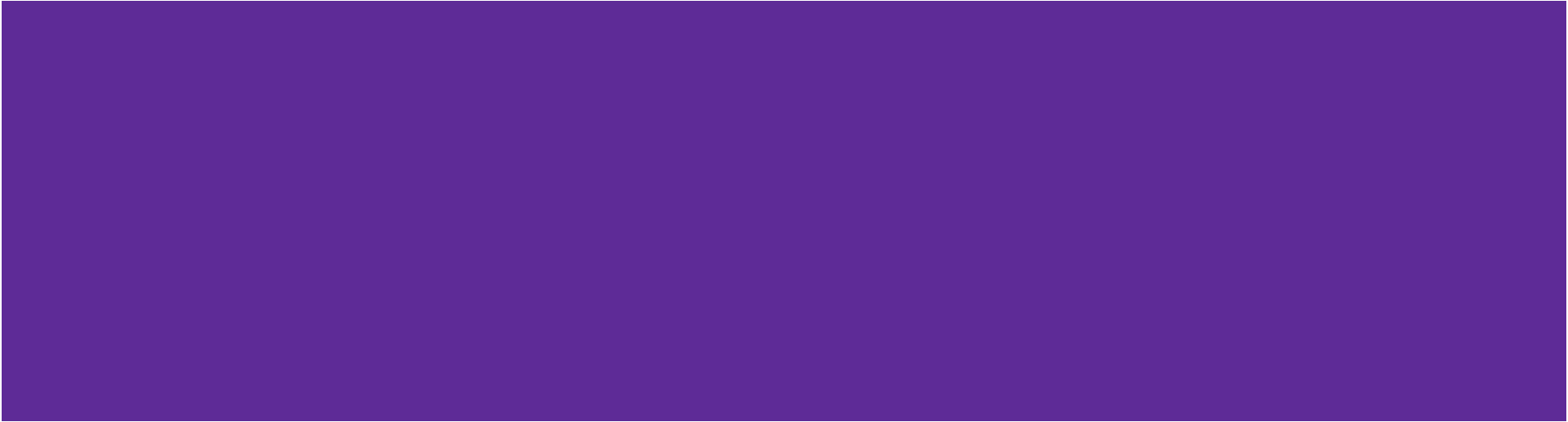# Ecommerce Capstone Project: Market Mix Modelling, Final Submission Report

- Objective: To Model the impact of different Levers on the Revenue/GMV(Gross Merchandise Value) of Company Eleckart
- Submitted by Achal Kagwad on 31st December 2021
- Email id: achal.kagwad@gmail.com Phone No:+91-9108302174
- Linkedin Profile: https://www.linkedin.com/in/achalkagwad/

Business Objective

1. **Performance Driver Analysis:** Which KPIs Drive the Top Line Performance?
2. **Impact Analysis of marketing ROI:** What is the quantitative impact of each commercial lever on revenue?
3. **Optimizing Marketing Spends:** How to best allocate the marketing budget to gain the highest outcome or return on investment?

Problem Solving Methodology: CRISP DM

1. Understanding the Business Data
2. Data Preparation and Feature Engineering
3. Exploratory Data Analysis and Visualizations
4. Preparing Regression Models For Prediction and Determining Important KPIs
5. Model Evaluation and Model Selection
6. Model Deployment(Out of Scope For the Project)

Understanding Data

1. **Main Consumer File** with Order Details at a daily basis
2. **Media Investment File** with amount invested in each advertising medium for the past year.
3. **Sale Calendar File** showing the dates from the past year where there was a promotional offer
4. **NPS Score** showing Net Promoter Score(Voice of Customer) and company stock value for last year
5. **Weather File** having detail of weather Reports from last year in the state of 'Ontario Canada'

Understanding Data: 4Ps of Marketing

1. **Product**:
    1.1. Number of Units Sold
    1.2. Delivery Days and SLAs
    1.3. Categories/Sub Categories
    1.4. Verticals
    1.5. Procurement SLAs
2. **Price**:
    2.1. GMV: Gross Merchandise Value
    2.2. Product MRP
3. **Place**:
    3.1. Pincode | Order Payment Type | Week of the Year(Seasonality) | Holiday Events | Sale Calendar
4. **Promotion**:
    4.1. Marketing Channel Investments | NPS-Customer Sentiment | Discounts | Adstocks

Data Preparation and Clean Up

1. **Handling Incorrect values in some columns**
   - **1.1.** **Imputing "\N"** value in deliverybdays & deliverycdays by 0
   - **1.2.** Treating incorrect GMV values **(where gmv > product_mrp * units**) by imputing the faulty MRP values with GMV/units
   - **1.3.** Handling **Negative values** for product_procurement_sla, deliverybdays & deliverycdays by dropping them
   - **1.4.** Handling **large values(0.3%)** for product_procurement_sla by dropping them
2. **De-Duplication Of Data:**
   - **2.1.** After converting all column values to lowercase, we see that there are around **99283 (6.33%) rows that are duplicates.** We went ahead and dropped them
3. **Treating Null Values and Whitespaces:**
   - **3.1.** 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
   - **3.2.** We first converted these whitespaces to NaNs and the dropped these values.
4. **Dropping Insignificant Columns**
   - **4.1.** Dropping Columns with **Single Unique Value** (as it doesn't add any information to the analysis)
   - **4.2.** Dropping some of the '**Id' Columns** which are insignificant to the analysis

Data Preparation and Clean Up Continued..

1. **Outlier Treatment:**
    1.1. 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
    1.2. 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**
    1.3. Thus the outliers couldn't affect the predictive model while at the same time there was enough data to build a generalizable model.

2. **Selecting One Year Data:**
    2.1. **Selecting 1 Year Dat**a from July, 2015 – June, 2016. In the process, 592 records were dropped.

3. **Converting Categorical Features to Numerical Form:**
    3.1. **Binary encoding** for categorical variable with 2 levels
    3.2. One Hot Encoding for categorical variable with multiple levels by creating **dummy variables**

4. **Additional Data Preparation For Model Building:**
    4.1. **Merging** Order dataset with all other secondary dataframes
    4.2. Extracting **3 separate dataframes for 3 product subcategories** - camera accessory, home audio and gaming accessory
    4.3. **Roll Up daily Order Data to Weekly Level** by aggregating the numeric variables based on Week#
    4.4. **Train Test Split -Dividing** the master data frames into train and test datasets for all 3 product subcategories followed by **Scaling(we used standard scaler)** of values for appropriate interpretation of features.

Feature Engineering: Creation of New KPIs

1. **Week#**: Generating Week Column from Order Date
2. **List Price**: List Price= GMV*Units
3. **Payday Week**: If Payday falls within the week then payday week=1 else 0
4. **Holiday Week**: If Holiday falls within the week then payday week=1 else 0
5. **Product Type** (Luxury/Mass Market): Is GMV value is greater than 80 percentile, then luxury else mass market
6. **Discount%**: Discount%= 100*(product_mrp-listprice)/product_mrp
7. **SMA(Simple Moving Average)**: 3 and 5 Week SMA for all Advertising Media Channels, NPS and Stock Exchange
8. **EMA(Exponential Moving Average)**: 8 Week EMA for all Advertising Media Channels
9. **Ad Stock Values:** for all Advertising Media Channels(assuming adstock rate as 60%)

Visualizations: An Insight into Data

Analyzing number of Items Sold & Revenue across the 3 Product Sub Categories
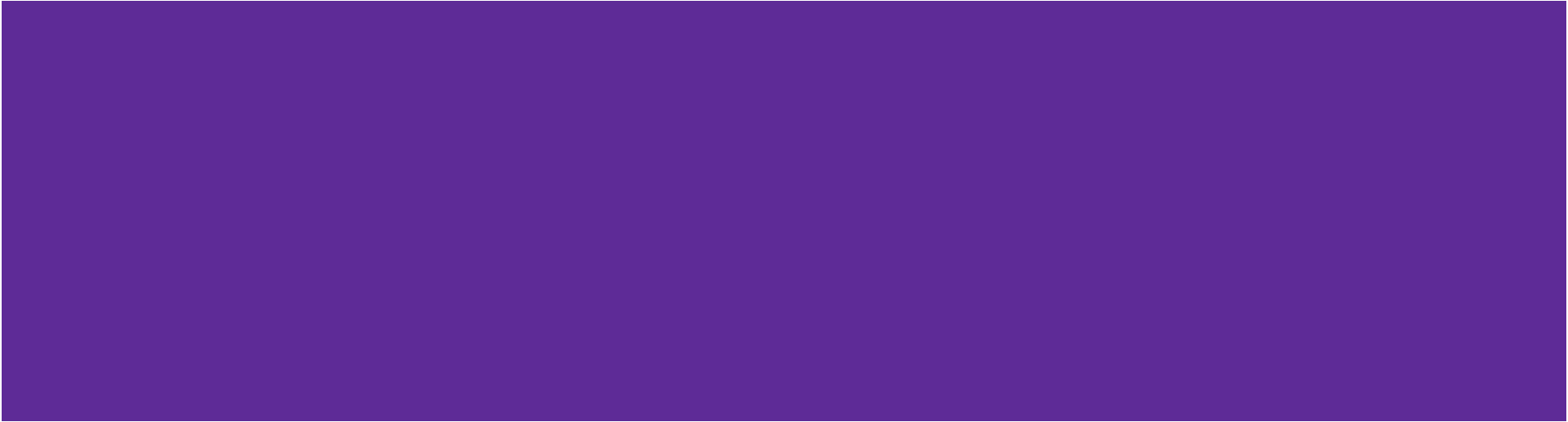
Units Sold Across 3 categories

- Most of the units sold belonged to the mass market category.

- Among mass market products sold, Camera and Gaming Accessories related products were sold the most.

- Home Audio products were the most popular among the luxury products sold
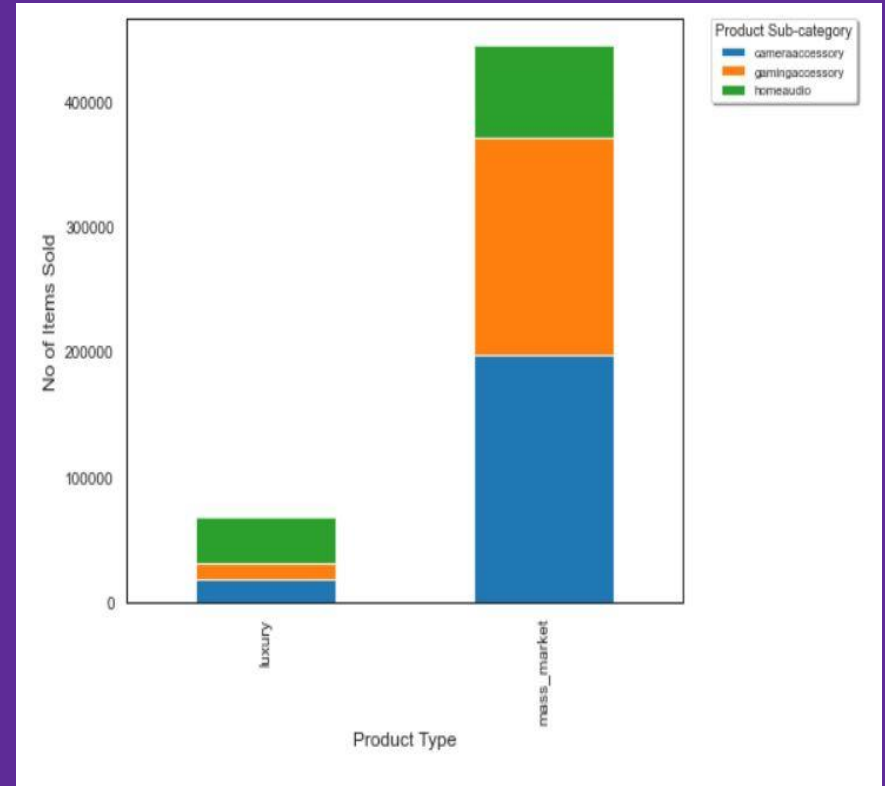
Percentage of Units Sold across 3 categories

-The percentage of number of orders tagged `mass market` is higher compared to orders tagged `luxury`.
-
- Percentage of luxury products under `HomeAudio` is higher compared to the other sub categories. This implies people are willing to buy more of luxury items for HomeAudio category.

Percentage of Luxury & Mass-market Product for different Sub-categories
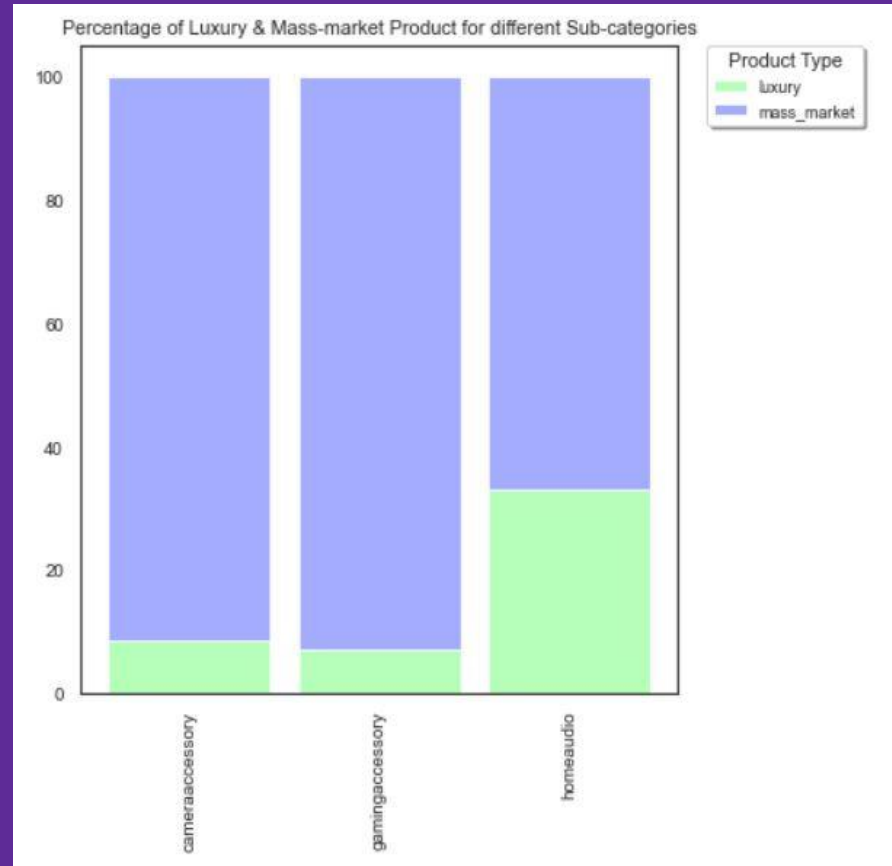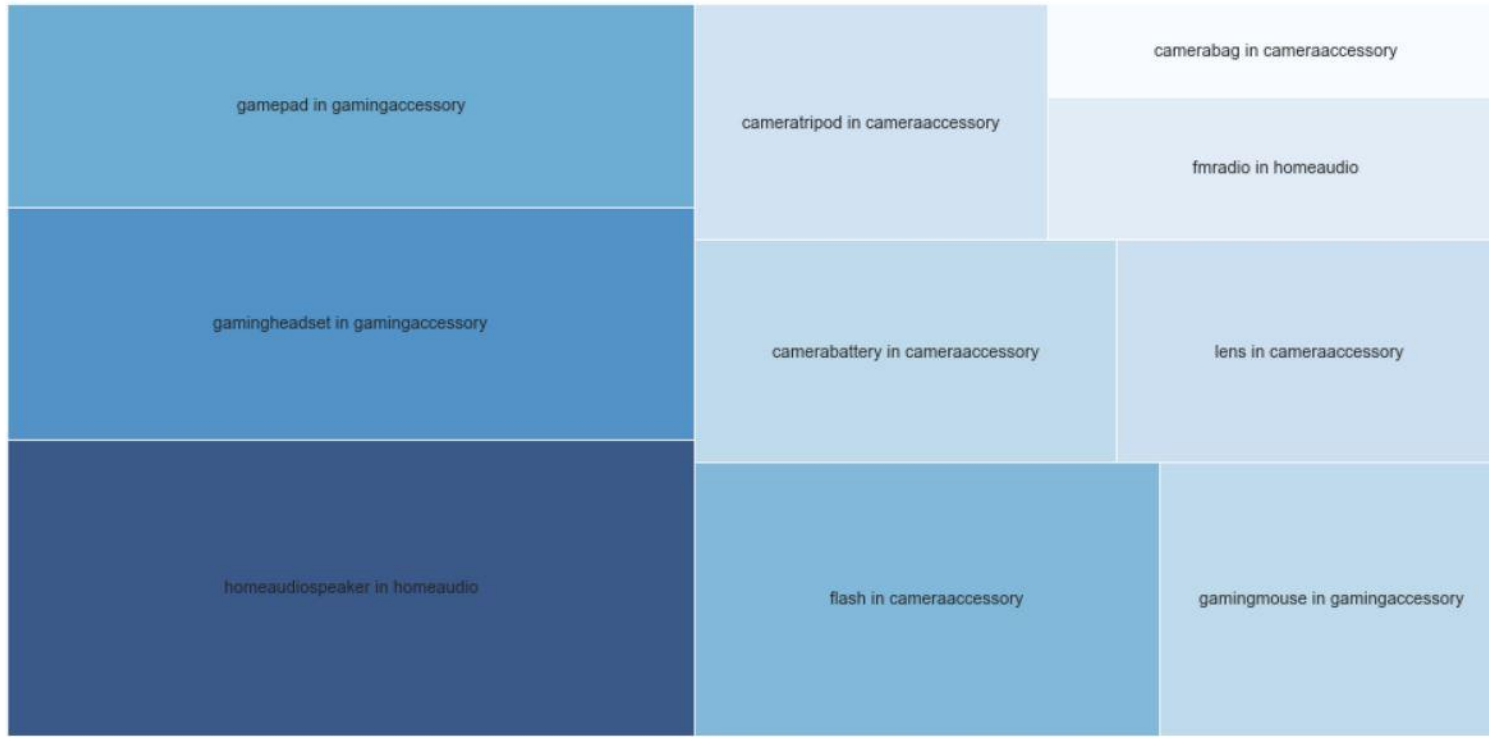
Insights: Above Plot shows top 10 product verticals with highest **ORDERS** across product categories. We can see `homeaudiospeaker in homeaudio` had the most no of sales followed by `gamingheadset & gamepad in gamingaccessory`
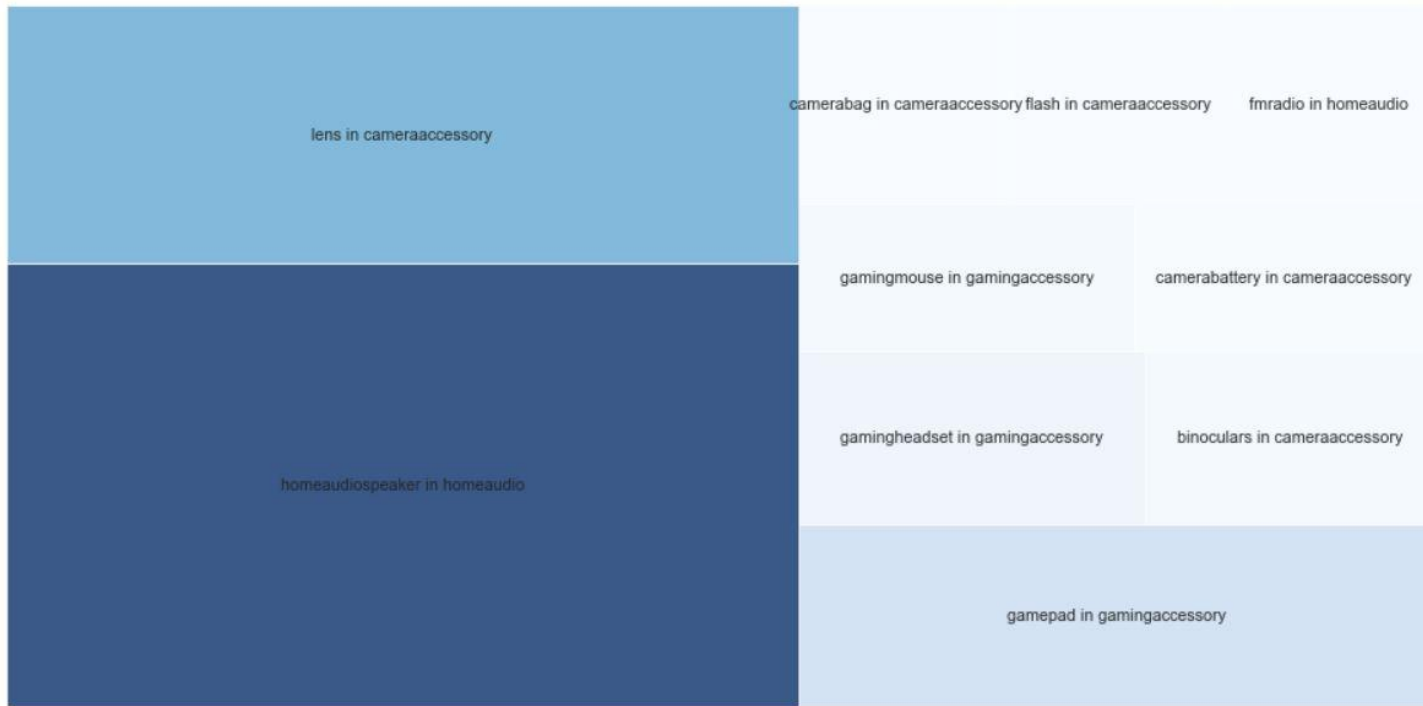
Insights: Above Plot shows top 10 product verticals with highest **REVENUE** across product categories. `homeaudiospeaker in homeaudio` brought the largest revenue followed by `lens in cameraaccessory` & `gamepad in gamingaccessory`.

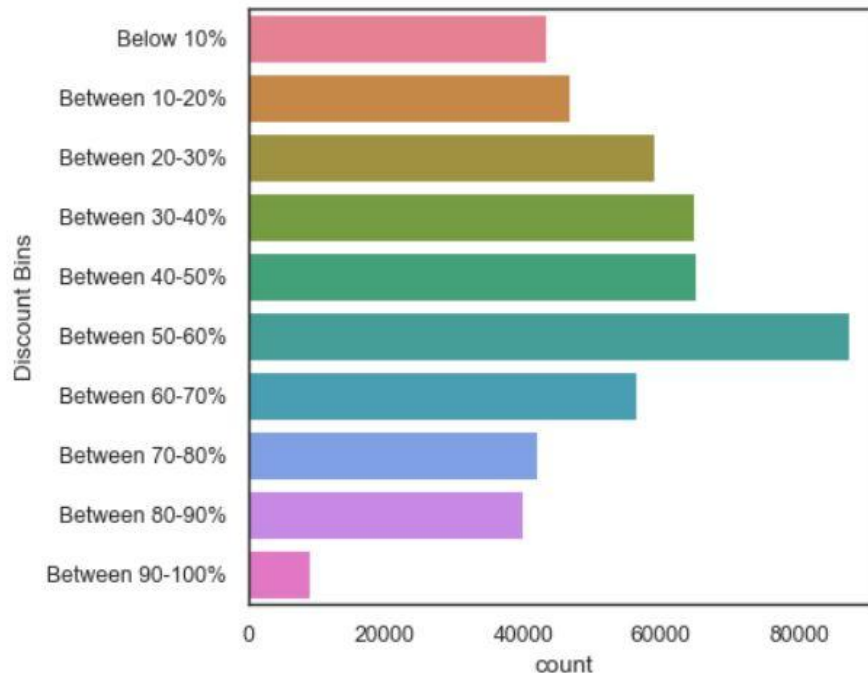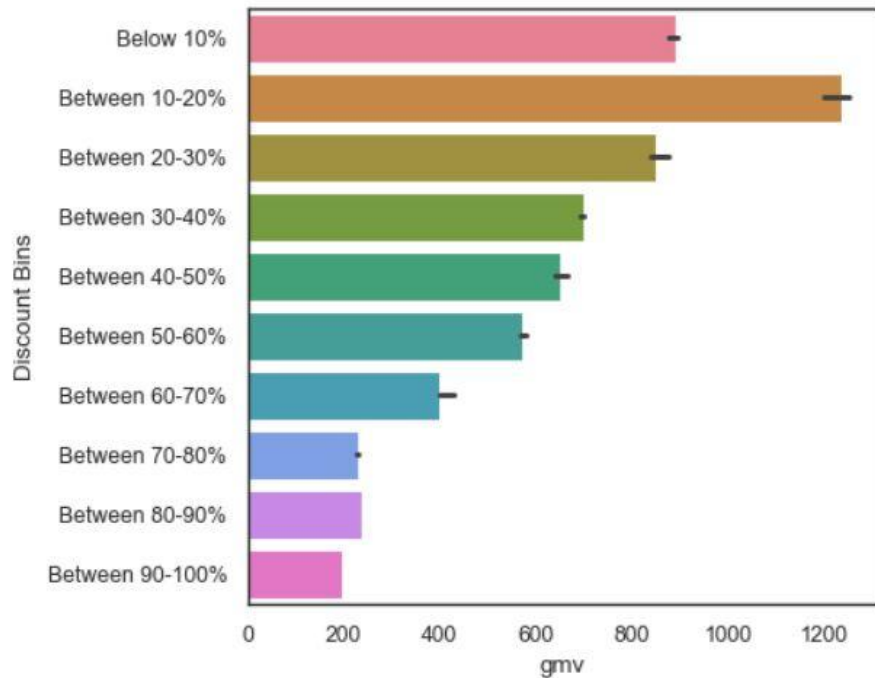# Analyzing how Revenue and Orders/Sales vary based on discount%

Impact of Discount% on No of Orders/Sales

Most of the sales takes place when the discount is between 50-60%


No of Orders at Different Discount%
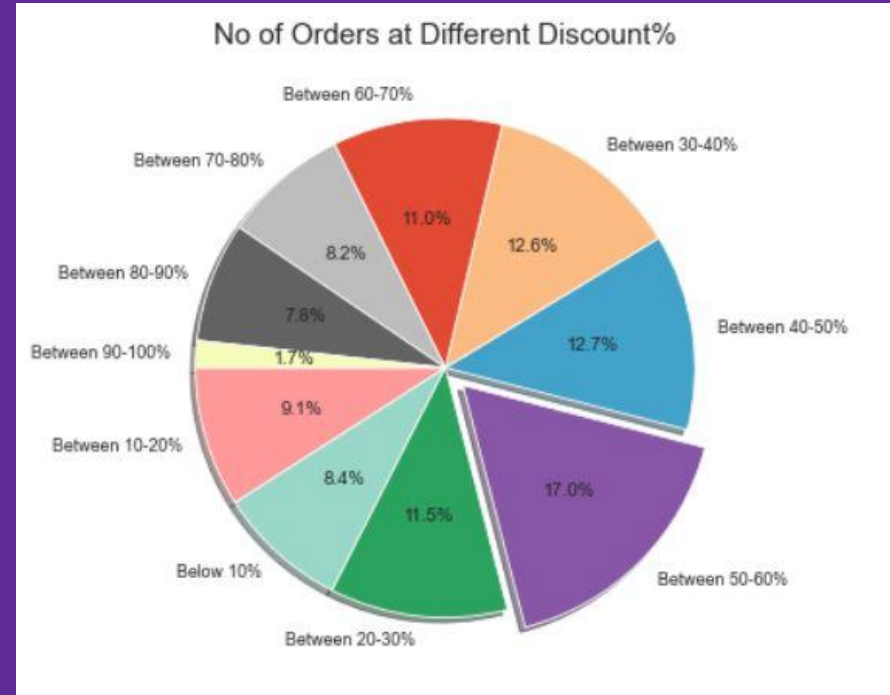
Insights: -Median **Revenue** is maximum when Average discount% is between 10-20%. But beyond that, average revenue slowly starts to decline.
- The **Sales** on the other hand shows a steady increase with increase in Discount percentage till it peaks at 50-60% after which it starts to fall again.
- This 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.¶

Impact of Discount% on Revenue

- The median discount percentage offered for luxury items is less compared to that of Mass Market Products.

- This is a known trend among luxury products or luxury brands to offer limited or no discounts to retain the exclusivity of their products.
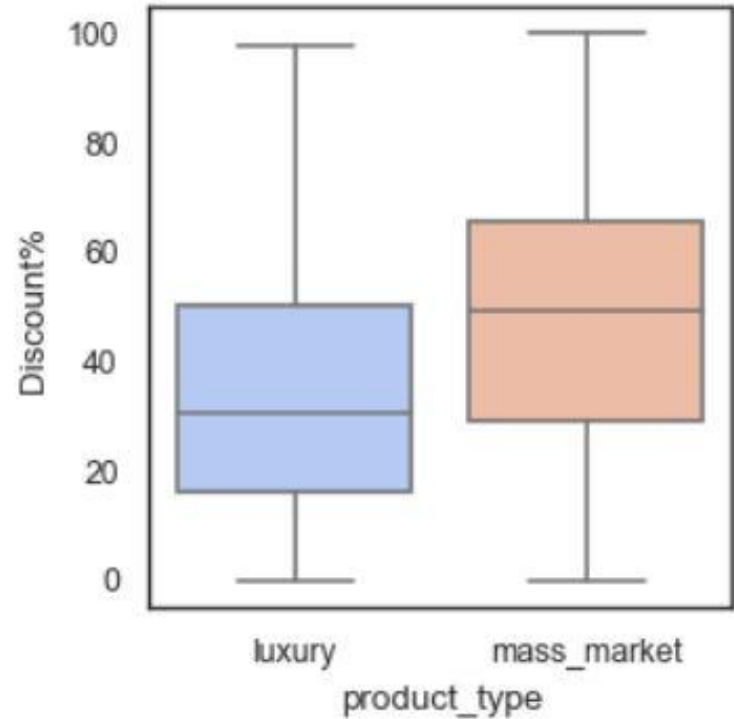
Percentage of Luxury & Mass-market Product under different Discount Group
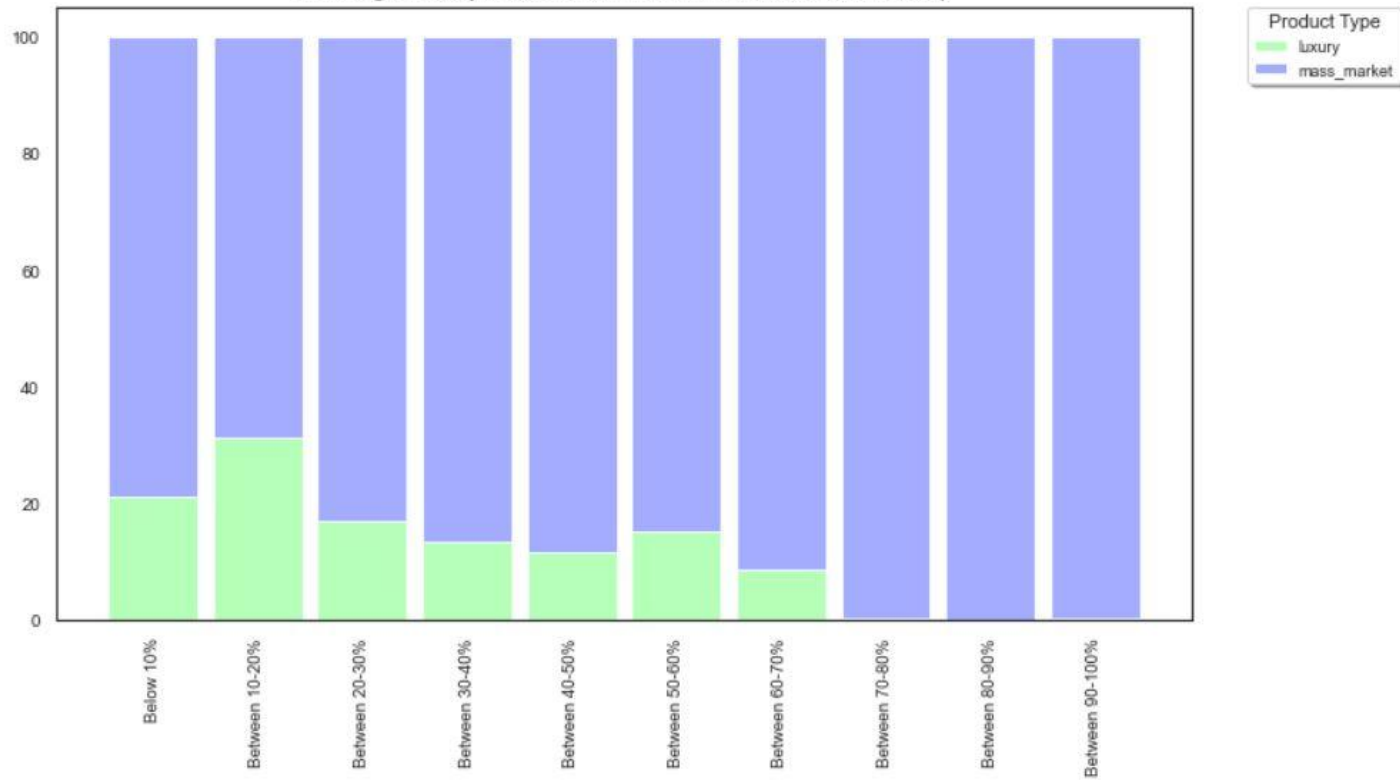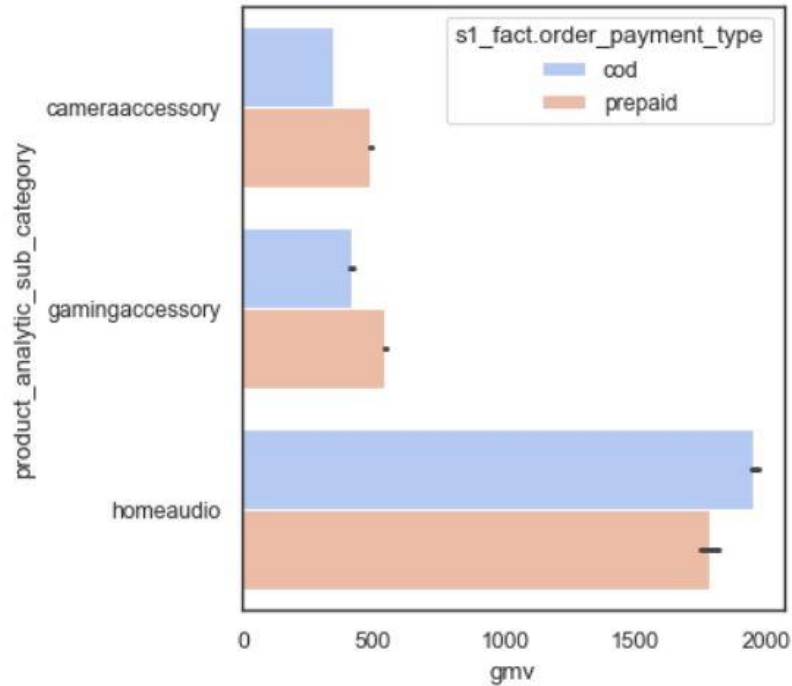
Insights:
We will find around 35% of products where a discount given between 10-20% are luxury products. This is bin which has the highest amount of products where luxury items are sold which implies a discount of 10 to 20% for luxury items usually works and gets sold and thus maintains its premiumness.

General Strategy to Optimize

1. Companies usually optimize on metrics : **Market Share, Revenue and Profit**.
2. Ecommerce companies prioritize optimizing on first Market Share, than Revenue and later on Profit in their long term business plan and success.
3. Here for Eleckart, if it is in the phase of optimizing on sales rather than profit it can continue to offer discounts in the range of 50 to 60% for mass market products to ensure high orders and sales.
4. In later years if Eleckart wants to optimize on Revenue and Profit it can offer discounts in the range of 10 to 20% to ensure high revenue as inferred from our visualizations above.

# Analyzing how Revenue and Orders/Sales vary based on Payment Types
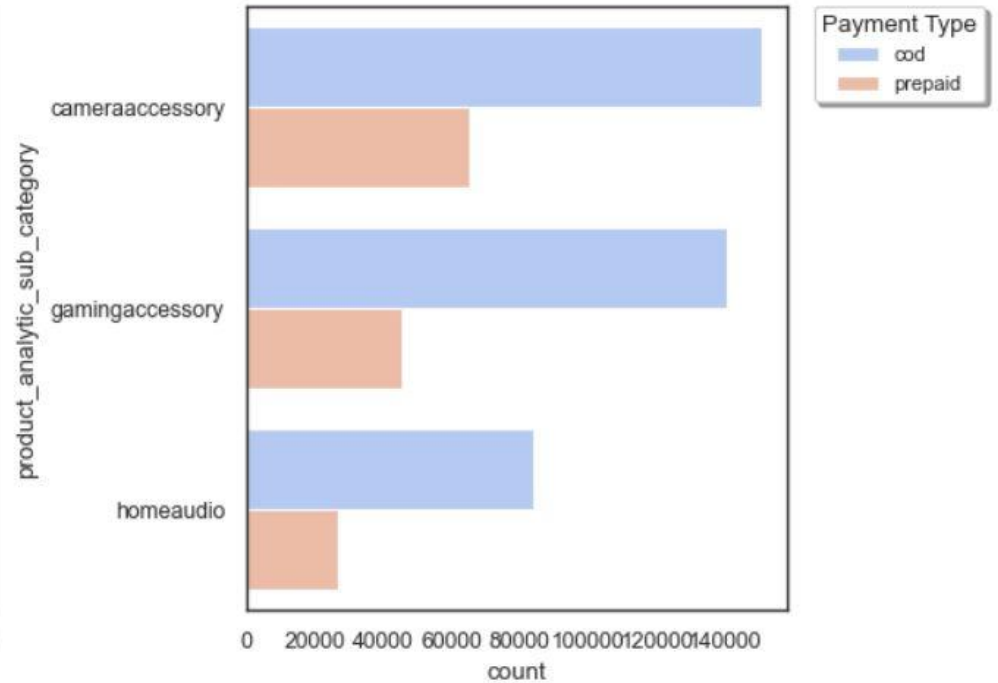
Insights:
- Clearly the Revenue/gmv for homeaudio category is greater compared to others categories
- The Number of Orders/sales is highest in Cameraaccessory category
- The Number of Orders/sales when 'cod(cash on delivery) is almost twice as much more than ''prepaid' in all categories.

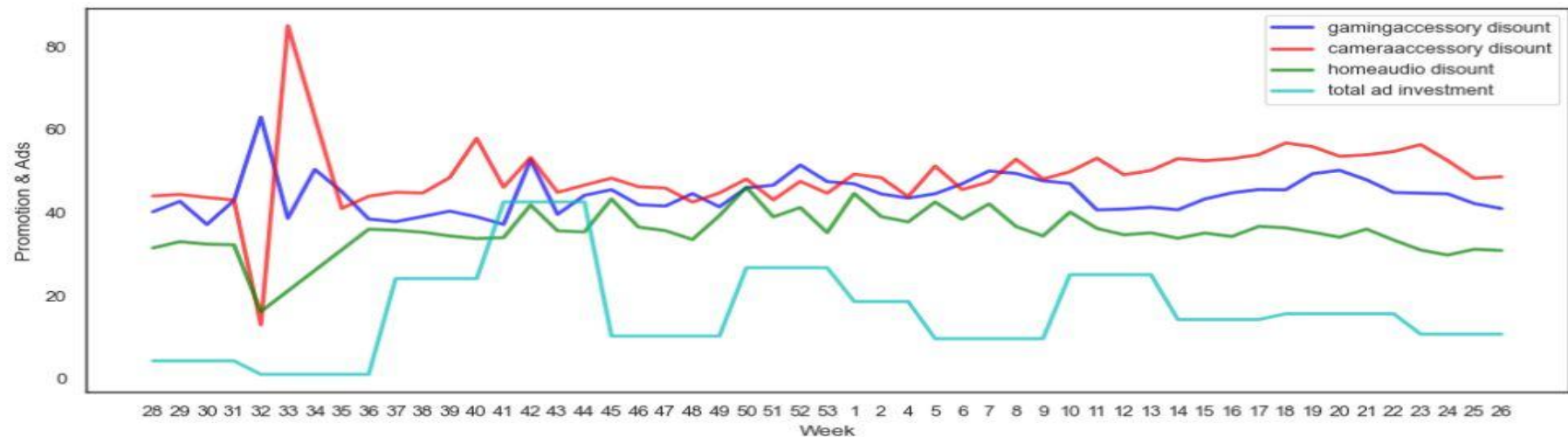# Analyzing Trends of various media channel investments by week
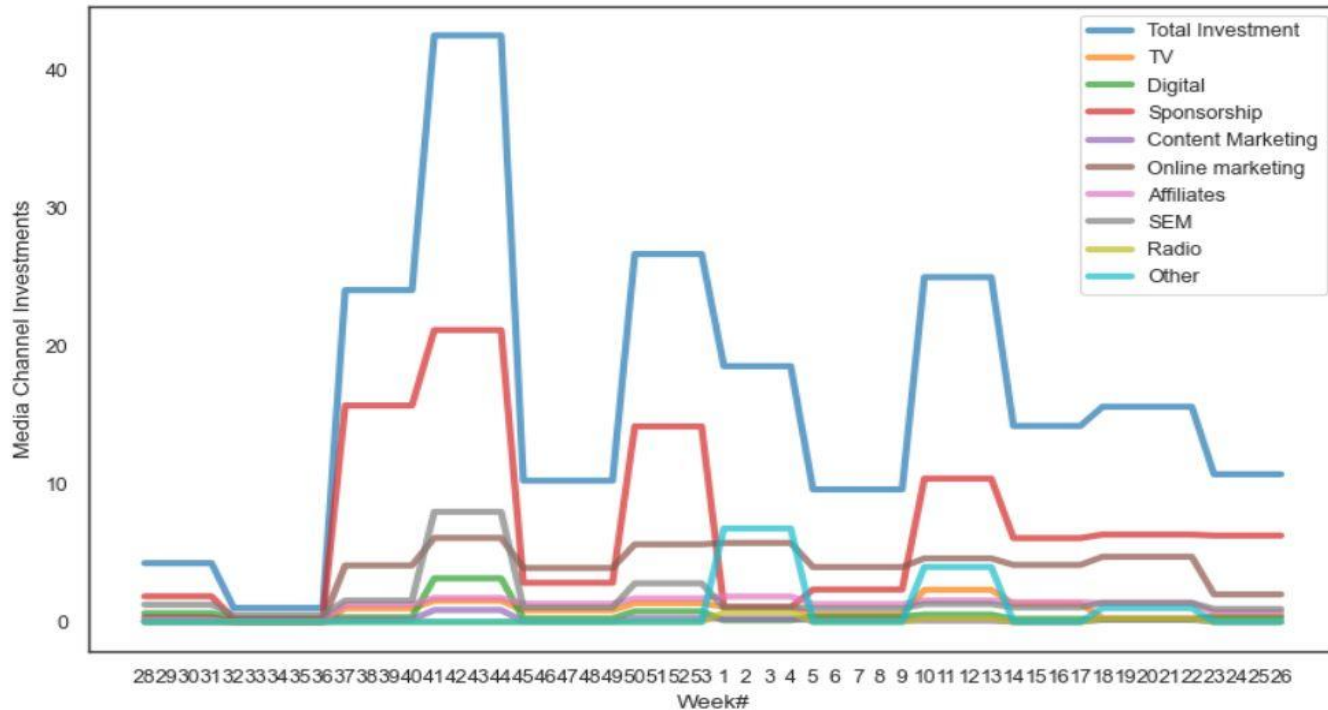
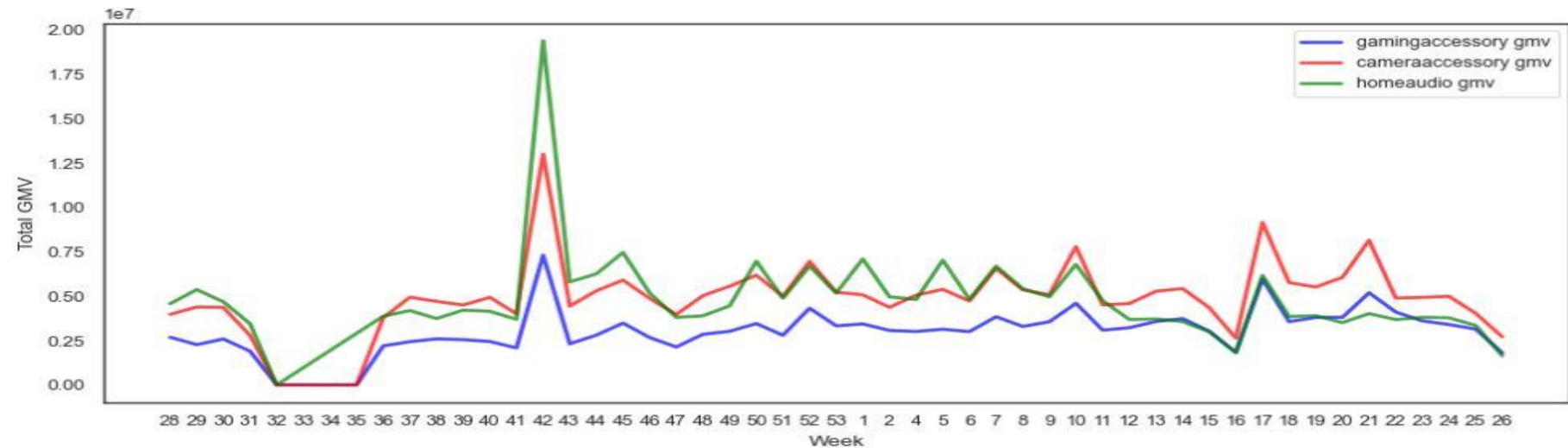**Insights**:
- Over the past year, bulk of the Ad Investment has been made in **Sponsorships** followed by **Online Marketing** & **Search Engine Marketing**.

Effect of Media Investment and Discount% on Revenue

Insights of Effect of Media Investment and Discount% on Revenue

1. For the `week# 42 (during Thanksgiving)`, all the graphs show a steep rise. Revenue increased because of both higher discount% and increased Ad Investment.
2. 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.
3. In general the average discount% offered for home audio products is lesser compared to that of the other product subcategories.

Brief Description of Linear Regression Models Built

Additive Linear Regression Model

1. Linear model is used to capture the current effect of several KPIs.

2. This model assumes an additive relationship between the different KPIs. Hence their impacts are also additive towards the dependent Y variable.

3. The equation can be represented as:
   - $Y = \alpha + \beta_1 A_t + \beta_2 P_t + \beta_3 D_t + \beta_4 Q_t + \beta_5 T_t + \epsilon$

Multiplicative Linear Regression Model

1. Multiplicative model is used when there are interactions between the KPIs.
2. To fit a multiplicative model, take logarithms of the data(on both sides of the model), then analyse the log data as before.
3. The equation can be represented as:
   - 3.1. **Y = e^α .X1^β1 . X2^β2 . X3^β3 . X4^β4 . X5^β5 + ϵ**
   - 3.2. **lnY = α + β1ln(X1) + β2ln(X2) + β3ln(X3) + β4ln(X4) + β5ln(X5) + ϵ'**

Metrics Impacting Revenue in LR Equations

1. - `A` stands for Advertising Impact on Revenue/GMV
2. - `P` stands for Pricing Impact on Revenue/GMV
3. - `D` stands for Discounts or Promotion Impact on Revenue/GMV
4. - `Q` stands for Quality or Product Assortment Impact on Revenue/GMV
5. - `T` stands for Industry Trend and Seasonality Impact on Revenue/GMV

Model Dashboard and Model Selection

All Values are on Test Data, as we are interested how models perform on test data; we choose model which has high R2 and low MSE

	R2 Add	R2 Mul	MSE Add	MSE Mul	Model Chosen	Top Features
Camera Accessory	0.772	0.974	1255947.4	0.311	Mul	10
Gaming Accessory	0.753	0.990	630689.24	0.257	Mul	5
Home Audio	0.958	0.993	351391.67	0.149	Mul	5

Top Features suggested by Multiplicative Model for the 3 product categories-Coefficient Values

best_multiplicative_cameraaccessory_coef_df

	Features	Coefficients
0	constant	14.931
5	product_vertical_lens	1.109
3	product_vertical_camerabattery	1.062
7	Affiliates	0.744
1	is_mass_market	0.526
8	Other_SMA_3	0.152
2	product_vertical_cameraaccessory	0.082
10	NPS	-0.064
9	Other_SMA_5	-0.182
4	product_vertical_camerabatterycharger	-0.678
6	Online marketing	-0.771

best_multiplicative_gamingaccessory_coef_df

|	Features	Coefficients
0	constant	14.576
2	product_vertical_gamepad	1.012
4	product_vertical_gamingheadset	0.568
5	product_vertical_gamingmouse	0.310
1	Discount%	-0.121
3	product_vertical_gamingaccessorykit	-0.223

best_multiplicative_homeaudio_coef_df

	Features	Coefficients
0	constant	15.363
2	product_vertical_homeaudiospeaker	0.236
1	is_mass_market	0.162
3	Total Investment_EMA_8	0.055
5	Sponsorship	0.045
4	Total_Investment_Ad_Stock	-0.095

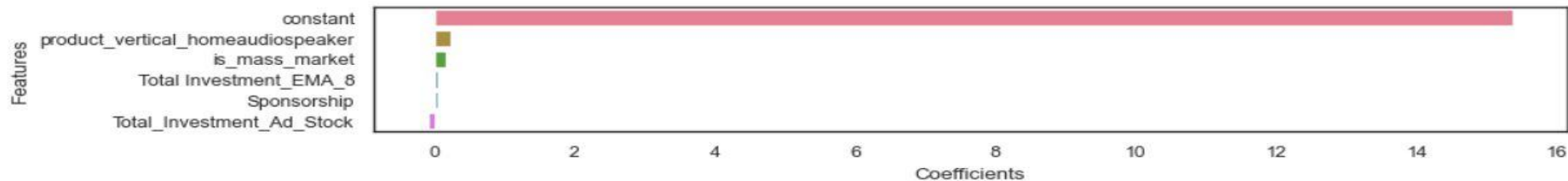# Top Features suggested by Multiplicative Model for 3 product categories-Coefficient Value Plots

Multiplicative Model: Coeffs of Imp Features: cameraaccessory

Multiplicative Model: Coeffs of Imp Features: Gamingaccessory
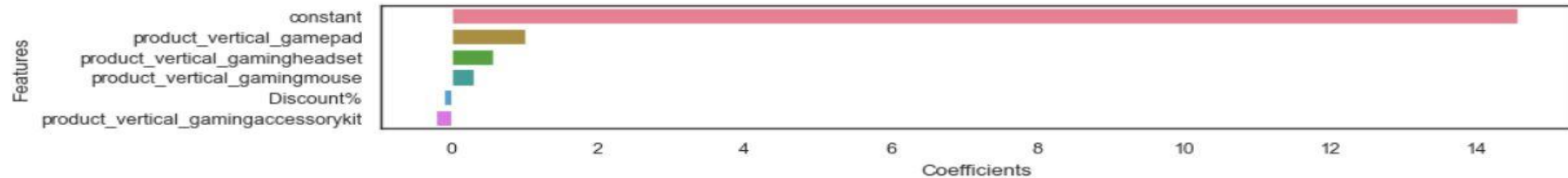
Multiplicative Model: Coeffs of Imp Features: Homeaudio

Equation of Best Fit Lines for 3 product categories

Best fit line equations of Linear Regression(Multiplicative Model) for three product categories:

1. **Camera Accessory: Revenue** = 14.931 + (1.109 × **product_vertical_lens**) + (1.062 × **product_vertical_camerabattery**) + (0.744 × **Affiliates**) + (0.526 × **is_mass_market**) + (0.152 × **Other_SMA_3**) + (0.082 × **product_vertical_cameraaccessory**) - (0.064 × **NPS**) - (0.182 × **Other_SMA_5**) - (0.678 × **product_vertical_camerabatterycharger**) - (0.771 × **Online marketing**)

2. **Gaming Accessory: Revenue** = 14.576 + (1.012 × **product_vertical_gamingheadpad**) + (0.568 × **product_vertical_gamingheadset**) + (0.310 × **product_vertical_gamingmouse**) - (0.211 × **Discount%**) - (0.223 × **product_vertical_gamingaccessorykit**)

3. **Home Audio: Revenue** = 15.363 + (0.236 × **product_vertical_homeaudiospeaker**) + (0.162 × **is_mass_market**) + (0.055 × **Total Investment_EMA_8**) + (0.045 × **Sponsorship**) - (0.095 × **Total_Investment_Ad_Stock**)

This equation implies how much the Revenue *will grow with a unit growth in any of these independent KPIs with all other KPIs held constant.*

Recommendations: Major Features to look out for:

1. For each product category of "Camera Accessory", "Gaming Accessory" and "Home Audio" we need to look out for these features:
2. Which **product verticals** contribute positively towards increasing the revenue and which product verticals affect adversely to revenue?
3. Which **media investment channels** contribute positively towards increasing the revenue and which channels affect adversely to revenue?
4. For the given product category whether promotion of "**Mass market**" products or "**Luxury**" Products lead to increased revenue?

Recommendations- Camera accessory

1. Company should promote "**Lens**" and "**Camera Battery**" as they contribute in increasing revenue.
2. Out of the various Media Investment Channels company should focus more on "**Affiliates**" channels by introducing well researched and known to bring high revenue schemes. "**Online Marketing**" Channel on the other hand impact negatively and thus investment in that channel should be reduced.
3. "**Mass market**" products are better contributors to increased revenue in comparison to luxury products.

best_multiplicative_cameraaccessory_coef_df

	Features	Coefficients
0	constant	14.931
5	product_vertical_lens	1.109
3	product_vertical_camerabattery	1.062
7	Affiliates	0.744
1	is_mass_market	0.526
8	Other_SMA_3	0.152
2	product_vertical_cameraaccessory	0.082
10	NPS	-0.064
9	Other_SMA_5	-0.182
4	product_vertical_camerabatterycharger	-0.678
6	Online marketing	-0.771

Recommendations- Gaming accessory

1. "**Gamepad**", "**Gamingheadset**", "**Gaming Mouse**" should be promoted as they contribute in increasing revenue. On the contrary "**gamingaccessorykit**" results in loss.
2. Advertising spends on "**Affiliates**" and "**TV**" are known to bring positive impact on revenue.
3. "**Mass market**" products are better contributors to increased revenue in comparison to luxury products.
4. Higher percentage of **discounts** for this Gaming accessory category works adversely towards bringing down the revenue.

best_multiplicative_gamingaccessory_coef_df

	Features	Coefficients
0	constant	14.576
2	product_vertical_gamepad	1.012
4	product_vertical_gamingheadset	0.568
5	product_vertical_gamingmouse	0.310
1	Discount%	-0.121
3	product_vertical_gamingaccessorykit	-0.223

Recommendations- Home Audio

1. "**Homeaudiospeaker**", and "**fmradio**" should be promoted as they contribute in increasing revenue.
2. Advertising spends on "**Sponsorship**" is seen to bring positive impact on revenue, while Investment in AdStock is seen to bring negative impact on revenue
3. "**Mass market**" products are better contributors to increased revenue in comparison to luxury products.

best_multiplicative_homeaudio_coef_df

	Features	Coefficients
0	constant	15.363
2	product_vertical_homeaudiospeaker	0.236
1	is_mass_market	0.162
3	Total Investment_EMA_8	0.055
5	Sponsorship	0.045
4	Total_Investment_Ad_Stock	-0.095

Appendix:Other Visualizations

Trends in Nps and Stock Exchange Over the weeks

- Consumer NPS score is highest in weeks 32-35, which coincides with the time when maximum discounts were offered.
- Company Stock Index has seasonal ups and downs over the span of 1 year
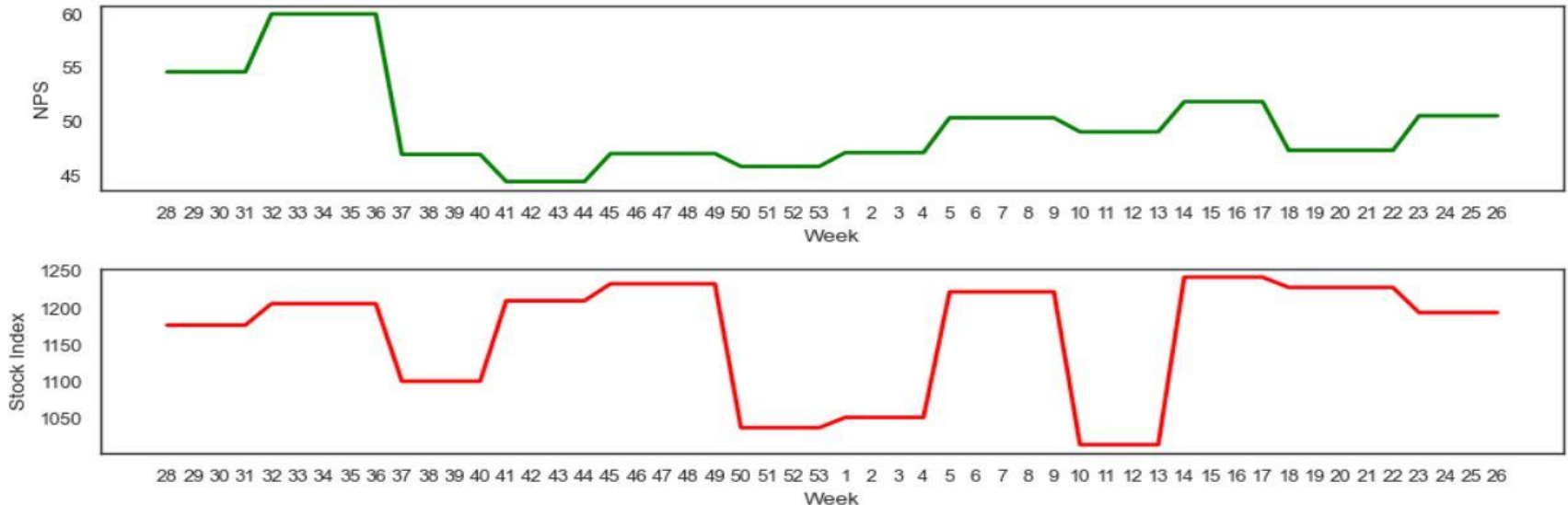
Impact of Holiday on Revenue
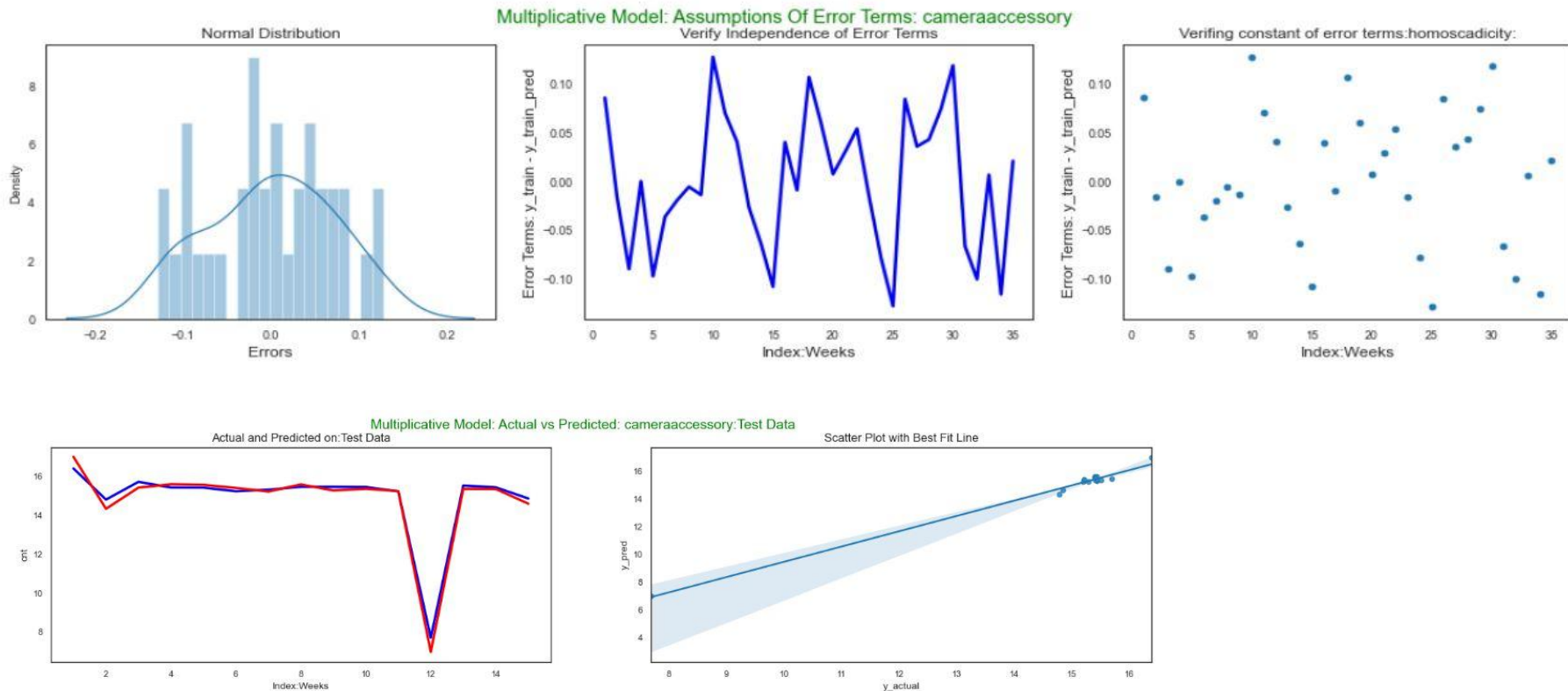
Presence of a Holiday does not impact revenue much!
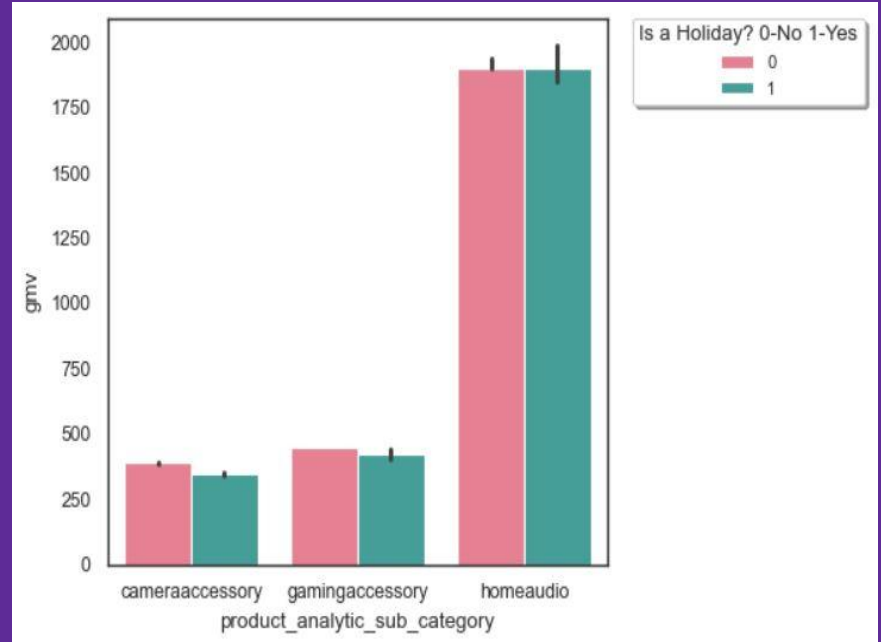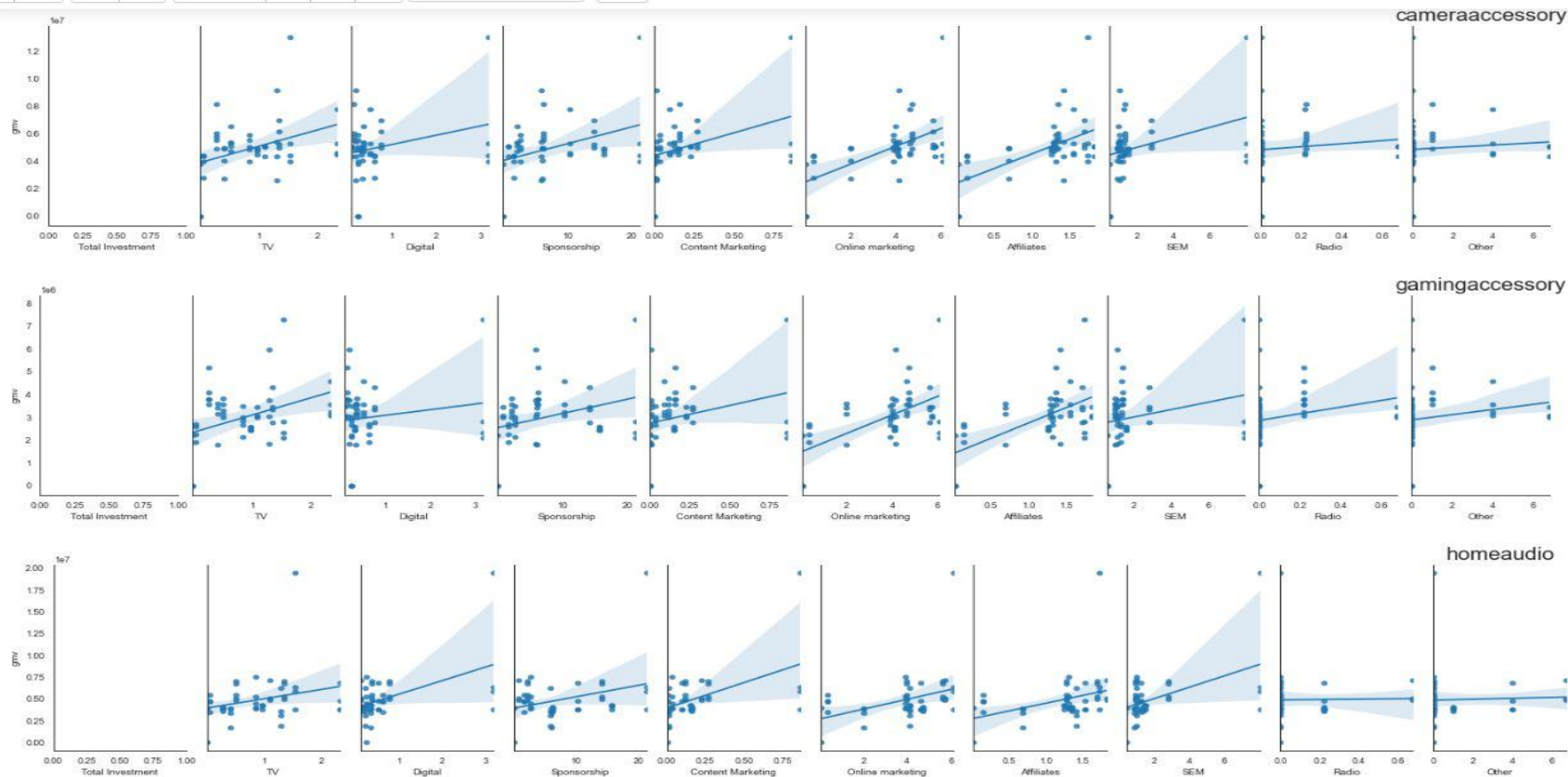
Insights: - **'Affiliates'** and '**Online Marketing**' seem to have relatively positive correlation with GMV(Revenue).
- **'Radio'** Seems to have least correlation with GMV(Revenue)
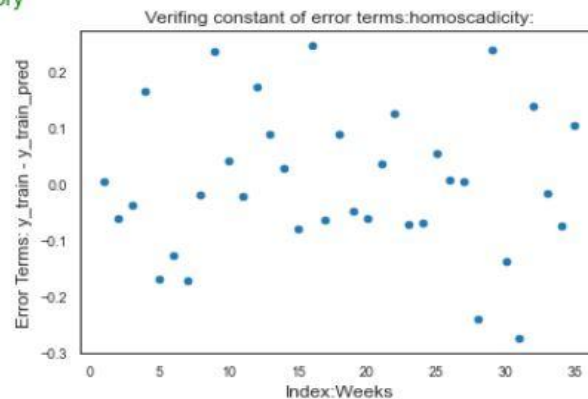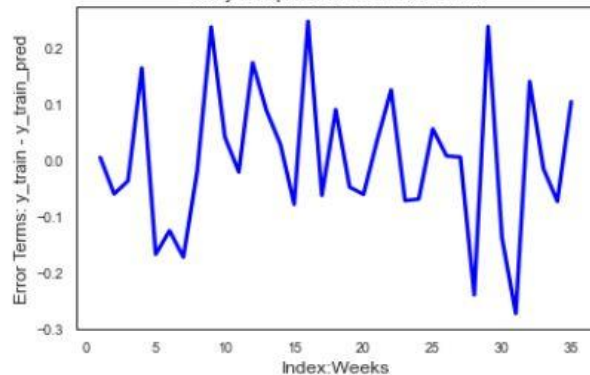
Model Validation:Visualizations
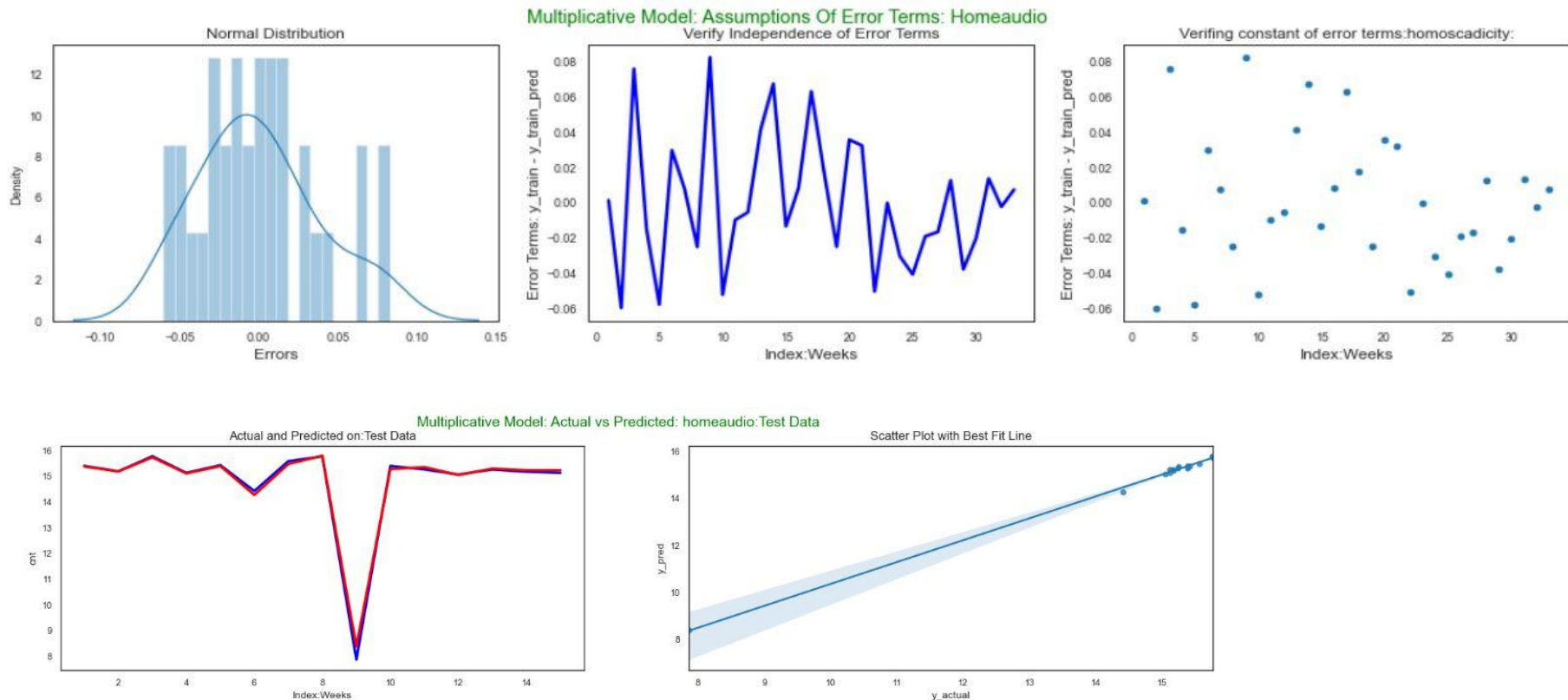
Error Terms & Actual vs Predicted Plots: Camera Accessory

Error Terms & Actual vs Predicted Plots: Gaming Accessory
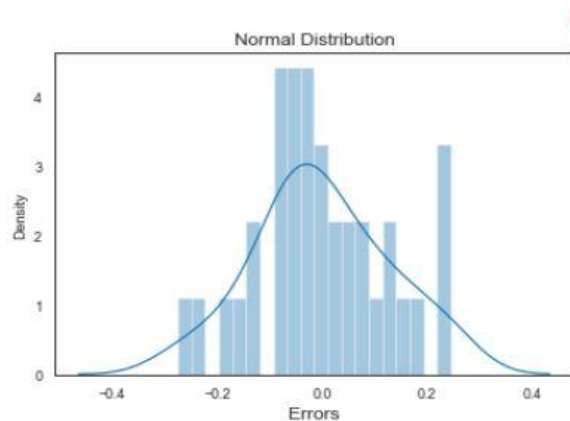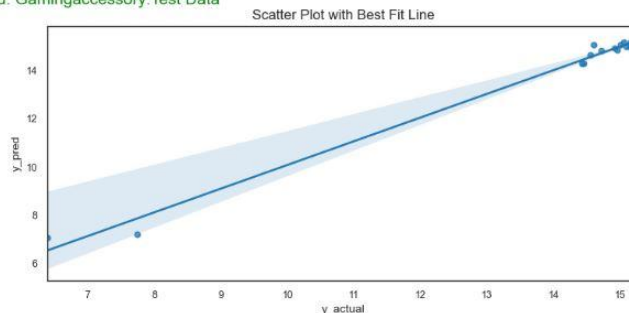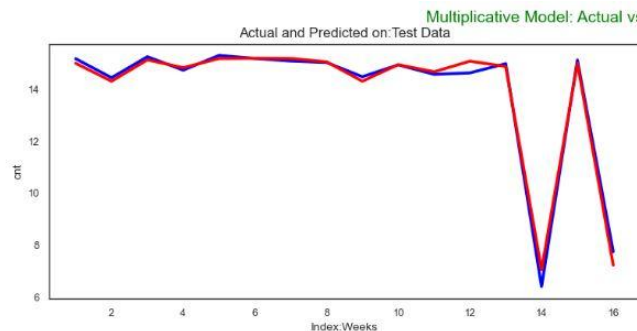
Error Terms & Actual vs Predicted Plots: Home Audio

Thank You!

Presenter: Achal Kagwad
Phone No: +91-9108302714
Email: achal.kagwad@gmail.com
Linked In:
https://www.linkedin.com/in/achalkagwad/