment in Ads. Discount was also lowest for all products apart from camera accessories. Post this dip in revenue, discount% was increased to bring about higher sales. This increase in Discount% was observed most in the case of gaming accessories. However, barring home audio products, the revenue from other products was seen to be constant for the next 3 weeks after which, the revenue started to pick up.
- In general the average discount % offered for home audio products is lesser compared to that of the other product subcategories.

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

EDA Analysis & Visualization:

Finding the percentage of Luxury & Mass-market Products from 3 sub-categories

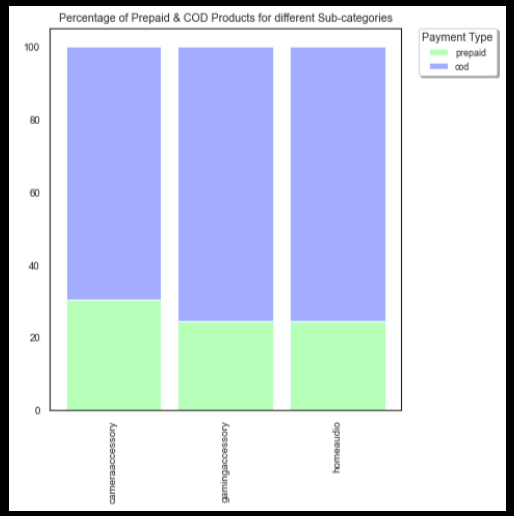
	product_analytic_sub_category	luxury	mass_market
0	cameraaccessory	18423	197478
1	gamingaccessory	13007	172869
2	homeaudio	36747	74314



Percentage of luxury products under Home Audio is much more compared to the other sub categories.

Finding the percentage of COD & Prepaid Products from 3 sub-categories

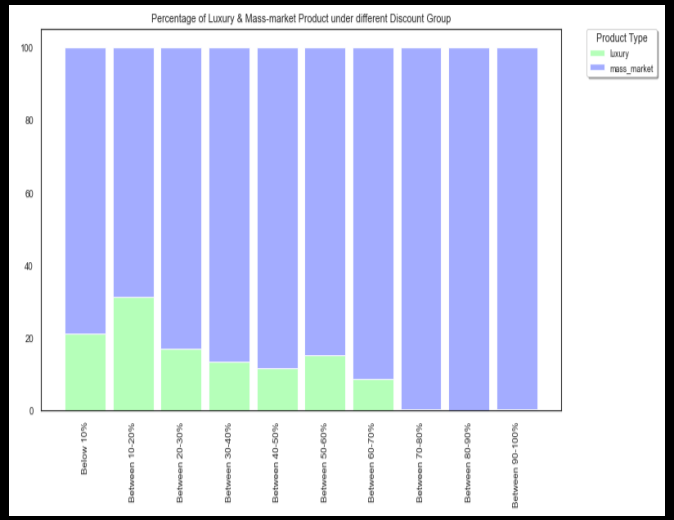
	product_analytic_sub_category	prepaid	cod
0	cameraaccessory	65462	150439
1	gamingaccessory	45273	140603
2	homeaudio	27123	83938



Percentage of prepaid payments under Camera Accessory was observed to be slightly more compared to that of the other sub categories.

Finding the percentage of Luxury and Mass_market Products under different Discount groups

	Discount Bins	luxury	mass_market
0	Below 10%	9110	34093
1	Between 10-20%	14604	31972
2	Between 20-30%	10077	48907
3	Between 30-40%	8593	56255
4	Between 40-50%	7477	57617
5	Between 50-60%	13335	73808
6	Between 60-70%	4847	51506
7	Between 70-80%	68	41789
8	Between 80-90%	48	39781
9	Between 90-100%	18	8933



Percentage of luxury products were given a discount between 10-20%.

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Description of Model Building

The primary objective of the case study being Revenue prediction and determination of important KPIs that influence the revenue growth, we have to build the following Linear Regression models:

1

Additive Model

Linear model is used to capture the current effect of several KPIs. This model assumes an additive relationship between the different KPIs. Hence their impacts are also additive towards the dependent Y variable. The equation can be represented as:

$Y = \alpha + \beta1At + \beta2Pt + \beta3Dt + \beta4Qt + \beta5Tt + \epsilon$

2

Multiplicative Model

Multiplicative model is used when there are interactions between the KPIs. To fit a multiplicative model, take logarithms of the data(on both sides of the model), then analyze the log data as before.

$Y = e^{\alpha} .X1^{\beta1} . X2^{\beta2} . X3^{\beta3} . X4^{\beta4} . X5^{\beta5} + \epsilon$ and $\ln Y = \alpha + \beta1\ln(X1) + \beta2\ln(X2) + \beta3\ln(X3) + \beta4\ln(X4) + \beta5\ln(X5) + \epsilon'$

3

Koyck Model

Koyck model is used to capture the carry-over effect of different KPIs, ie.to model the current revenue figures based on the past figures of the KPIs. The Koyck tells us that the current revenue generated is not just influenced by the different independent attributes, but also because of the revenue generated over the last periods.

$Yt = \alpha + \beta1X1 + \beta2X2 + \beta3X3 + \beta4X4 + \beta5X5 + \epsilon$ and $Yt = \alpha + \mu Yt\text{-}1 + \beta1X1 + \beta2X2 + \beta3X3 + \beta4X4 + \beta5X5 + \epsilon$

4

Distributive Lag Model (Additive)

This is a more generalizable distributed lag model that captures the carry-over effect of all the variables:

$Yt = \alpha + \mu1Yt\text{-}1 + \mu2Yt\text{-}2 + \mu3Yt\text{-}3 + ... \bullet + \beta1X1t + \beta1X1t\text{-}1 + \beta1X1t\text{-}2 + ... \bullet + \beta2X2t + \beta2X2t\text{-}1 + \beta2X2t\text{-}2 + ... \bullet + \beta3X3t + \beta3X3t\text{-}1 + \beta3X3t\text{-}2 + ... \bullet + \beta4X4t + \beta4X4t\text{-}1 + \beta4X4t\text{-}2 + ... \bullet + \beta5X5t + \beta5X5t\text{-}1 + \beta5X5t\text{-}2 + ... \bullet + \epsilon$

5

Distributive Lag Model (Multiplicative)

Distributive Lag Model(Multiplicative) will help us capture the interactions between current and carry over effects of the KPIs.

$Yt = \alpha + \mu1\ln(Yt\text{-}1) + \mu2\ln(Yt\text{-}2) + \mu3\ln(Yt\text{-}3) + ... \bullet + \beta1\ln(X1t) + \beta1\ln(X1t\text{-}1) + \beta1\ln(X1t\text{-}2) + ... \bullet + \beta2\ln(X2t) + \beta2\ln(X2t\text{-}1) + \beta2\ln(X2t\text{-}2) + ... \bullet + \beta3\ln(X3t) + \beta3\ln(X3t\text{-}1) + \beta3\ln(X3t\text{-}2) + ... \bullet + \beta4\ln(X4t) + \beta4\ln(X4t\text{-}1) + \beta4\ln(X4t\text{-}2) + ... \bullet + \beta5\ln(X5t) + \beta5\ln(X5t\text{-}1) + \beta5\ln(X5t\text{-}2) + ... \bullet + \epsilon'$

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Dashboard:

The following table contains the details of all models built, 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	

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Model Selection:

- The criteria of choosing the model is based on the accuracy parameters - R2 score & MSE score - and the business relevance of the important attributes chosen by the model.
- Also we tried to choose models with cross validation because even though the ones without, sometimes give us good scores, they are not very dependable & generalizable, owing to limited dataset.
- By referring to the model dashboard, we finalize the following models for the 3 mentioned product subcategories – Camera Accessory, Gaming Accessory and 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)

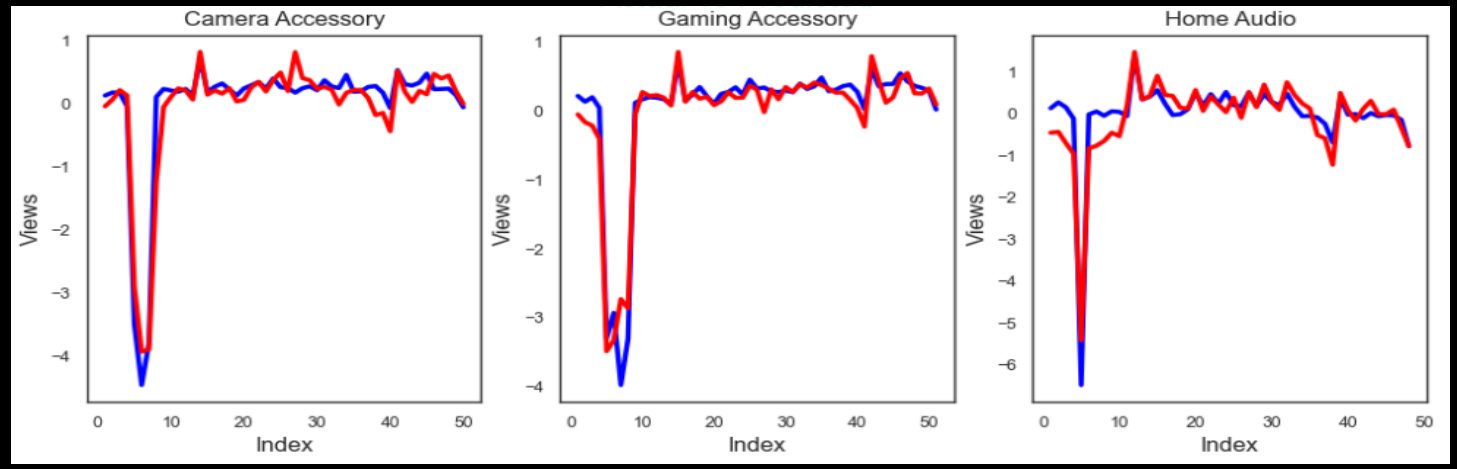
- We notice that all the 3 chosen models for the 3 sub-categories are Multiplicative models.
- This fact tells us that there exists some interaction between the KPIs for all the 3 model.
- These models tell us about the growth of revenue vs the interactive growth of the KPIs

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Model Validation:

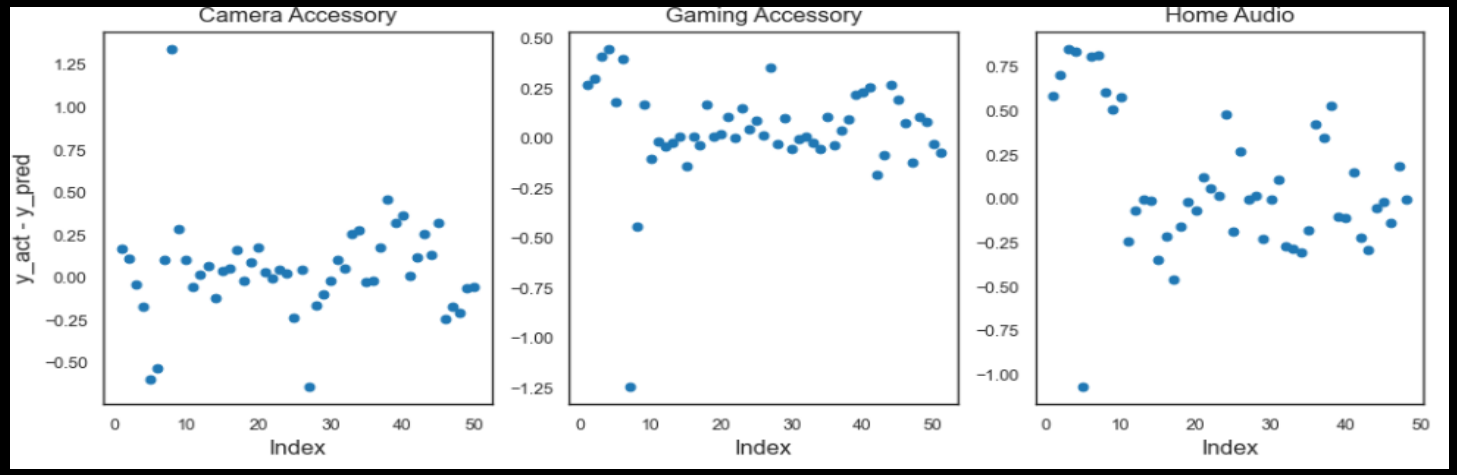
Actual vs. Predicted

Plotting the distribution of the error terms. The error terms follow a normal distribution with mean at 0 barring a few outlier values.



Error Terms

Drawing a scatter plot of the Error Terms to check the spread to ensure that the error terms have constant variance (homoscedasticity). The variance doesn't increase or decrease or follow a pattern as the error values change.

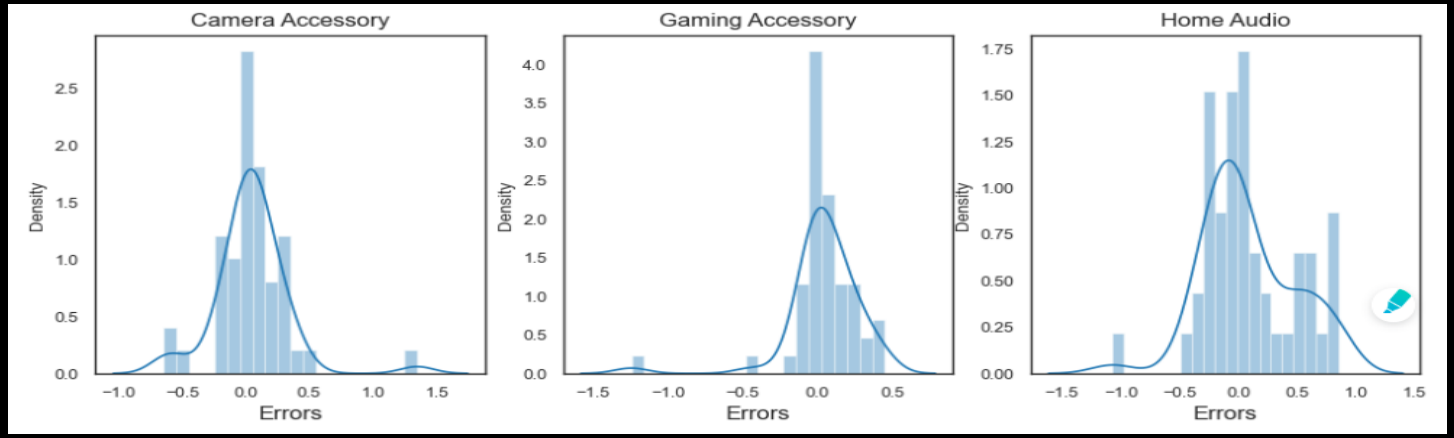


Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Model Validation (contd.):

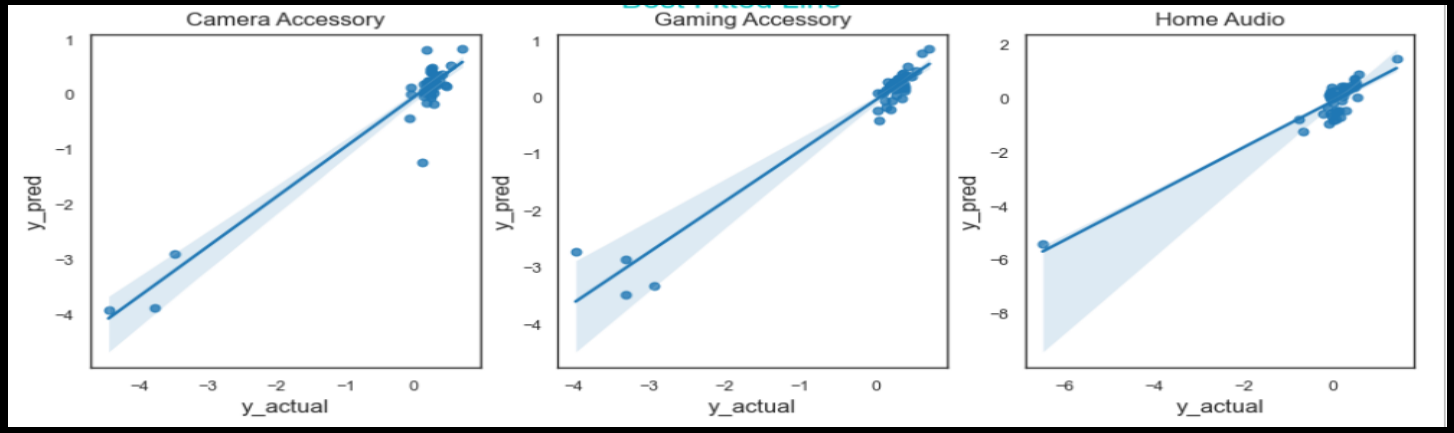
Distribution of Error Terms

Plotting the distribution of the error terms. The error terms follow a normal distribution with mean at 0 barring a few outlier values.



Best Fitted Line

Plotting a scatter plot with actual and predicted price values from the dataset to check the spread and drawing the best fitted line through it.



Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Equation of Best Fitted Line

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 the revenue can grow with a unit growth in any of these independent KPIs with all other KPIs held constant

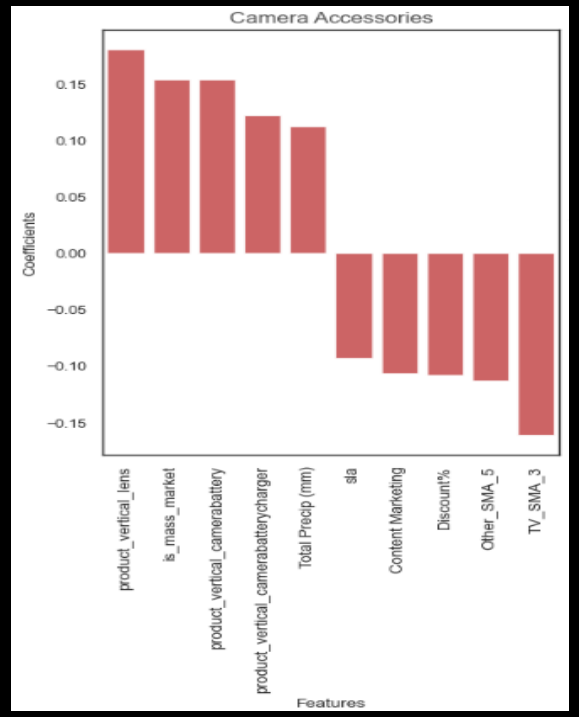
Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Recommendations:

Top 5 features that affect Each of the 3 Product Sub-categories(both positively and adversely) as per our Chosen Models'

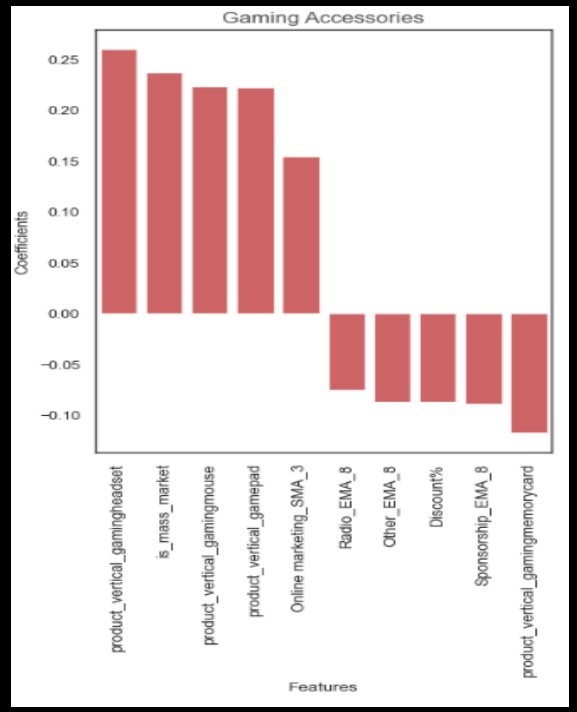
Camera Accessory

- Company should promote `Lens`, `Camera Batteries` & `Camera Battery Chargers` as they fetch the highest revenue.
- Advertisement spends on TV has a positive impact on revenue. One unit of TV spend can boost the revenue by 0.105 units. Content Marketing spends on the other hand impacts negatively.
- `Mass-market` products are better contributors to the increased revenue in comparison to the Luxury products.
- Higher percentage of Discounts in general given for this sub category works adversely towards bringing down the revenue.



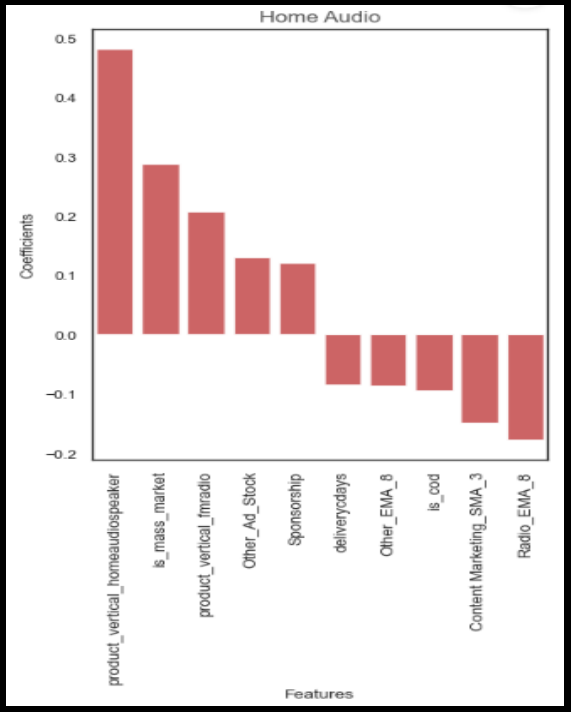
Gaming Accessory

- Company should promote `Gaming Headset`, `Gaming Mouse` & `Gamepad` as they fetch the highest revenue. On the contrary, `Gaming Memory Cards` results in loss.
- Advertisement spends on Online Marketing, Radio & Others have a positive cumulative impact on revenue. Sponsorship spends on the other hand has a negative cumulative effect.
- `Mass-market` products are better contributors to the increased revenue in comparison to the Luxury products.
- Higher percentage of Discounts in general given for this sub category works adversely towards bringing down the revenue



Home Audio

- Company should promote `Home Audio Speakers` & `FM Radios as they fetch the highest revenue.
- `Mass-market` products are better contributors to the increased revenue in comparison to the Luxury products.
- Radio Ad Stock (carry over effect of Radio Advertisement) spends helps to boost the revenue to a significant extent.
- Advertisement spends on Sponsorship has a positive impact on revenue. Content Marketing spends on the other hand impacts negatively.
- COD payments in general for this sub category are bad in bringing down the revenue.



Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Recommendations (contd.):

General Recommendations:

- Most of the sales take place when Discount% is between 50-60%. However, that doesn't necessarily help in boosting the revenue. EDA shows that an average discount% between 10-20% is the most profitable for the company specially among luxury items.
- In general most of the Home Audio items sold are luxury items and hence, customers prefer to use COD instead of paying upfront.
- During festive time(e.g. Thanksgiving) more investment is made on Advertisement and good promotional offers were rolled out. This usually boosts the revenue. However just providing discounts without properly advertising for it on several media channels doesn't help. We have seen that for the weeks 32 - 35(August), revenue generated was the lowest from all 3 product subcategories even though median discount% was raised after the initial drought. In fact, this dip in revenue can be observed as a direct relation to minimum amount of total investment in Ads during the given timeframe.

- Camera Accessories – Recommendation based on elasticity of KPI's
  - Focus on ads spend on Digital Means
  - Weekly sales have better impact
  - Procurement of products should be taken seriously

- Game Accessories – Recommendation based on elasticity of KPI's
  - Focus on digital ad stocks
  - As SEM investments are least effective, so further do not add revenue
  - Procurement of products should be taken seriously

- Home Audio – Recommendation based on elasticity of KPI's
  - Need more focus on cash on delivery customers
  - More advertisements through Television channels
  - Launch marketing schemes towards colder regions, it might help the sales

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Challenges Faced:

- Deciding on number of derived KPI's needed which are important as well as logical.
- Arriving at a base dataset because many iterations were performed as the model results were leading to perfectly fitting model.
- To solve this problem, correlation concept was used to remove correlated variables
- Selection of significant variables which are finally decided by correlation matrix and non zero variance
- Identifying the functions/ packages to perform ad stock and creating Lag Variables
- Limited availability of the sample implementation of Kyock, Lag and Multiplicative models in Python
- Decision on removal of variables from analysis

Codes: https://github.com/veer2701/Market-Mix-Model/tree/main

Thank You!