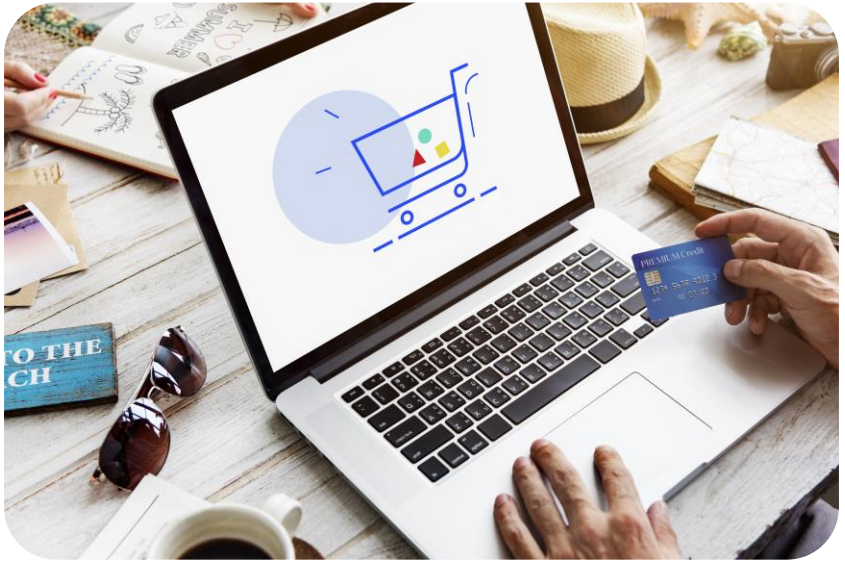
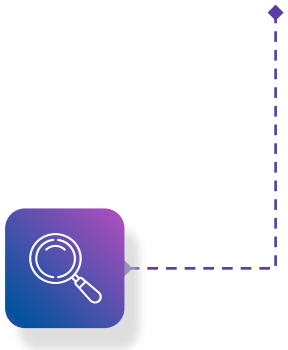
E-Commerce Capstone Project



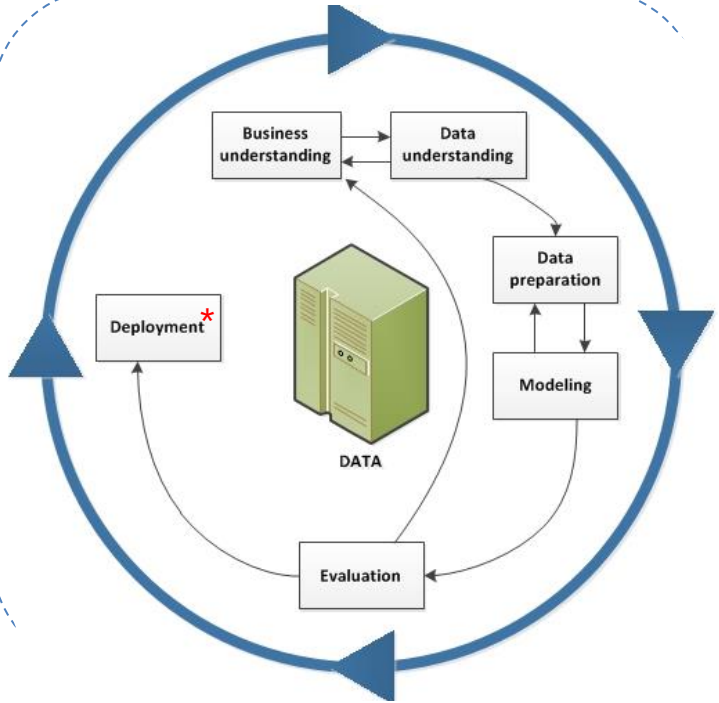
Market Mix Modelling

Anand Umrani

TABLE OF CONTENTS



CRISP-DM Framework



Problem Solving Flow

The project's methodology adheres to the CRISP-DM Framework, with its distinct phases organized in a sequential progression.

* Out of scope

Business Understanding

- ElecKart is a leading e-commerce chain based in Ontario, Canada, specializing in electronic products.
- It sells all major electronic items such as TVs, cameras, OLED screens, and microwaves.
- The company follows the marketplace model and acts as a platform where buyers and sellers can interact with each other.
- The brand has shown commendable growth in just a few years of its operation.
- ElecKart experienced a revenue decline in the past year despite significant marketing and promotional spending.

Objectives

Develop comprehensive market mix models for the Camera Accessory, Home Audio, and Gaming Accessory subcategories at a weekly level, utilizing data from July 2015 to June 2016. These models will integrate information on consumer purchases, monthly spends on advertising channels, climatic data, and NPS/stock index to optimize marketing expenditures.

The primary goal is to create an effective model that maximizes revenue.

Performance Driver Analysis

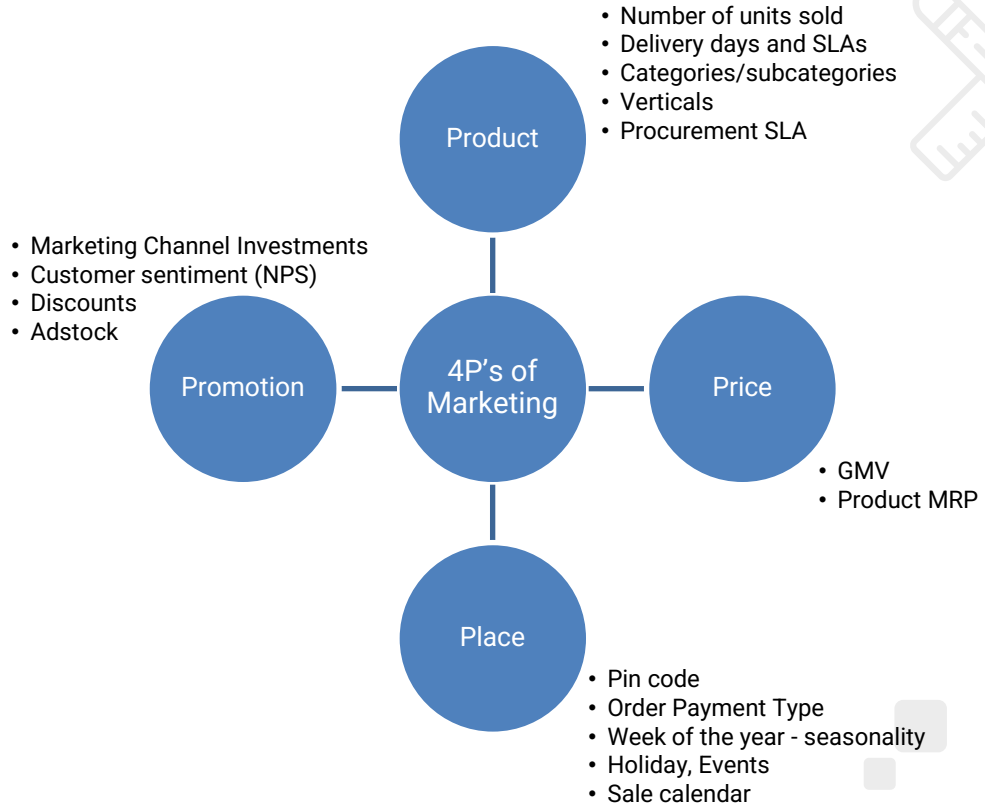
Which KPI's drive the top line performance?

Impact Analysis on Marketing ROI

What is the quantitative impact of each commercial lever on revenue?

Optimizing Marketing Spends

How to best allocate the marketing budget to gain the highest outcome



Data Understanding

Following info/data files have been provided:

- Main Consumer file: order details at a daily basis
- Media Investment file: amount invested in each advertising medium for the past year
- Sale Calendar file: having dates from past year when there was a promotional offer
- NPS file: having 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

Data Clean Up & Preparation

1. Handling in correct values in columns
    a. Imputing “\N” value in deliverybdays & delivelycdays by 0
    b. Treating incorrect GMV values (where gmv > product_mrp * units) by imputing such values with gmv/units
    c. Handling Negative values for product_procurement_sla, deliverybdays & delivetycdays by dropping them
    d. Handling large values (0.30%) for product_procurement_sla by dropping them
2. Duplication of data
    a. Converted all column data into lower case
    b. Dropped duplicate (6.33%) rows
3. NULL values and White spaces
    a. Initially there were not NULL values in the dataframe. But, there were quite a few whitespaces present in some of the columns. These whitespaces were converted into NaNs and then such values were dropped.
4. Dropping insignificant columns
    a. Dropping columns with single unique value (such columns add no value to model building or analysis)
    b. Similarly, dropping ID columns which are of no use in analysis

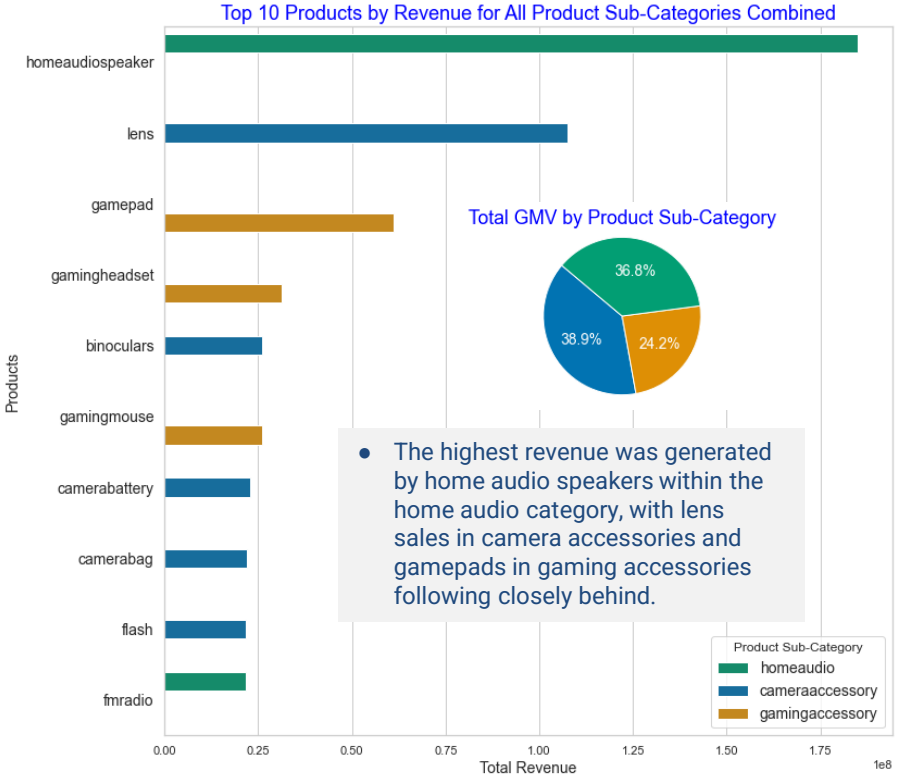
Data Clean Up & Preparation

5. Outliers Treatment
    a. Following variables had outliers – ‘sla', 'deliverybdays', 'deliverycdays', 'gmv', 'product_mrp', 'list_price’
    b. For these variables values above 99 percentile value were capped. This ensured that data loss is kept to minimum
6. Retaining only 1 year data
    a. As per the objective statement, only data between July 2015 to June 2016 was kept. As result records were dropped
7. Transforming categorical variables to numerical
    a. 2 level binary encoding was done for categorical variables
    b. One hot encoding was performed for categorical variables with multiple levels by creating dummy variables
8. Data preparation before model building
    a. Merging Order dataset with all other secondary dataframes
    b. Extracting 3 separate dataframes for 3 product subcategories - camera accessory, gaming accessory and home audio
    c. Daily Order Data rolled up to Weekly Level by aggregating the numeric variables based on Week numbers
    d. Scaling and dividing the master data-frames into train and test datasets for all 3 product subcategories

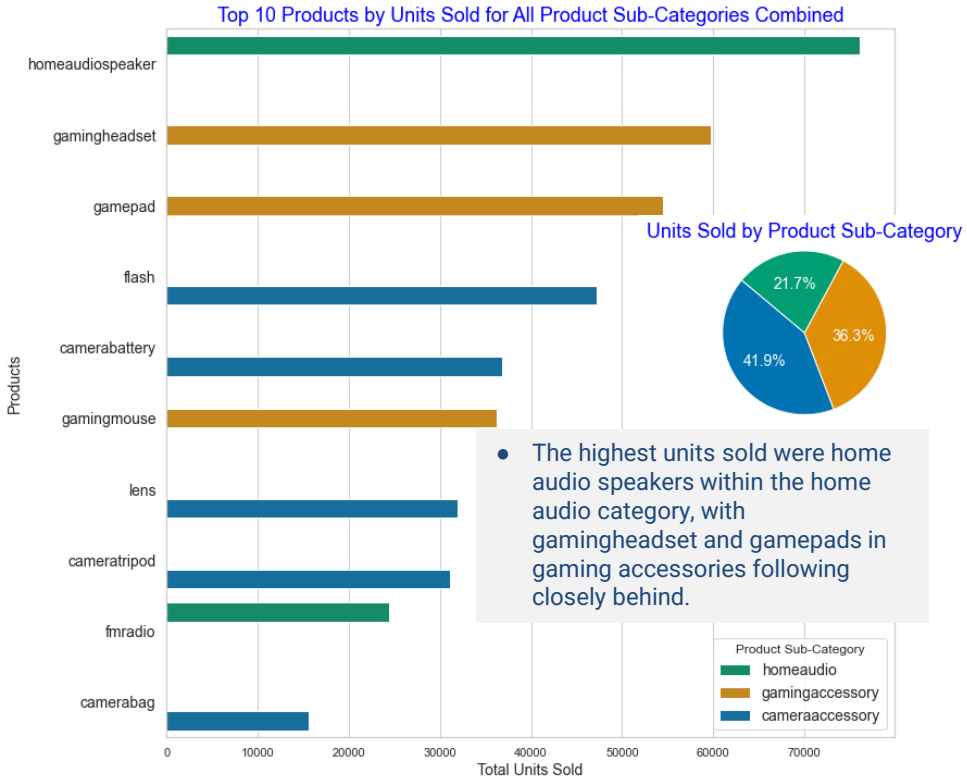
Feature Engineering & Deriving new KPI's

#	Feature	Description
1	Week Number	Creating Week Number column from order date
2	Payday Week	If payday falls within the week, then payday week = 1 else 0
3	Holiday Week	If holiday falls within the week, then payday week = 1 else 0
4	List price	gmv*units
5	Discount%	Discount% = 100*(product_mrp-list price)/product_mrp
6	Product Type	If gmv > 80 percentile, then premium else mass
7	Lag Variables	Lag variables (lag of 1, 2, 3 days) for all KPI's for distributive lag models
8	Adstock Values	Calculate Adstock values for all advertising media (assuming Adstock rate as 60%)
9	Simple Moving Avg	3 & 5 weeks SMA for all Adv media channels, NPS and Stock Index
10	Exp Moving Avg	8 weeks exponential moving average for all Adv media channels

Insights: EDA

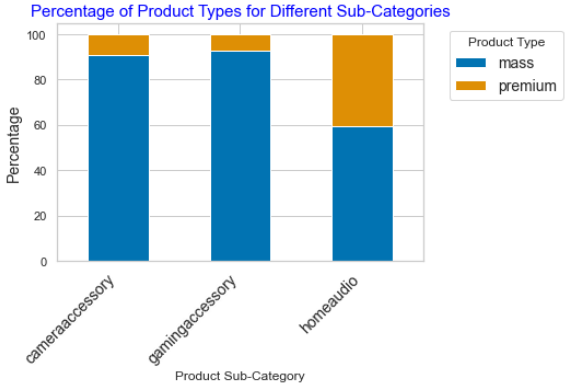


- The highest revenue was generated by home audio speakers within the home audio category, with lens sales in camera accessories and gamepads in gaming accessories following closely behind.

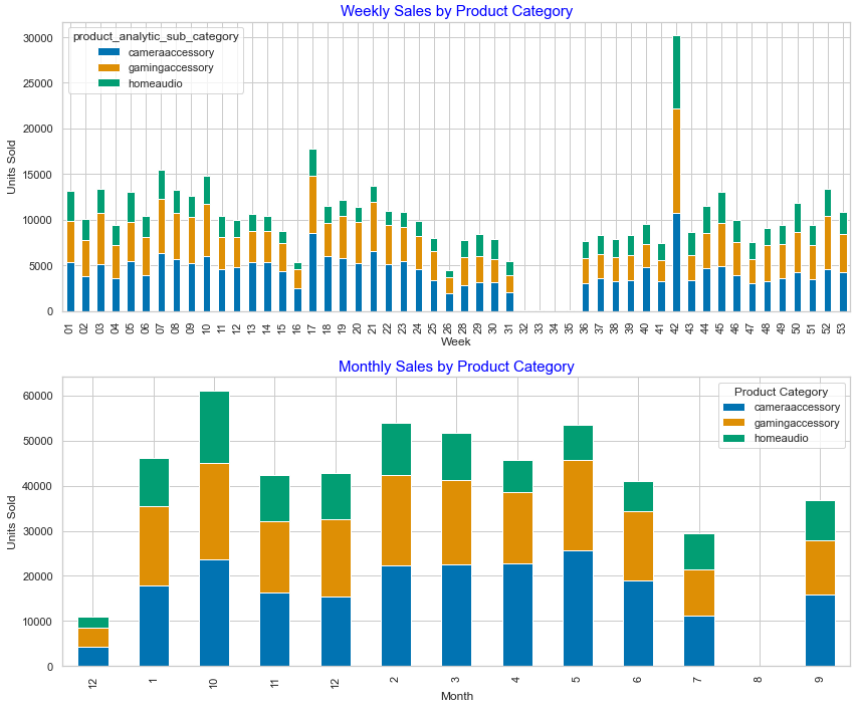


- The highest units sold were home audio speakers within the home audio category, with gamingheadset and gamepads in gaming accessories following closely behind.

Insights: EDA

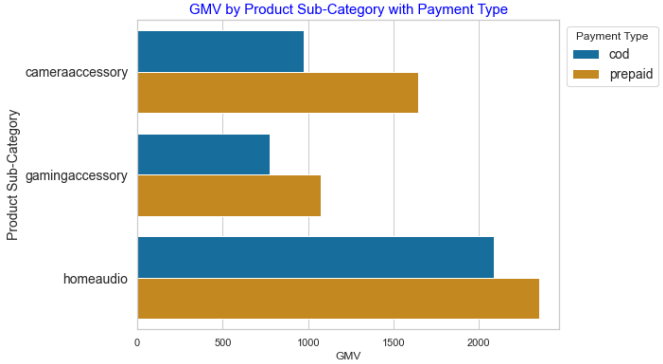


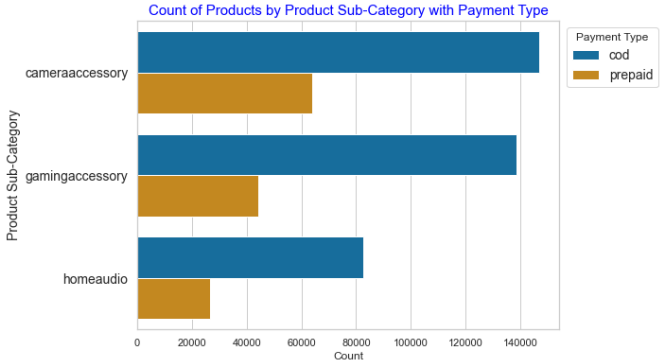
- The majority of units sold were from the mass market category, with Camera and Gaming Accessories being the top-selling subcategories within this segment.
- Among luxury products, Home Audio items stood out as the most favored by consumers.

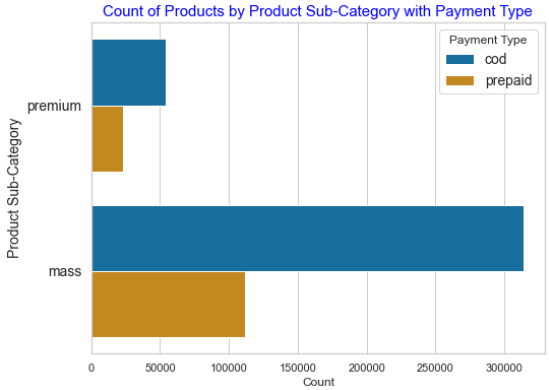


- Maximum units have been sold in the week number 42 during the thanksgiving week in the month of October

Insights: EDA

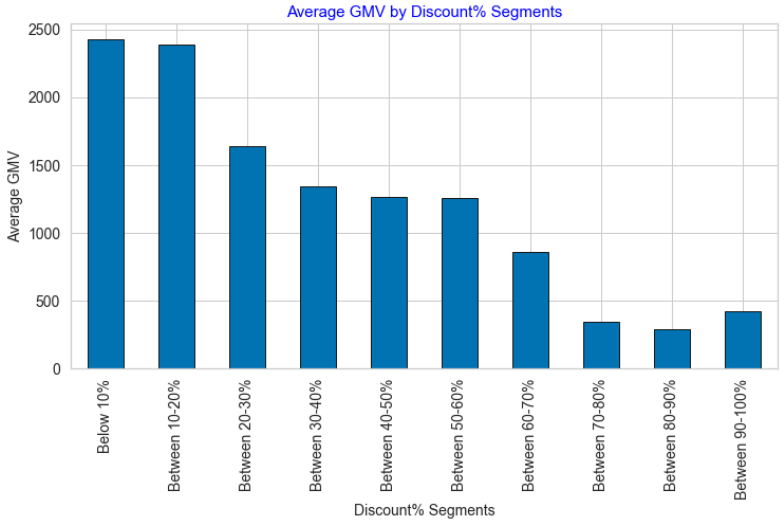


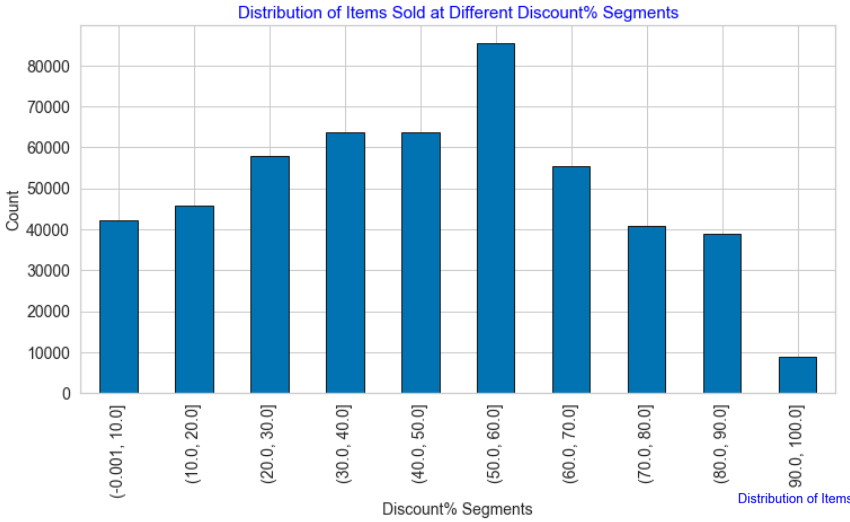


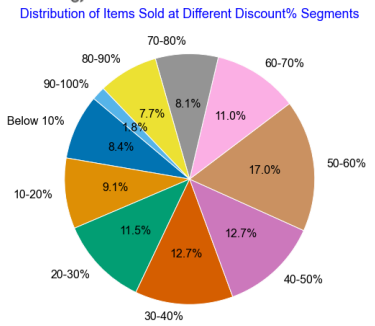


- Cash on Delivery (COD) was the most preferred payment and majority of the sales has come from it, however, the max revenue has come from prepaid payment option.
- Both premium and mass type of products were also paid through COD option.

Insights: EDA

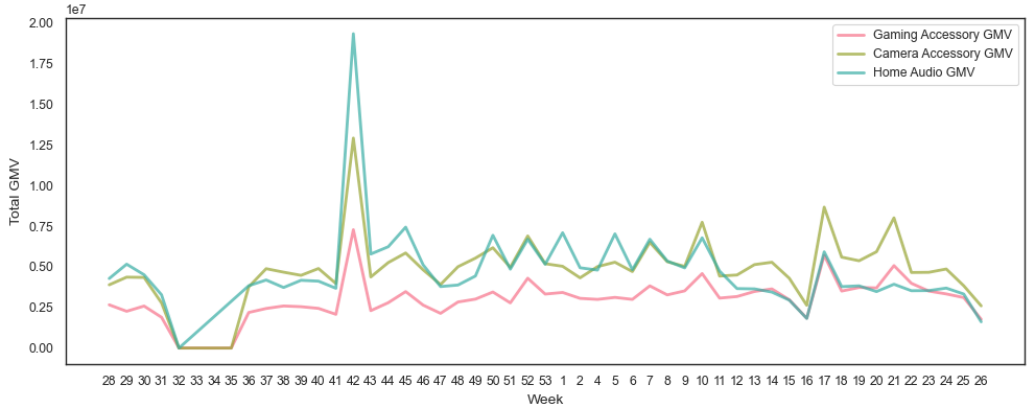


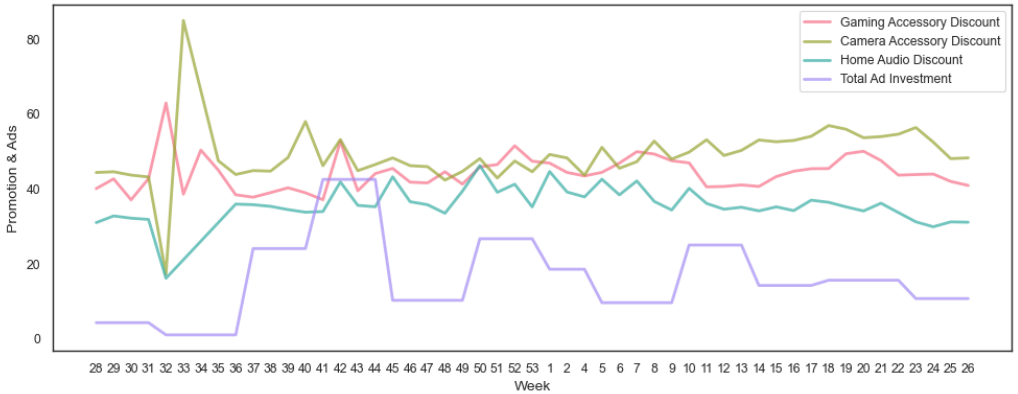


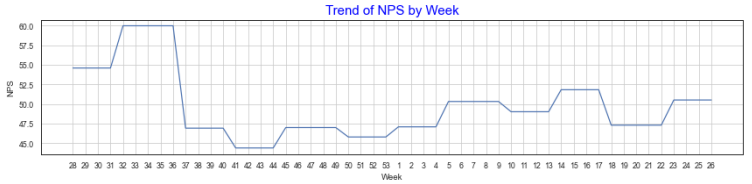


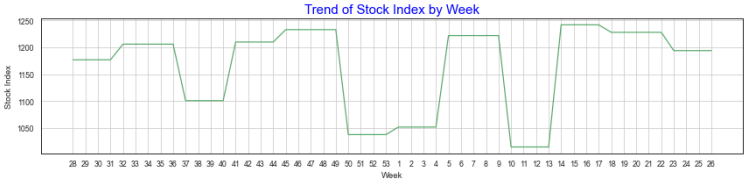
- Revenue was highest when the discount is up to 20%, after that it starts falling.
- This indicates that the company should offer up to 20% discount which will be profitable for it.
- Maximum sale of items (17%) happened when the discount is between 50 to 60%
- Sales increased as % discount increased 50-60% then started dropping again.

Insights: EDA









- NPS during week 32-35 was highest, when discounts were also highest
- The avg discount provided on home audio products is the lower that other categories.
- Week 32 to 35 saw lowest total ad investment and the revenue was also lowest. During the same period discounts were also low except for camera accessory.
- As seen in previous analysis, week 42 (thanksgiving week) revenue for all three categories was highest and during same duration the total amount spend on ads was also highest.

Modeling

Preparing Regression Models for Prediction and Determining Important KPI's

The main goal of the case study is to predict revenue and identify key performance indicators (KPIs) that have an impact on revenue growth.

Additive Model	To predict revenue by summing the individual effects of various factors, making it suitable for scenarios where factors contribute to revenue independently.
Multiplicative Model	This model predicts revenue by multiplying the effects of different factors, making it useful when factors interact with each other to influence revenue.
Koyck Model	The Koyck Model focuses on capturing the dynamic relationship between past & present values of KPIs & revenue, providing insights into how past performance impacts future revenue.
Distributive Lag Model (additive)	This model considers a time delay in the impact of KPIs on revenue, where each time period's effect is added to the previous ones. It's valuable for scenarios where the effect of KPIs accumulates over time.
Distributive Lag Model (multiplicative)	The Distributive Lag Model (Multiplicative) also accounts for time delays in the impact of KPIs on revenue, but in this case, the effects multiply over time. This model is suitable for situations where the impact of KPIs intensifies or weakens as time progresses.

Models – Summary

Product Sub-Category	Linear Regression Model	Cross Validation	R2	MSE
Camera Accessory	Additive Model	No	0.684	0.314
		Yes	-0.857	1.857
	Multiplicative Model	No	0.869	0.283
		Yes	0.906	0.094
	Koyck Model	No	0.799	0.200
		Yes	0.403	0.597
	Distributive Lag Model (additive)	No	0.864	0.136
		Yes	0.832	0.168
	Distributive Lag Model (multiplicative)	No	0.807	0.418
		Yes	0.843	0.157
Gaming Accessory	Additive Model	No	0.924	0.061
		Yes	0.472	0.528
	Multiplicative Model	No	0.933	0.105
		Yes	0.935	0.065
	Koyck Model	No	0.924	0.062
		Yes	0.530	0.470
	Distributive Lag Model (additive)	No	0.874	0.101
		Yes	0.910	0.090
	Distributive Lag Model (multiplicative)	No	0.931	0.108
		Yes	0.898	0.102
Home Audio Accessory	Additive Model	No	0.855	0.348
		Yes	0.605	0.395
	Multiplicative Model	No	-0.302	0.279
		Yes	0.910	0.090
	Koyck Model	No	0.862	0.331
		Yes	0.618	0.382
	Distributive Lag Model (additive)	No	0.375	1.498
		Yes	0.521	0.479
	Distributive Lag Model (multiplicative)	No	-0.136	0.243
		Yes	0.652	0.348

For R2 (R-squared):
- Green: R2 >= 0.7 (Good fit)
- Amber: 0.5 <= R2 < 0.7 (Moderate fit)
- Red: R2 < 0.5 (Weak fit)

For MSE (Mean Squared Error):
- Green: MSE <= 0.1 (Low error)
- Amber: 0.1 < MSE <= 0.3 (Moderate error)
- Red: MSE > 0.3 (High error)

Model - Selection

Camera Accessory

Model Recommendation:
Multiplicative Model (Cross Validation: Yes)

- R2: 0.906
- MSE: 0.094
- Top 5 KPIs:
  product_vertical_camerabattery,
  product_vertical_lens,
  product_vertical_camerabatterycharger,
  mass_market,
  product_vertical_cameraremotecontrol

Gaming Accessory

Model Recommendation:
Multiplicative Model (Cross Validation: Yes)

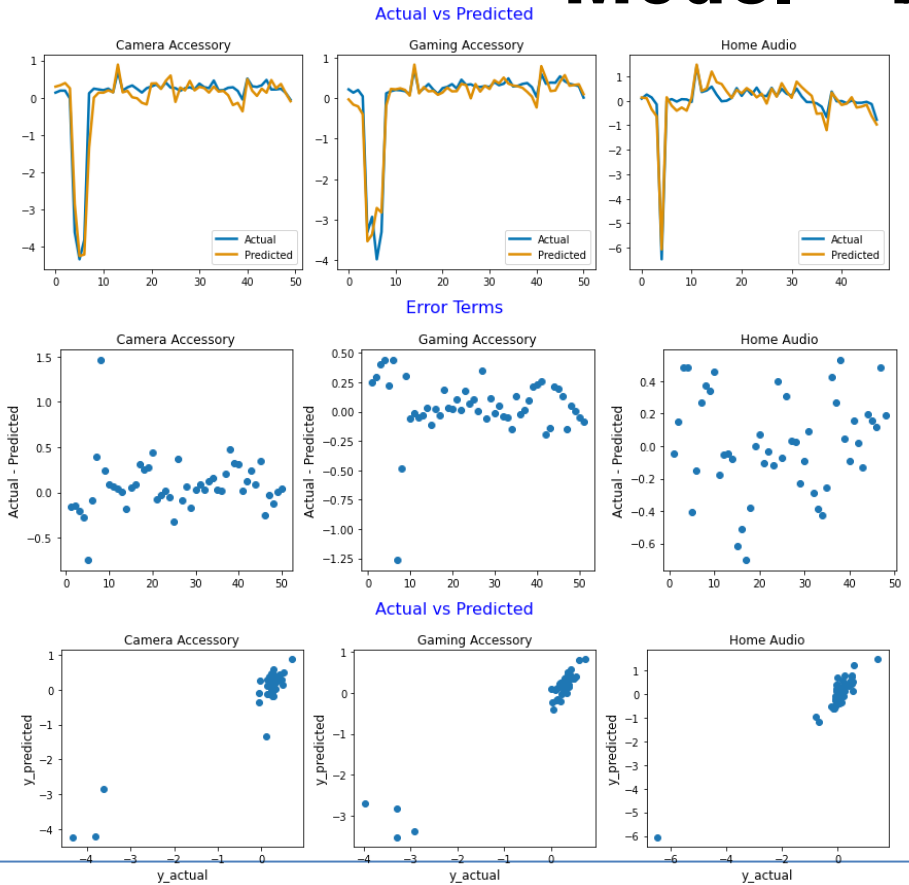
- R2: 0.935
- MSE: 0.065
- Top 5 KPIs:
  product_vertical_gamingheadset,
  product_vertical_gamepad,
  product_vertical_gamingmouse,
  mass_market,
  Online marketing_SMA_3

Home Audio

Model Recommendation:
Multiplicative Model (Cross Validation: Yes)

- R2: 0.910
- MSE: 0.090
- Top 5 KPIs:
  product_vertical_homeaudiospeaker,
  mass_market,
  Total Precip (mm)
  product_vertical_fmradio,
  Other_Ad_Stock

Model – Evaluation



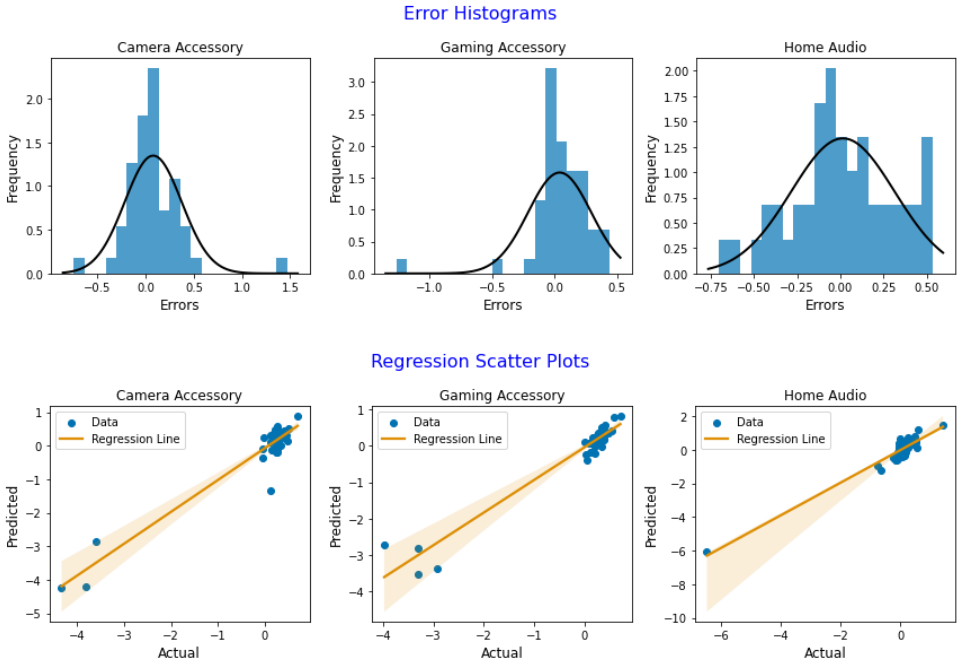
- Actual vs. Predicted Data Plot

This plot compares the actual target values with the predicted values generated by the model. It shows how well the model's predictions align with the actual data points.

- Scatter Plot of Actual vs. Predicted Data

A visual representation of the relationship between actual and predicted values shows that there is no pattern

Model – Evaluation



- Error Histogram Plot

Errors are normally distributed which is an important assumption for regression models

- Scatter plot with best fit line

Regression line fits the data well

Equations for Revenue Growth

From the best fit line, Top 5 features/KPI's are shown in the equation that have positive impact on Revenue(gmv)

Camera Accessory

Revenue = 0.0 + (0.174 * product_vertical_camerabattery) + (0.17 * product_vertical_lens) + (0.117 * product_vertical_camerabatterycharger) + (0.112 * mass_market) + (0.099 * product_vertical_cameraremotecontrol)

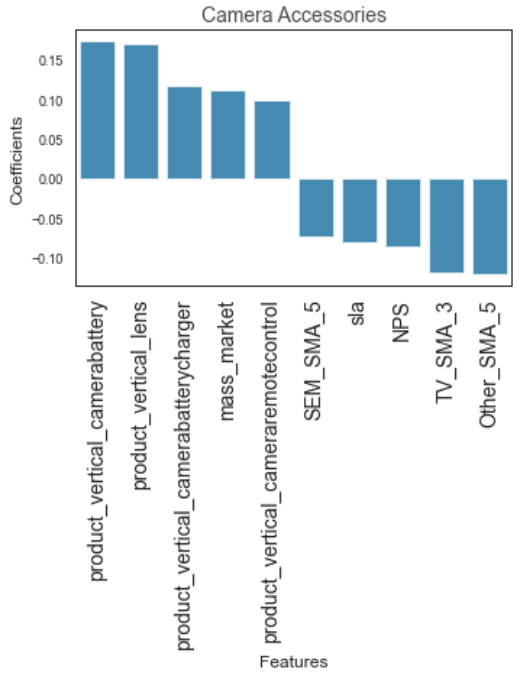
Gaming Accessory

Revenue = 0.0 + (0.294 * product_vertical_gamingheadset) + (0.24 * product_vertical_gamepad) + (0.238 * product_vertical_gamingmouse) + (0.196 * mass_market) + (0.164 * Online marketing_SMA_3)

Home Audio

Revenue = 0.0 + (0.522 * product_vertical_homeaudiospeaker) + (0.206 * mass_market) + (0.162 * Total Precip (mm)) + (0.151 * product_vertical_fmradio) + (0.149 * Other_Ad_Stock)

Model- Recommendations



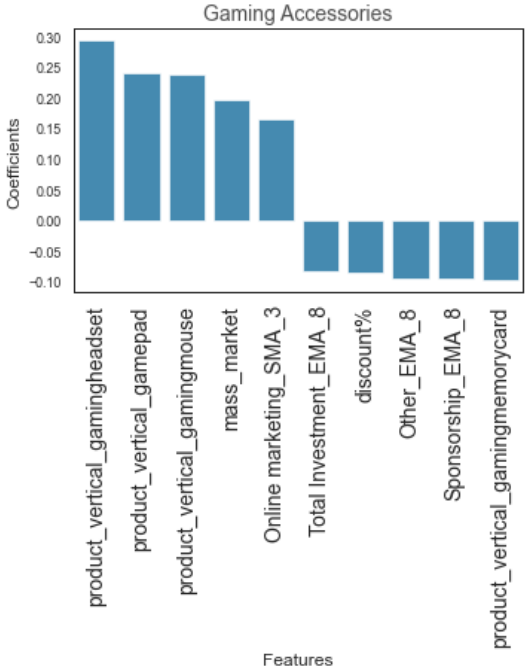
Camera Accessory

- Focus on promoting and selling camera batteries and lenses, as they have the highest positive impact on revenue.
- Include camera battery chargers and remote controls in your product offerings to capture additional revenue opportunities.
- Target the mass market products with your camera accessory products, as it has a positive influence on revenue.
- Evaluate and optimize your Search Engine Marketing (SEM) strategy to reduce its negative impact on revenue.
- Work on improving your Net Promoter Score (NPS) and optimizing marketing strategies, including television advertising, to counter their negative effects on revenue.

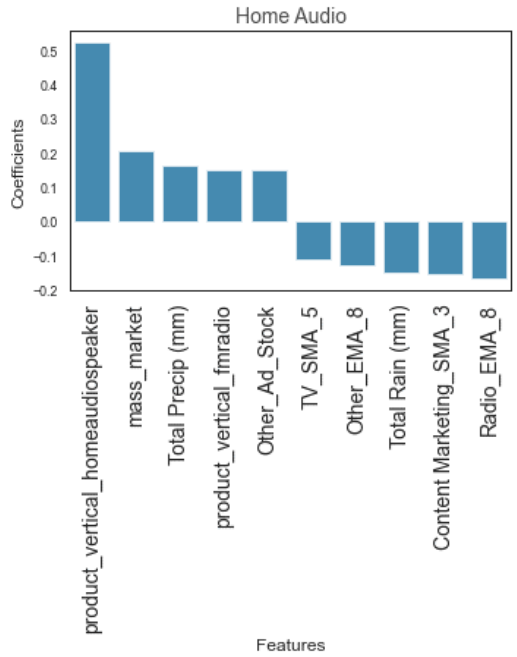
Model- Recommendations

Gaming Accessory

- Prioritize the promotion and sales of gaming headsets, gamepads, and gaming mice, as they have the highest positive impact on revenue.
- Target the mass market products with your gaming accessory products, as it has a significant positive influence on revenue.
- Invest in online marketing strategies, particularly those with a 3-day simple moving average (SMA_3), to boost revenue.
- Monitor and optimize your total investment, with an emphasis on an 8-day exponential moving average (EMA_8), to mitigate its negative impact on revenue.
- Pay attention to pricing strategies, such as discounts, to counteract their negative effects on revenue.
- Evaluate and refine sponsorship strategies (Sponsorship_EMA_8) to improve revenue performance.
- Consider revising or phasing out gaming memory card products if they are underperforming in terms of revenue generation.



Model- Recommendations



Home Audio

- Focus on promoting and selling home audio speakers, as they have the highest positive impact on revenue.
- Target the mass market with your home audio products, as it significantly contributes to revenue generation.
- Leverage advertising and marketing strategies to enhance revenue, with a particular emphasis on the FM radio product vertical.
- Allocate resources to manage and optimize your ad stock for other products (Other_Ad_Stock) to drive revenue growth.
- Keep an eye on TV advertising with a 5-day simple moving average (SMA_5) to balance its impact on revenue.
- Evaluate and improve your advertising strategies for other products (Other_EMA_8) to mitigate their negative impact on revenue.

Overall Recommendations

- **Focus on Mass Market Products:** Given that the majority of units sold come from the mass market category, Eleckart should continue to prioritize this segment. Camera and Gaming Accessories have been the top-selling subcategories within this segment, so efforts should be made to expand and optimize offerings in these areas.
- **Product Performance:** Home audio speakers, lens sales in camera accessories, and gamepads in gaming accessories have performed well. Eleckart should consider expanding and promoting these product categories.
- **Discount Strategy:** The data indicates that offering discounts of up to 20% tends to maximize revenue. Eleckart should carefully plan its discount strategies to maintain profitability while attracting customers.
- **Payment Options:** While Cash on Delivery (COD) is the most preferred payment method, it's essential to note that the maximum revenue comes from prepaid payment options. Encouraging and incentivizing prepaid payments can help boost revenue.
- **Advertising and NPS:** NPS (Net Promoter Score) was highest during weeks 32-35 when discounts were also highest. This suggests that customers respond positively to promotions. Eleckart should align advertising efforts with peak NPS periods to maximize impact.
- **Ad Investment:** Analyzing the relationship between total ad investment and revenue is crucial. Week 42, corresponding to Thanksgiving week, saw the highest revenue, but it also had the highest ad spend. Careful budget allocation and ROI analysis can help ensure efficient advertising strategies.
- **Leverage Luxury Home Audio:** Luxury Home Audio products have been favored by consumers. Eleckart can further capitalize on this trend by enhancing its luxury product offerings and marketing strategies.
- **Week 32-35 Analysis:** During weeks 32-35, both total ad investment and revenue were relatively low. This may suggest that this period experiences lower consumer interest. Eleckart should consider alternative marketing strategies or promotions during these weeks.

THANK YOU!

Anand Umrani