Capstone Project: Market Mix Model ElecKart Final Report for Submission

To simulate the influence of various factors on sales figures.

- Ojas Balsotra

Primary Business Goal

To develop a market mix model tailored for ElecKart, an e-commerce company located in Ontario, Canada, focusing on three product sub-categories: Camera Accessory, Gaming Accessory, and Home Audio. The primary objective is to assess the real-world effects of various marketing variables over a one-year period from July 2015 to June 2016. The ultimate goal is to provide recommendations for the most efficient allocation of marketing budgets across different levers for the upcoming year.

The objective can be subdivided into the following sub-goals:

1 Analysis of Performance Drivers

Identifying Key Performance Indicators (KPIs) that significantly influence top-line performance.

2 Impact Assessment on Marketing ROI

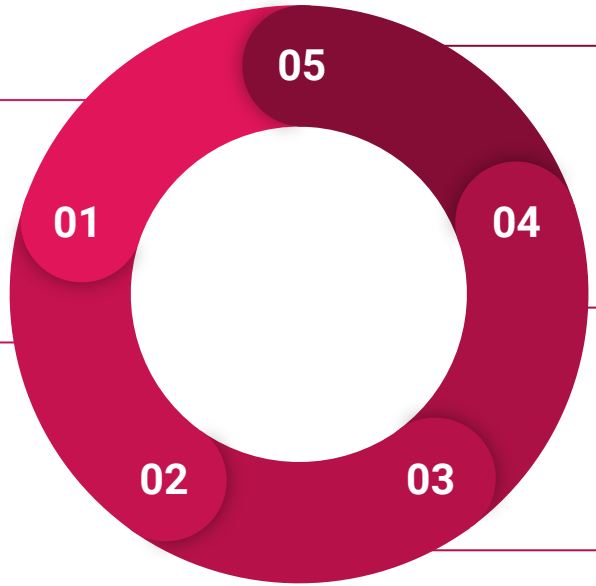
Measuring the exact influence of each commercial lever on revenue to assess their efficiency and effectiveness.

3 Marketing Budget Optimization

Identifying the optimal strategy for allocating the marketing budget to achieve the highest possible results.

Problem-Solving Methodology

The project approach has been structured to adhere to the CRISP-DM Framework, which consists of several sequential stages, as illustrated below:



Understanding the Business Data

Gaining insights into the business context and data sources to define project objectives.

Data Preparation & Feature Engineering

Cleaning, transforming, and enriching data to create suitable features for modeling.

Model Evaluation, Selection & Deployment.

Assessing model performance, selecting the best model, and deploying it for practical use.

Preparing Models for preparation

Building and fine-tuning predictive models to address specific business questions.

EDA & Visualization

Visualizing data patterns to uncover insights and trends.

Understanding Business Data

We had access to several data files for budget optimization analysis, including:

1. Main Consumer File: Contains daily order details.

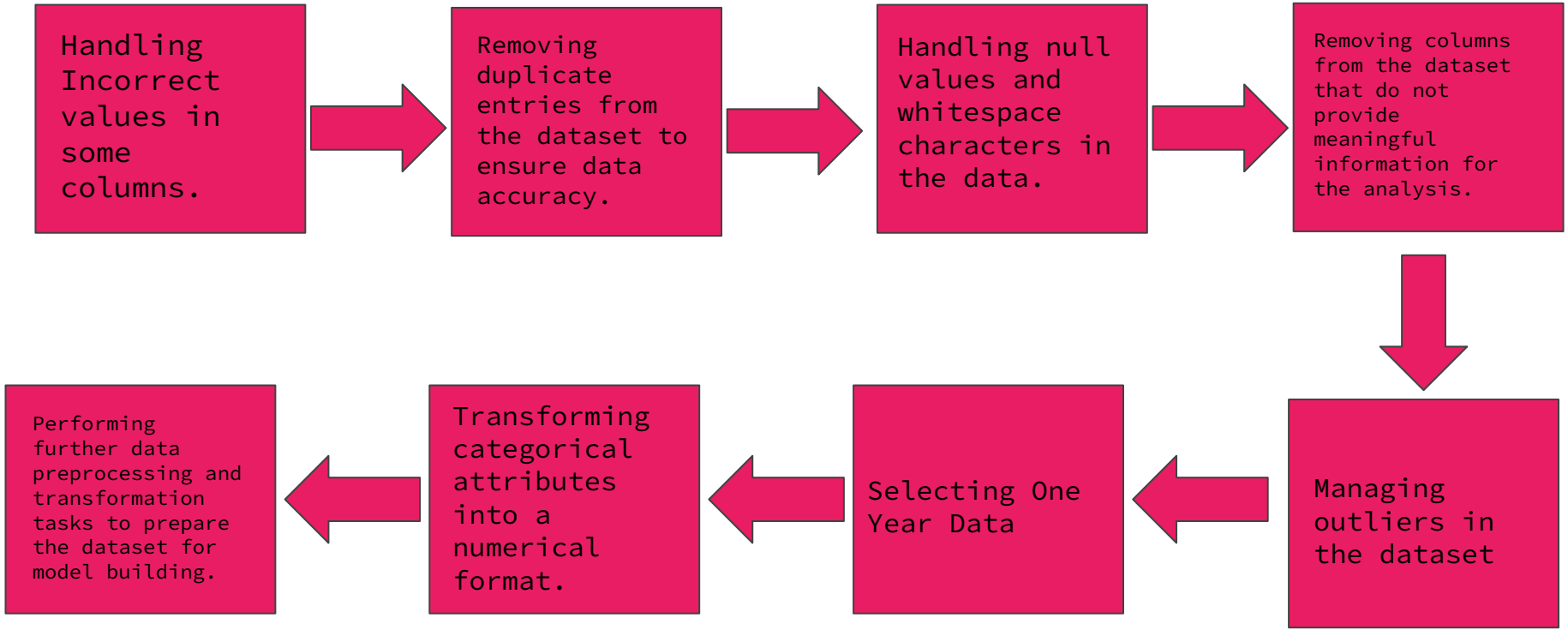
2. Media Investment File: Records advertising investments in various mediums over the past year.

3. Sale Calendar File: Displays dates of promotional offers from the previous year.

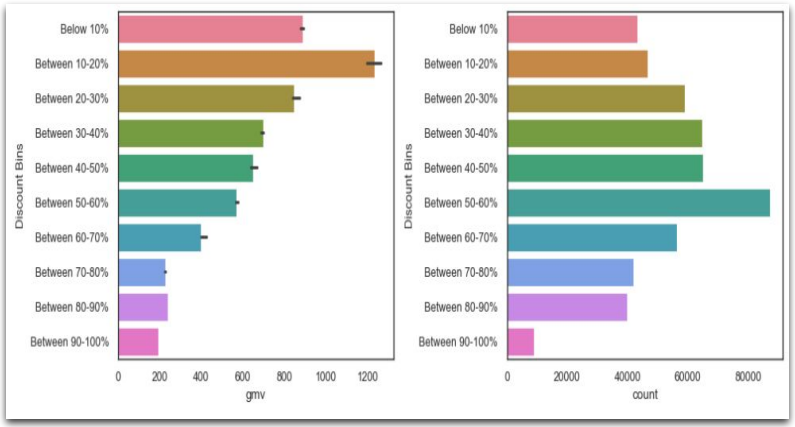
4. NPS File: Includes net promotion scores and company stock values from the last year.

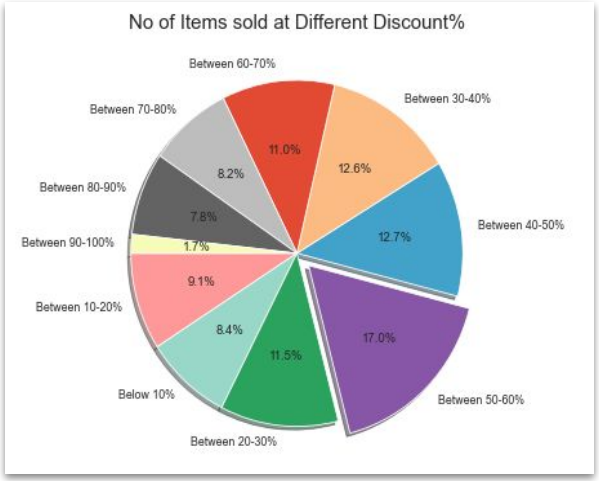
5. Weather File: Contains detailed weather reports for Ontario, Canada, from the previous year.

Data Preparation & CleanUp



Visualization and Insight into Data





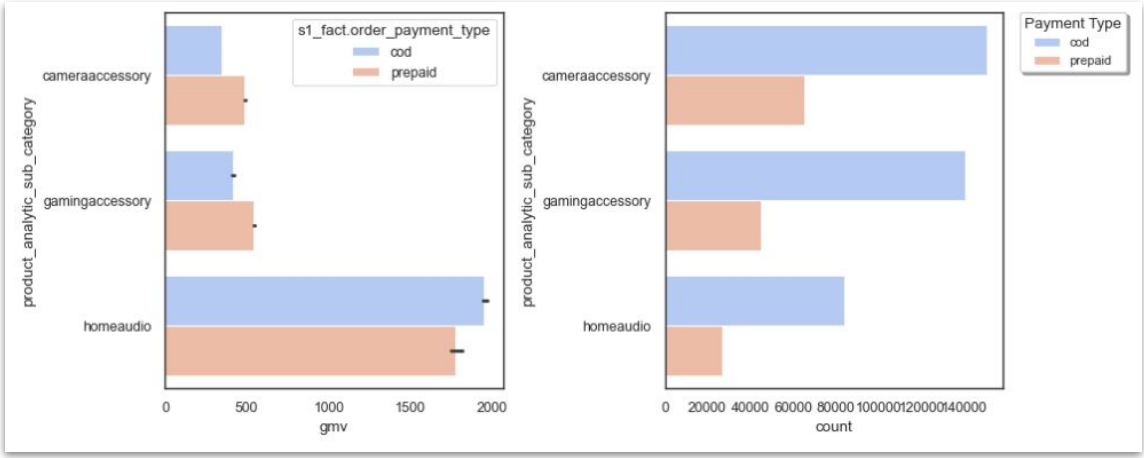
Analysing the Sales amount and Revenue based on Discount %.

Median Revenue is maximum when Average discount% is between 10-20%. But beyond that, average revenue slowly starts to decline.

The sales on the other hand shows a steady increase with increase in Discount percentage till it peaks at 50-60% after which it starts to fall again.

This shows that at higher discount, although the sales are good, the revenue collapses signifying a loss for the company. An average discount of 10-20% is the most profitable for the company.

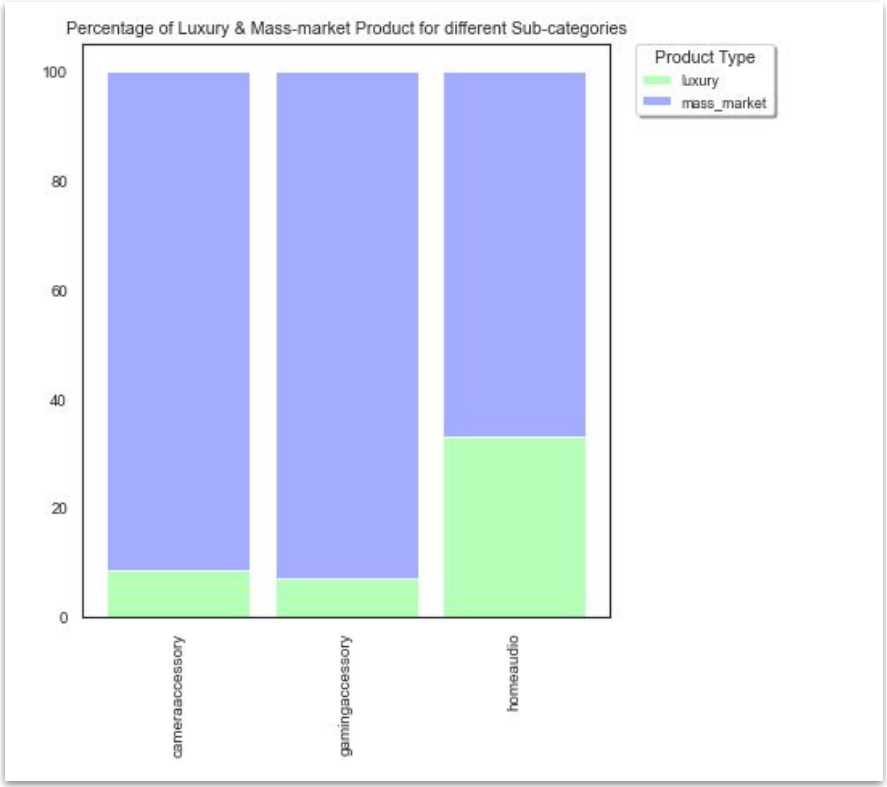
Visualization and Insight into Data



Analysing the Sales amount and Revenue based on Payment Type.

Except for Home Audio Products, for the other 2 product sub categories, we observe that the median Revenue from Prepaid orders is more than that from COD products even though the no of products sold is way higher in case of COD products for all categories.

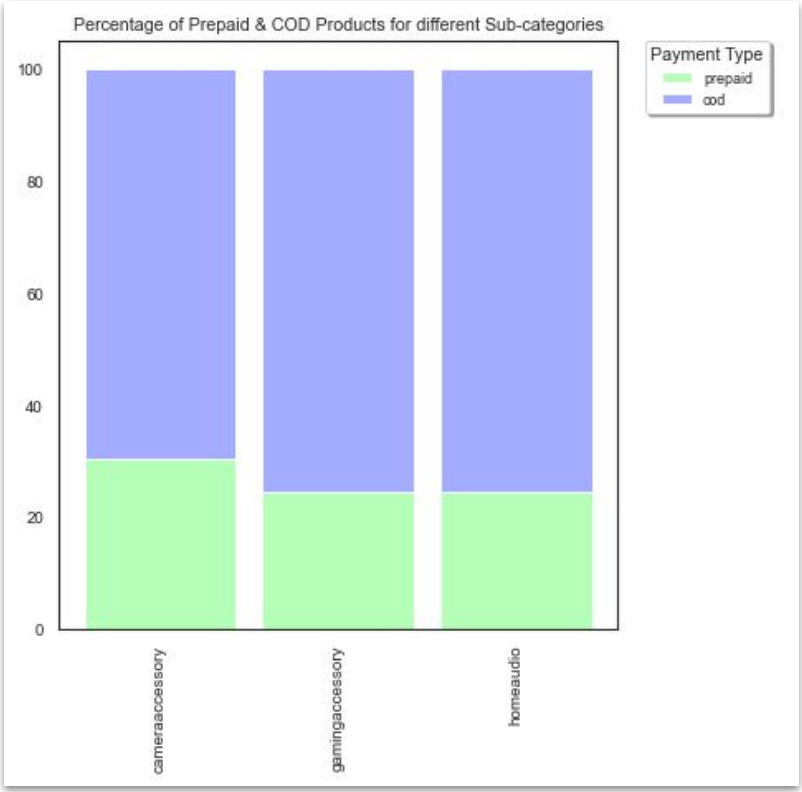
Visualization and Insight into Data



Analysing the Luxury & Mass-Market Product for different SubCategories.

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

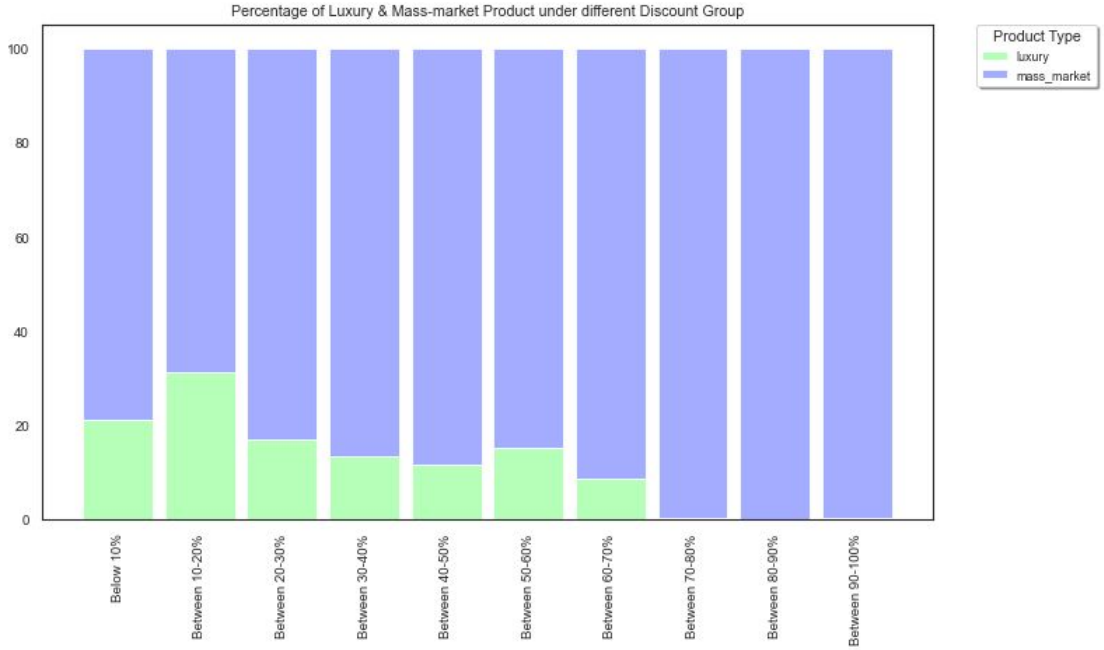
Visualization and Insight into Data



Analysing the % of Prepaid & Cod Products for different SubCategories.

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

Visualization and Insight into Data



Percentage of Luxury & Mass-market Product under different Discount Group

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

Description of Model Built

With the primary objective of this case study being revenue prediction and the identification of key performance indicators (KPIs) that have an impact on revenue growth, we have developed the following Linear Regression models:

Additive - With the primary objective of this case study being revenue prediction and the determination of important KPIs that influence revenue growth, we have built the Additive Linear Regression models.

The equation can be represented as:

$$Y = \alpha + \beta 1At + \beta 2Pt + \beta 3Dt + \beta 4Qt + \beta 5Tt + \epsilon$$

Multiplicative - With the primary objective of this case study being revenue prediction and the determination of important KPIs that influence revenue growth, we have built both Multiplicative Linear Regression models

The equation can be represented as:

$$lnY = \alpha + \beta 1ln(X1) + \beta 2ln(X2) + \beta 3ln(X3) + \beta 4ln(X4) + \beta 5ln(X5) + \epsilon'$$

Description of Model Built

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 equation can be represented as:

$$Yt = \alpha + \mu Yt-1 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \beta 4X4 + \beta 5X5 + \epsilon$$

Distributive Lag Model (Additive) - In the distributed lag model, not only is the dependent variable entered in its lagged version, but the independent variables are as well. This is a more generalizable model and captures the carry-over effect of all the variables:

The equation can be represented as:

$$Yt = \alpha + \mu 1Yt-1 + \mu 2Yt-2 + \mu 3Yt-3 + ... \ \beta 1X1t + \beta 1X1t-1 + ... \bullet + \beta 2X2t + \beta 2X2t-1 + ... + \ \epsilon$$

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

The equation can be represented as:

$$Yt = \alpha + \mu 1\ln(Yt-1) + \mu 2\ln(Yt-2) + \mu 3\ln(Yt-3) + ... + \beta 1\ln(X1t) + \beta 1\ln(X1t-1) + \beta 1\ln(X1t-2) + ... + \epsilon'$$

The following table contains the details of top 5 KPIs returned by GMV(Revenue) :-

The 5 most important features affecting GMV(Revenue) for cameraaccessory are:

Features	Coefficients
is_mass_market	0.061
product_vertical_lens	0.060
product_vertical_cameraaccessory	0.060
product_vertical_camerabattery	0.059
product_vertical_cameratripod	0.059

The 5 most important features affecting GMV(Revenue) for gamingaccessory are:

Features	Coefficients
product_vertical_gamepad	0.088
product_vertical_gamingmouse	0.085
is_mass_market	0.082
product_vertical_gamingkeyboard	0.074
is_cod	0.072

The 5 most important features affecting GMV(Revenue) for homeaudio are:

Features	Coefficients
product_vertical_homeaudiospeaker	0.136
is_mass_market	0.133
product_vertical_fmradio	0.122
is_cod	0.112
product_vertical_voicerecorder	0.104

Model Selection

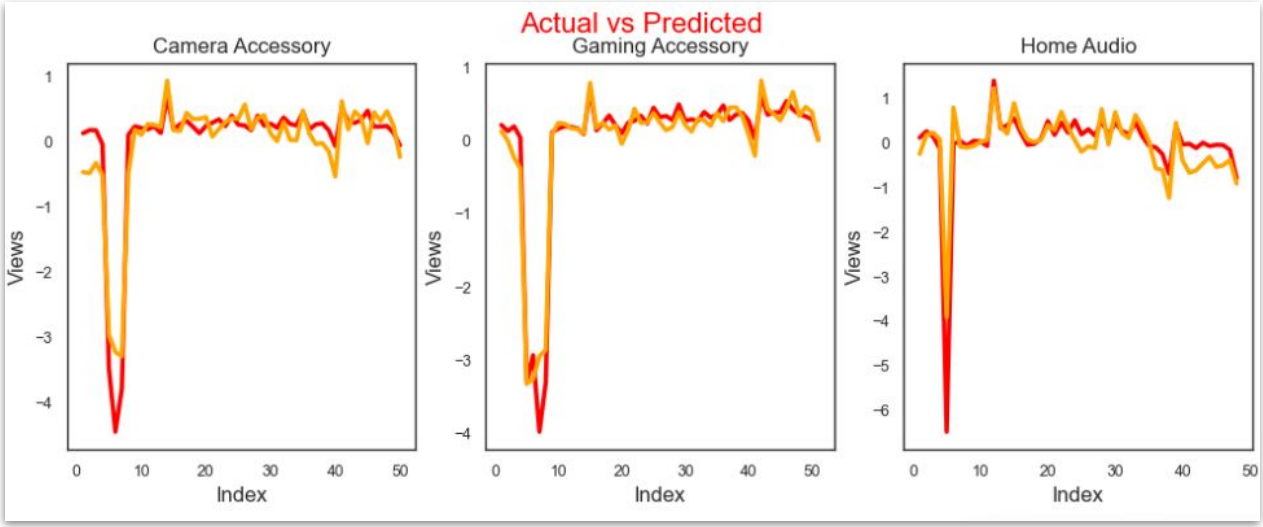
The criteria for selecting the models are primarily based on accuracy metrics, including the R2 score and MSE score, along with consideration of the business relevance of the key attributes identified by the model. Furthermore, we have prioritized models with cross-validation since those without it, although occasionally producing favorable scores, are not as reliable or generalizable due to the constraints of our limited dataset.

Upon review of the model dashboard chosen ultimately chosen the following models for the three specified 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)

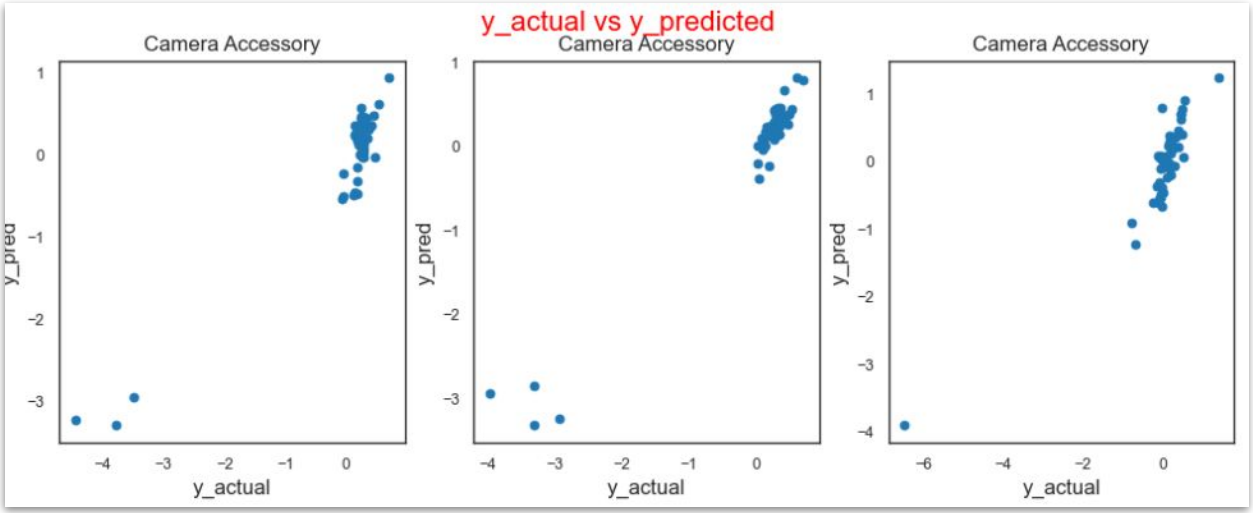
Model Validation

Plotting the actual and predicted price values from the dataset to check the likeness.



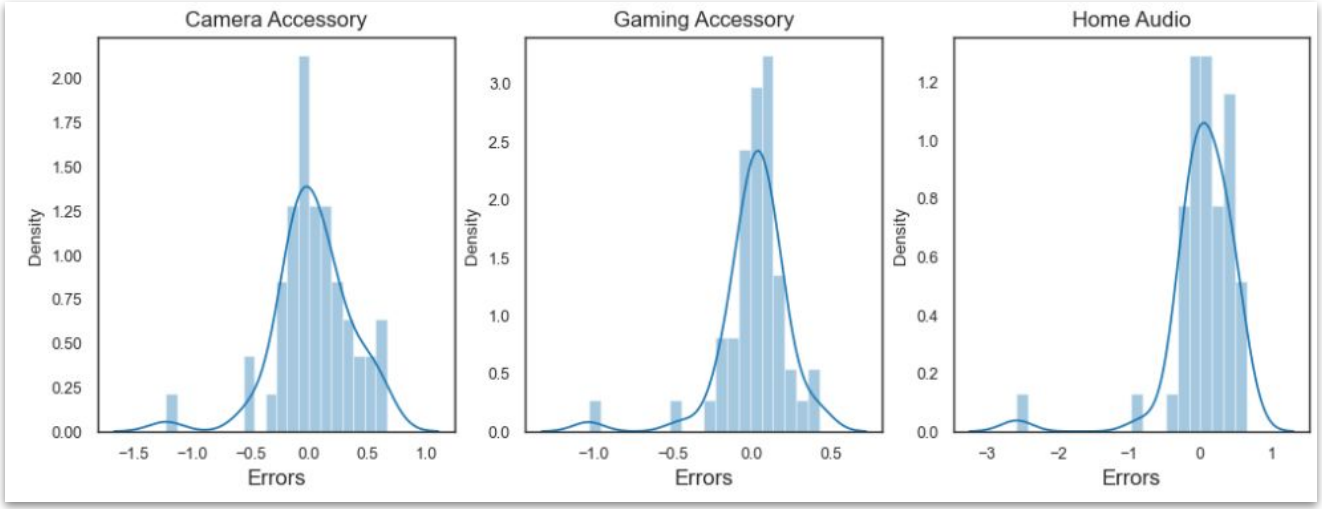
Model Validation

Drawing a scatter plot with actual and predicted price values from the dataset to check the spread.



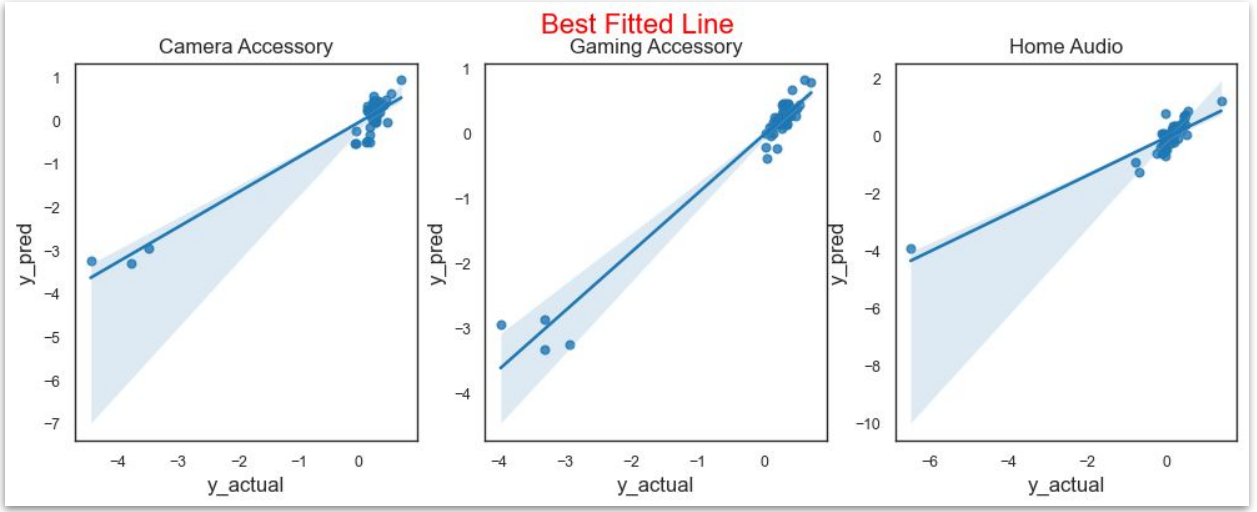
Model Validation

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



Model Validation

Creating a scatter plot that displays the actual price values against the predicted price values from the dataset. Additionally, fitting the best-fitting line through this plot to assess the spread and model performance.



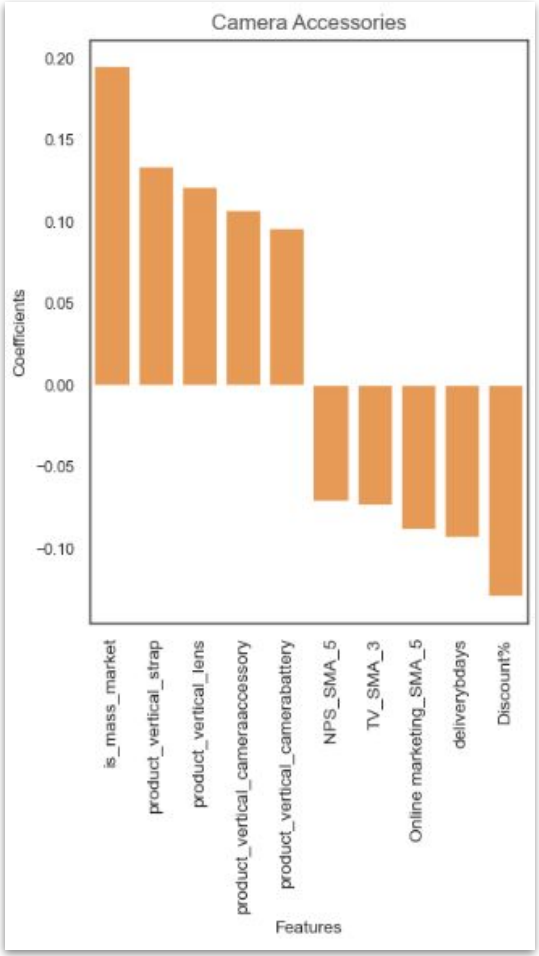
Recommendation for Camera Accessories.

- To maximize revenue, the company should prioritize promoting lenses, camera batteries, and camera battery chargers as they consistently generate the highest income.

- A Investing in television advertising has a positive impact on revenue, with each unit of TV spending resulting in a revenue increase of 0.105 units. In contrast, content marketing spending has a negative effect on revenue.

- Mass-market camera products tend to contribute more significantly to revenue growth compared to luxury items.

- Offering a higher percentage of discounts within this subcategory generally has a detrimental effect on revenue.



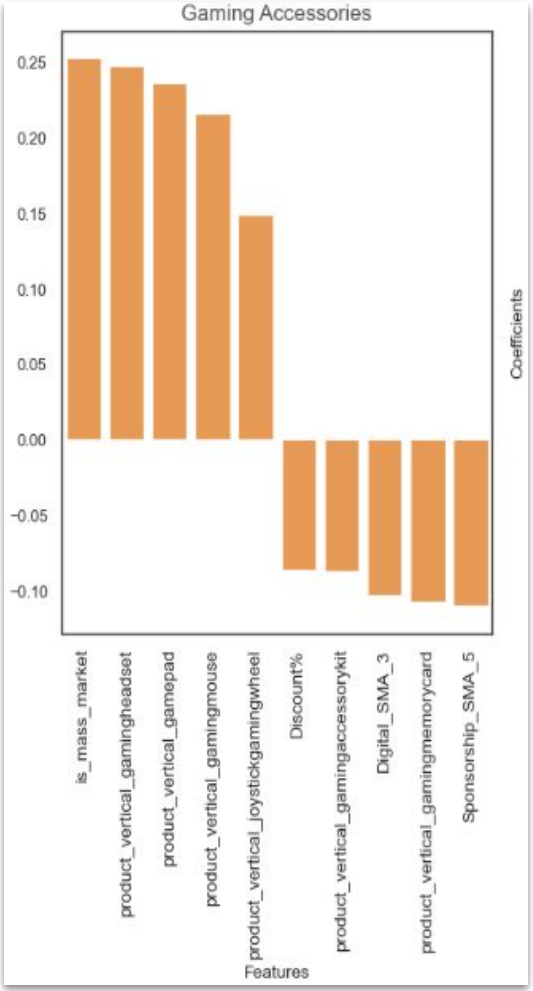
Recommendation for Gaming Accessories.

- To optimize revenue, the company should focus on promoting gaming headsets, gaming mice, and gamepads as they consistently yield the highest returns. Conversely, investing in gaming memory cards results in losses.

- Cumulative spending on online marketing, radio advertising, and other channels has a positive impact on revenue. However, sponsorship spending has a negative cumulative effect.

- Mass-market gaming products are more effective contributors to revenue growth compared to luxury products.

- Offering a higher percentage of discounts within this subcategory tends to have an adverse effect on revenue.



Recommendation for Home Audio.

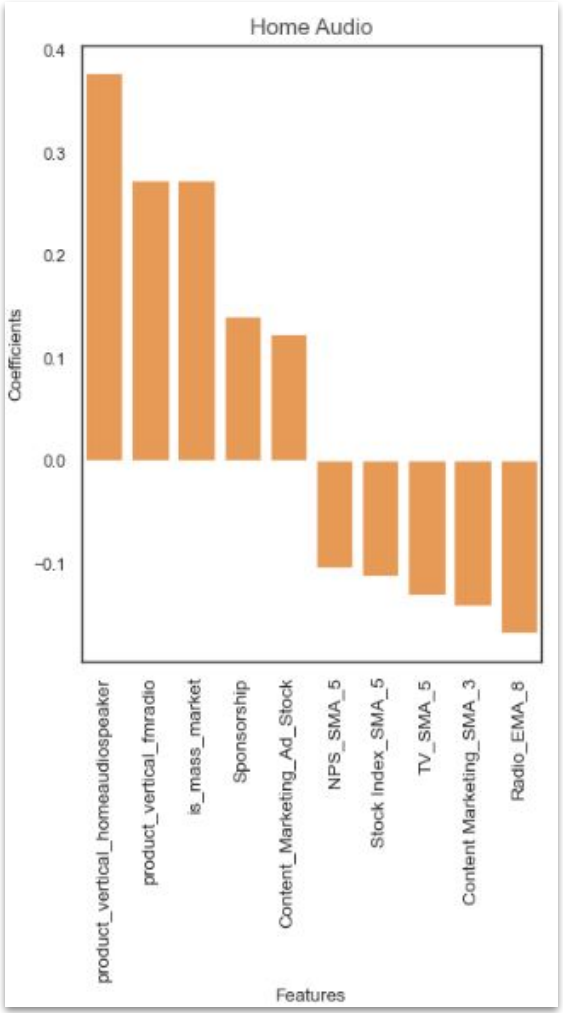
- To maximize revenue, the company should prioritize promoting home audio speakers and FM radios, as they consistently generate the highest income.

-Mass-market home audio products are more effective contributors to revenue growth compared to luxury products.

- Increasing spending on radio adstock (the carry-over effect of radio advertising) significantly boosts revenue.

- While sponsorship spending has a positive impact on revenue, content marketing spending has a negative effect.

- Cash-on-delivery (COD) payments within this subcategory generally have a negative impact on revenue.



Summary

- **Most sales occur when the discount percentage falls between 50-60%. However, it's important to note that this doesn't necessarily lead to a boost in revenue. Exploratory data analysis (EDA) shows that an average discount percentage between 10-20% is the most profitable, especially for luxury items.**
- **Luxury items are predominant among home audio products, leading customers to prefer COD payments instead of paying upfront.**
- **During festive times like Thanksgiving, increased investment in advertising and attractive promotional offers typically boost revenue. However, merely providing discounts without effective advertising across various media channels is not as effective. Notably, during weeks 32-35 (August), revenue hit its lowest point across all three product subcategories, despite a subsequent increase in median discount percentage. This decline in revenue can be directly linked to the minimal overall ad investment during that timeframe.**