Capstone Project Ecommerce - Eleckart

Submitted by –

Sargam Agarwal, Kruti Patil, Krishna Mohan, Bhaskar Ghosh

Problem Statement

As a Data Scientist or an Analyst, working for ElecKart, we need to develop a market mix model based on the given information and we need to observe the actual impact of different marketing variables over the last year and recommend the optimal budget allocation for different marketing levers for the next year.

Objective

We are required to create market mix models for the three different product sub-categories:

- camera accessory
- home audio
- gaming accessories

To ensure that the budget is used effectively, we must monitor the actual effects of various marketing variables for the year (2015–2016) and make recommendations for the best budget allocation for the various marketing levers for the next year.

Business Objective

Business Understanding and Data Cleaning

Exploratory Data Analysis

Feature Engineering

Model Building

Deriving insights based on the results

Recommendations to the business

Summary of Business Understanding

We have consumer data that includes details of each purchase, user zip code, and order ID.

Media investment data provides information on spending across various channels such as TV and radio.

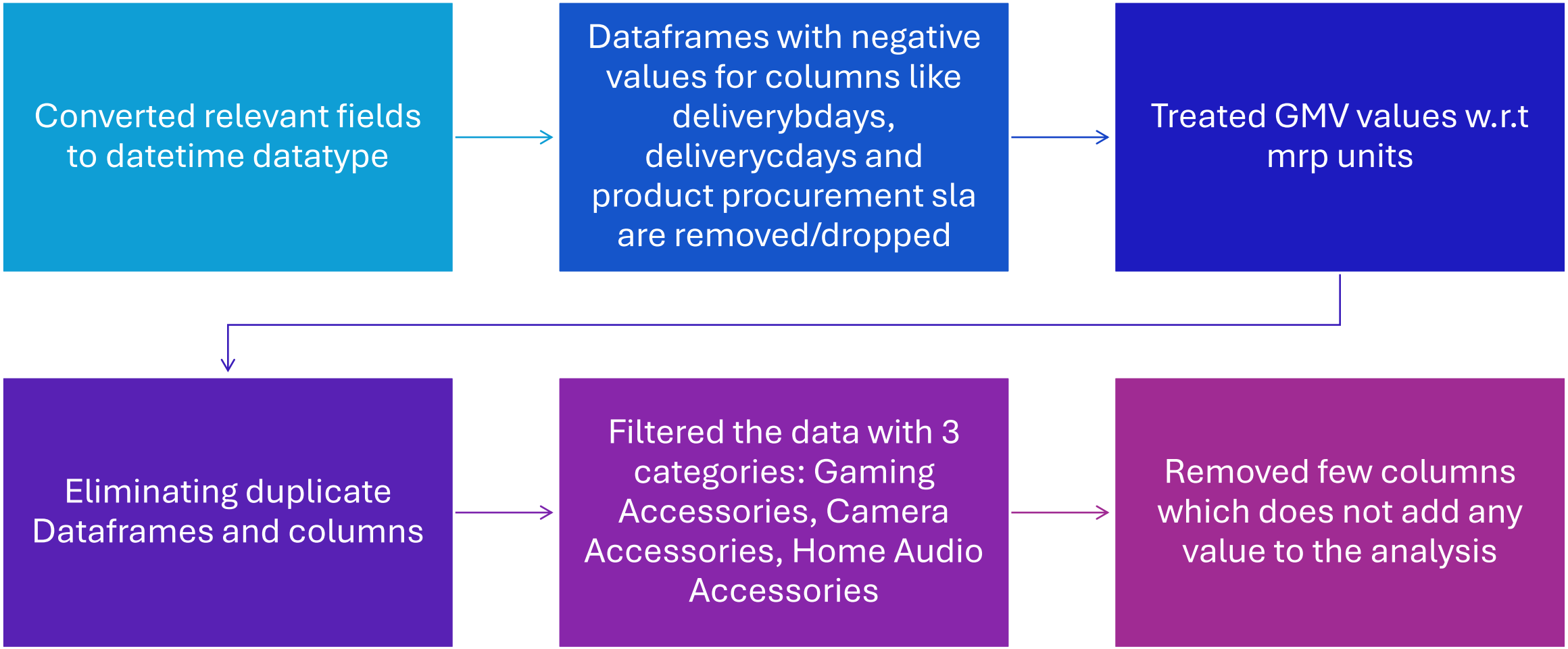
Climate data for the Ontario region is available for the years 2015 and 2016.

NPS score data offers insights into customer satisfaction, which can impact stock market prices.

A list of local holidays is provided, with the first and fifteenth of each month designated as pay dates.

Data Cleaning Steps

The sole purpose of data understanding and data cleaning is to deal with inconsistencies in the dataset such as missing values, exponential values present in certain fields etc. We also convert the various datatypes to a common datatype which will help us in analysis.



KPIs and Expected Results

Gross Merchandise Value (GMV), also known as revenue, is our target variable (independent variable).

We need to identify characteristics that lead to an increase in GMV.

Dependent variables directly affecting GMV include the amount and percentage of discounts offered on products, NPS rating, etc.

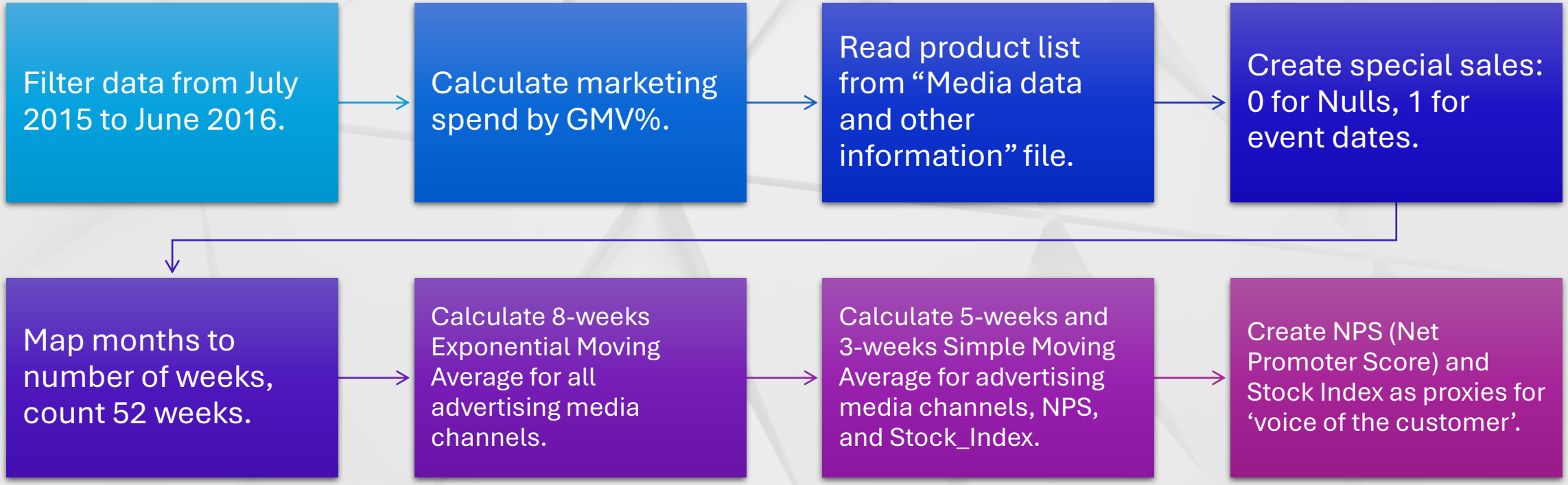
Our codes indicate that NPS score, holidays, and paydates (first and fifteenth of each month) have a positive impact on sales.

We need to determine which of the three subcategories—gaming, audio, and camera—the company should invest in to boost revenue in the next financial year.

Steps of Feature Engineering

- Create a Week column.
- Update year for week 53 to 2015, despite it being 2016 in other datasets.
- Drop columns with single value or all different values.
- Create a feature: GMV divided by units to get the list price.
- Create Payday column: flag as 1 if near salary day (1st and 15th), else 0.
- Create holiday_flag: flag as 1 if holiday or occasion in Ontario, else 0.
- Create Product Type column: flag as Premium_product if GMV > 80%, else mass_market.
- Create dummy variables for order_payment_type.

Feature Engineering



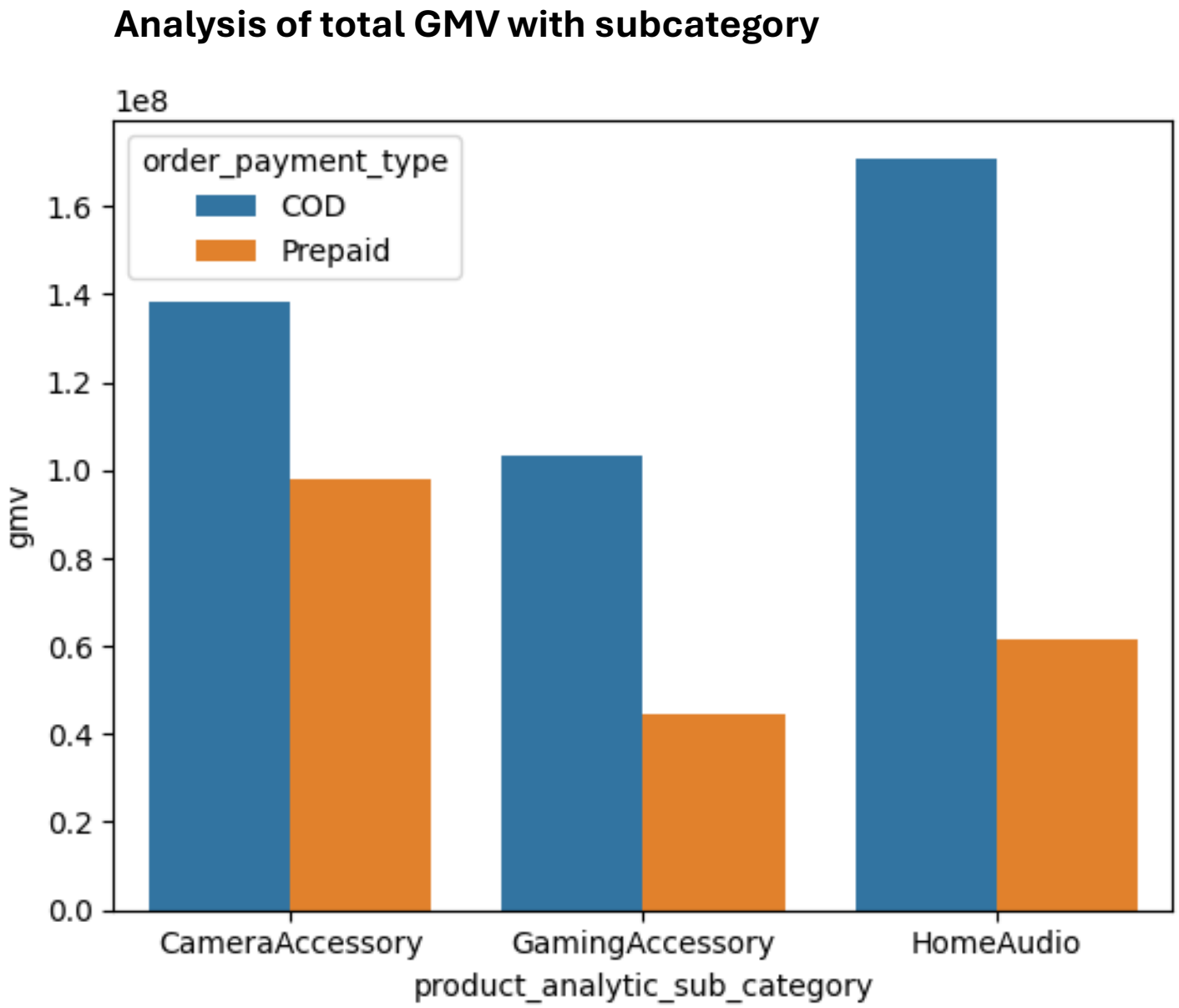
Steps and Visualization

Data Understanding

	DESCRIPTION
FSN ID	The unique identification of each SKU
Order Date	Date on which the order was placed
Order ID	The unique identification number of each order
Order item ID	Different products under the same order generates different order Item IDs
GMV	Gross Merchandise Value or Revenue
Units	Number of units of the specific product sold
Order payment type	How the order was paid
SLA	Number of days it typically takes to deliver the product
Cust id	Unique identification of a customer
Product MRP	Maximum retail price of the product
Product procurement SLA	Time typically taken to procure the product
Additional Information	
Monthly spend on various advertising channels	
Days when there was any special sale on products	
Monthly NPS score (this may work as a proxy to the ‘voice of the customer’)	
Stock index of the company on a monthly basis	
Climatic information of Ontario during 2015 and 2016	

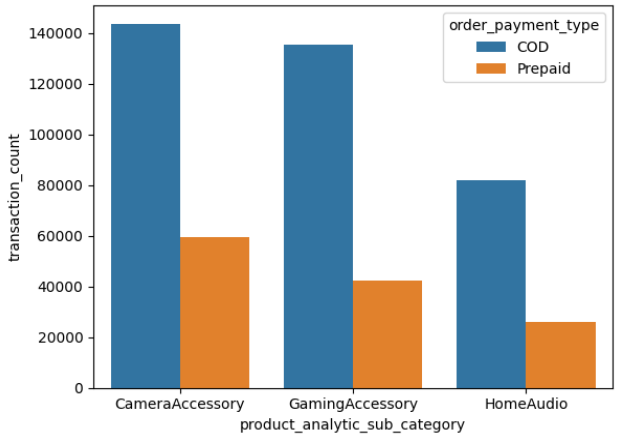
Exploratory Data Analysis

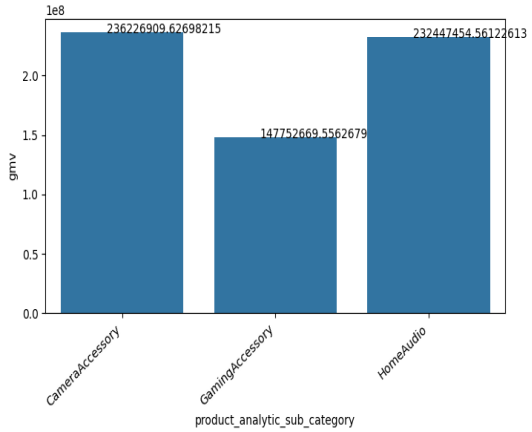
- Max revenue for COD order is from the class Home Audio followed by Camera
- For prepaid orders, the maximum revenue is from Camera Accessory, followed by Home Audio and a slight
- decrease in the category of Gaming Accessories



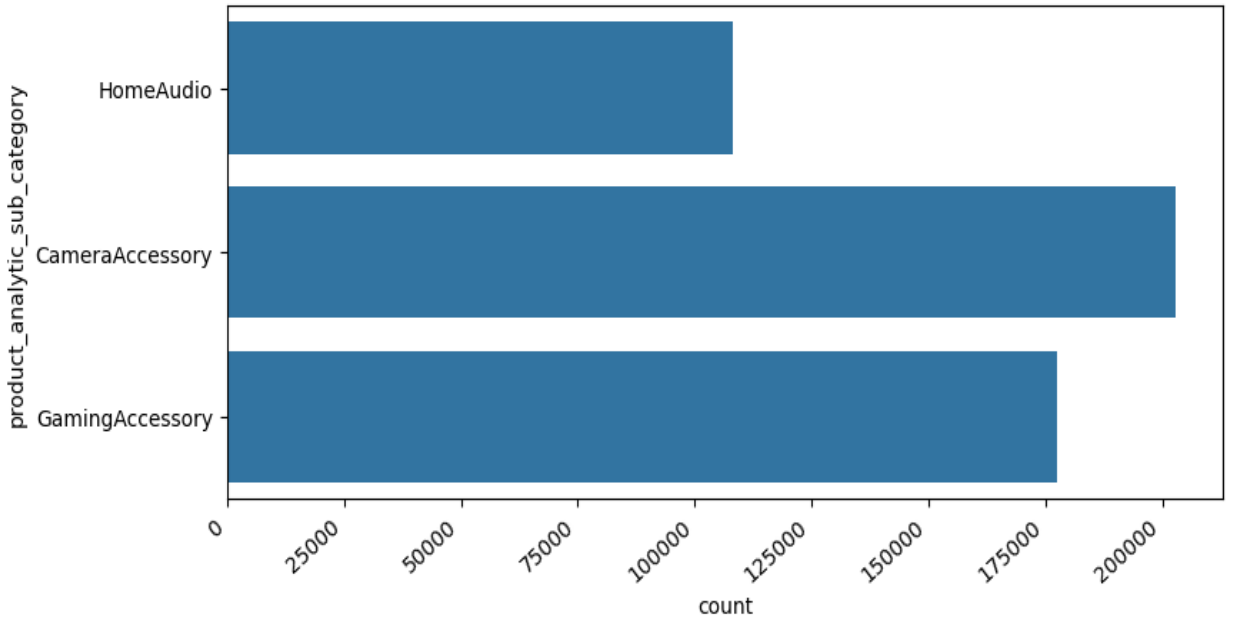
Exploratory Data Analysis

According to analysis, maximum number of transactions were from Camera Accessory category followed by Gaming Accessory category and Home Audio category

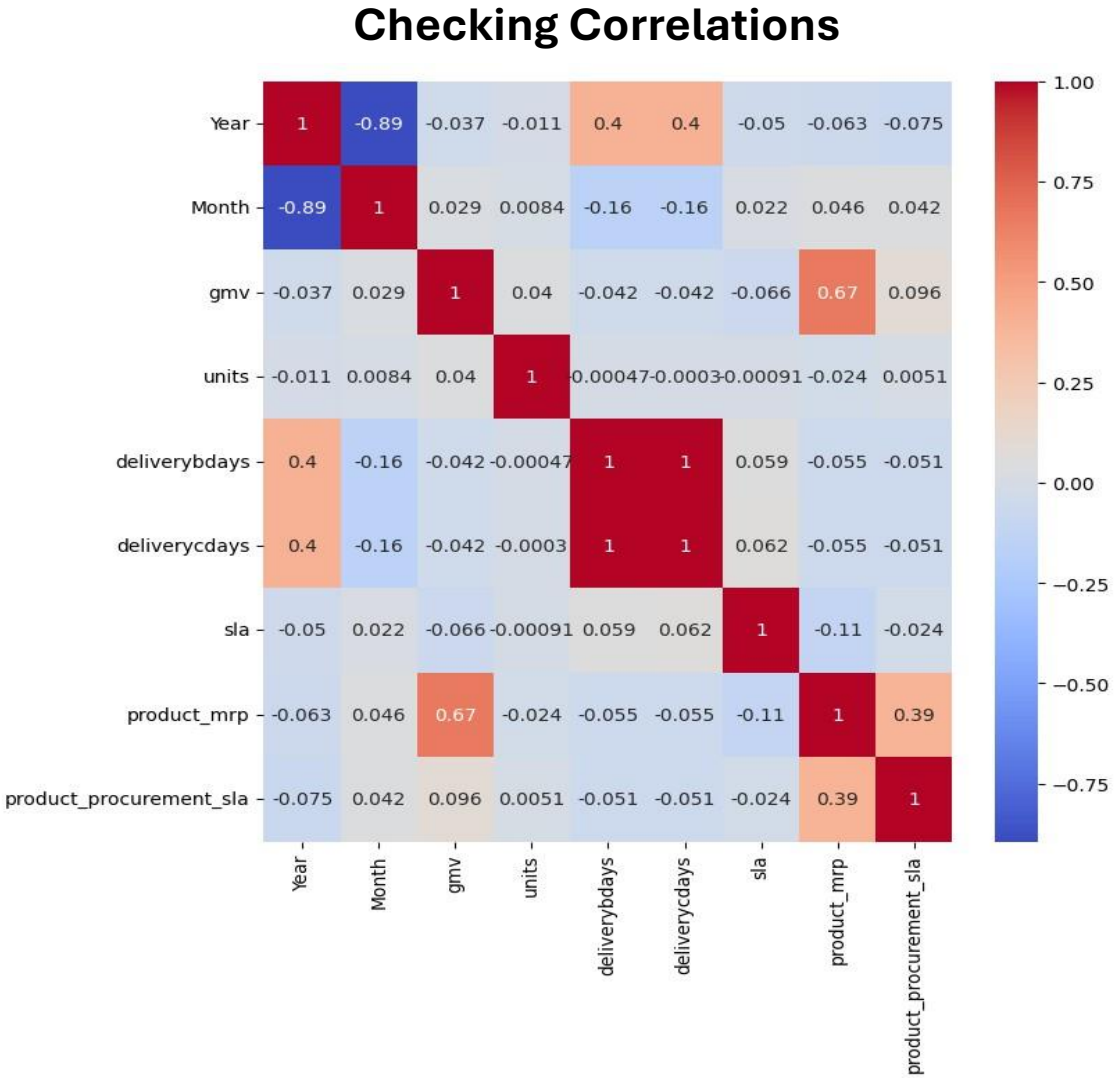




Analysis of no. of transactions based on subcategory



Exploratory Data Analysis

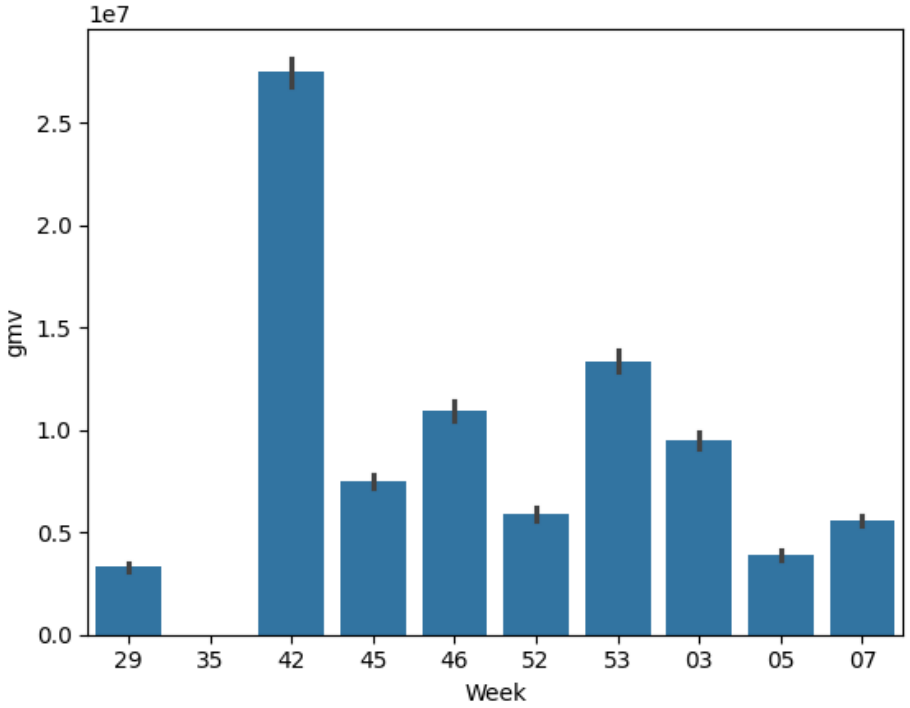


CORRELATION MATRIX : Checking the correlation Between the variables in the data set

Here we can observe that Month & Year are highly correlated.

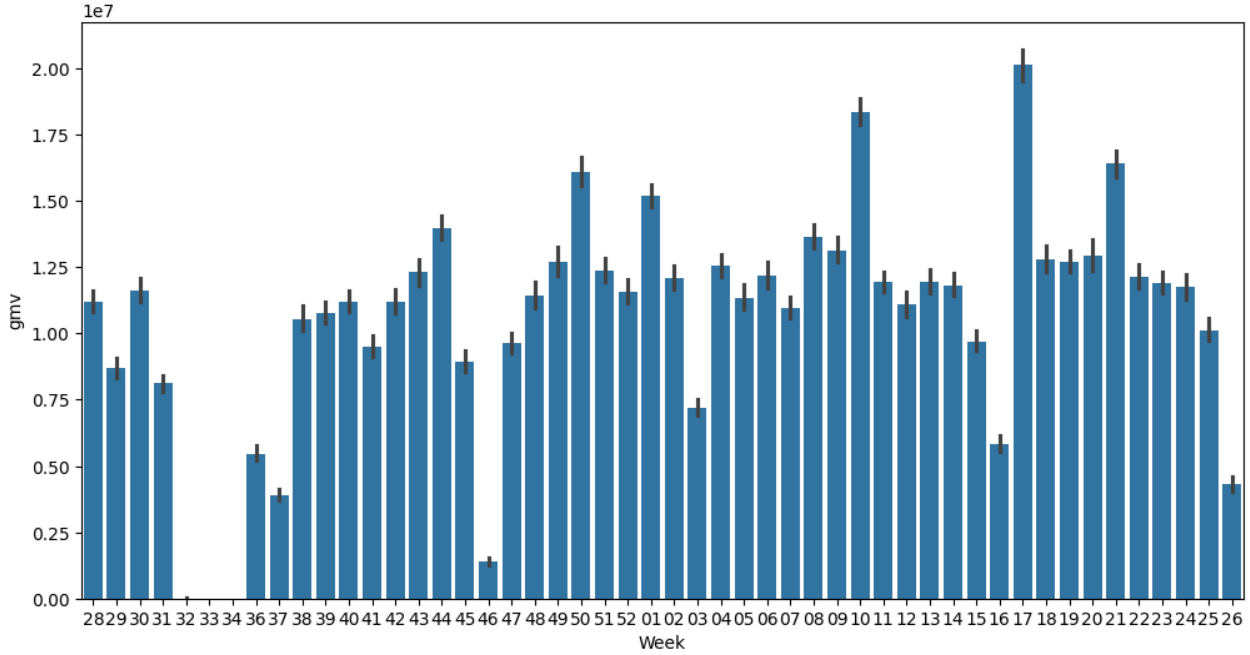
Exploratory Data Analysis

Analysis of Sales on Special Day



Plot for special sales flag = 1 (Special Sale)

Insights: The special sale day's week 42 biggest sales were noted.

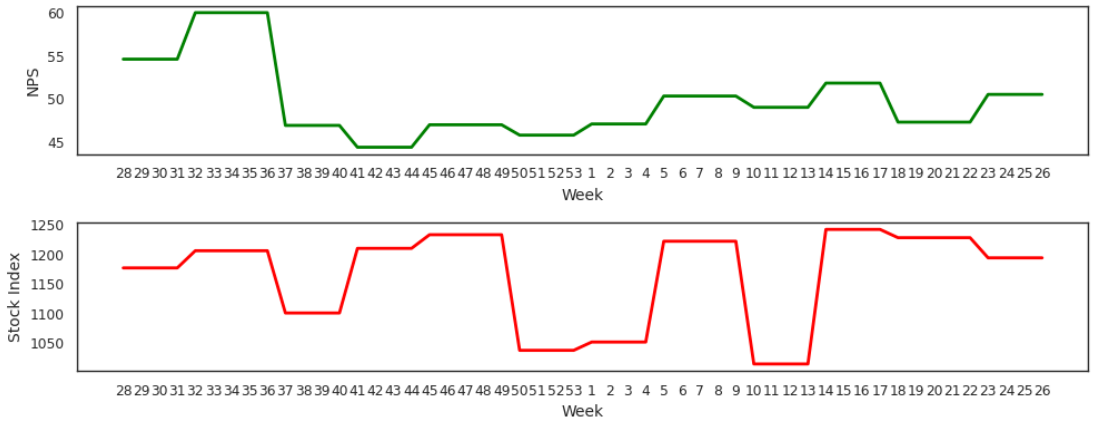


Plot for special sales flag = 0 (No Special Sale)

Insights: The special sale day's week 17 has the highest observed sales of 0

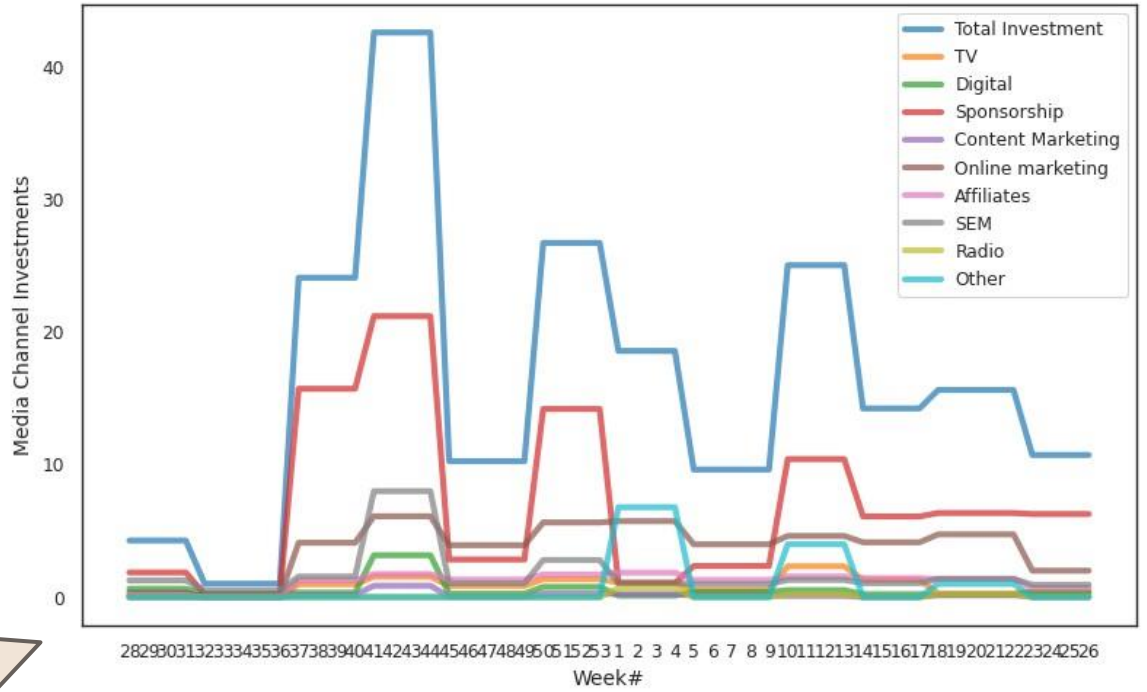
Exploratory Data Analysis

Displaying Trends



NPS and Stock Index by Week

Consumer NPS Score is highest in weeks 32 – 35, which coincides with the time when maximum discounts were being offered

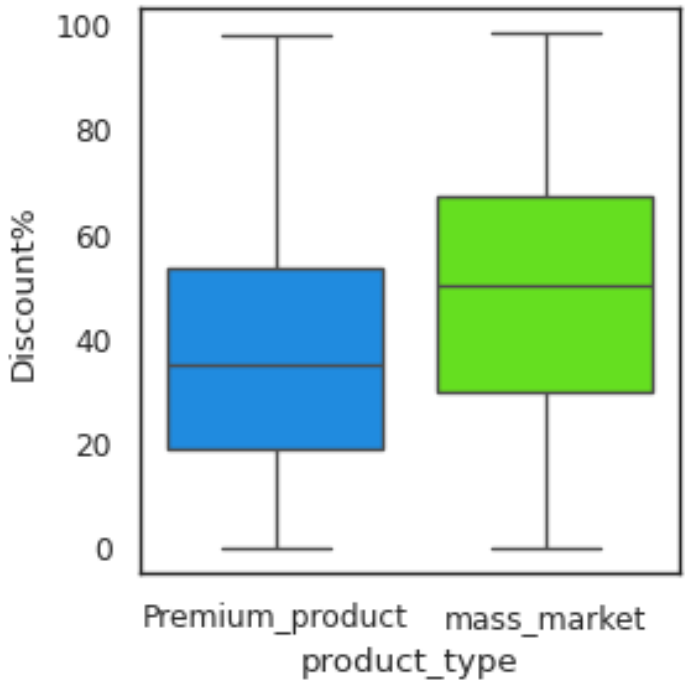


Various Media Channel Investments by Week

Over the past year, most Ad Investments have gone to sponsorships, followed by Online and Search Engine Marketing, especially during Thanksgiving

Exploratory Data Analysis

Comparing Distribution of Discount% for Product Types – Premium Products and Mass Market

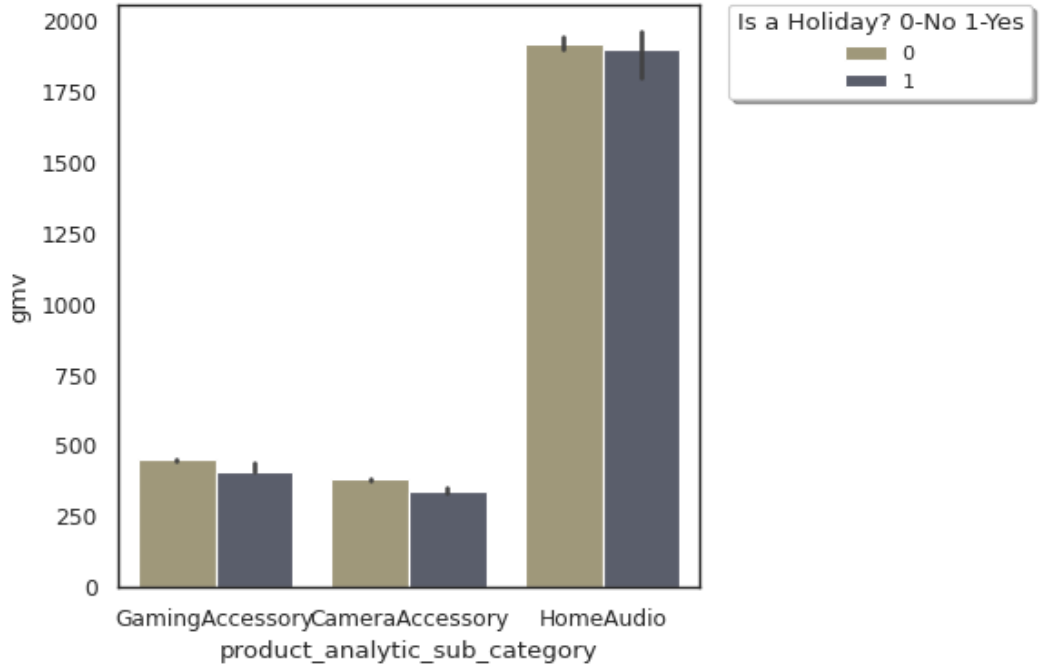


Median Discount% offered

Luxury Items < Mass Market Products

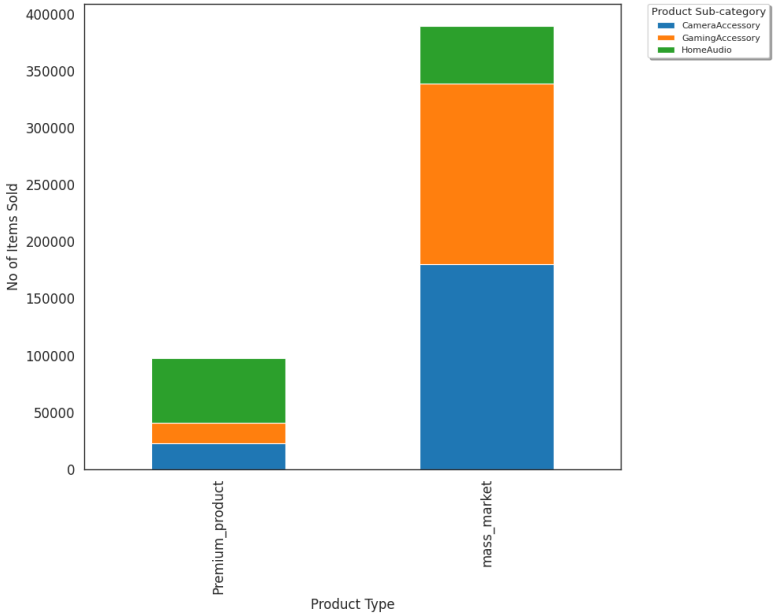
*This is a known trend among luxury products or luxury brands to offer limited or no discounts to retain the exclusivity of their products.

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

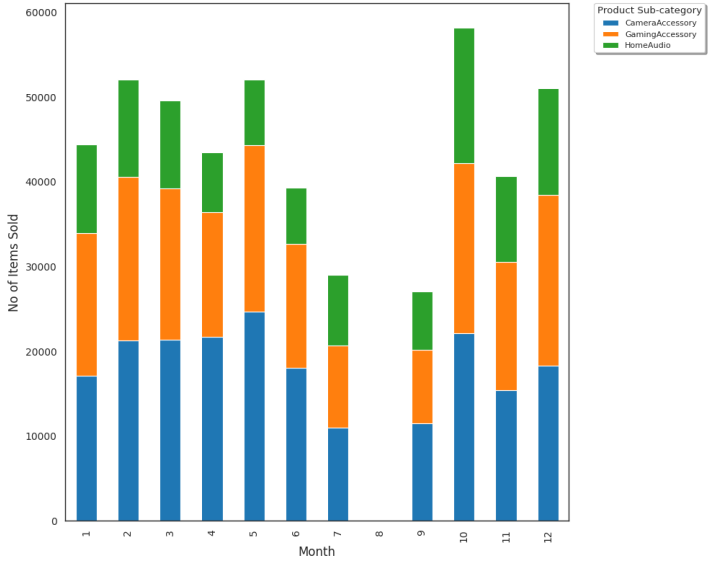


Average(Median) Revenue for 3 product subcategories from holiday and non-holiday days are more or less comparable

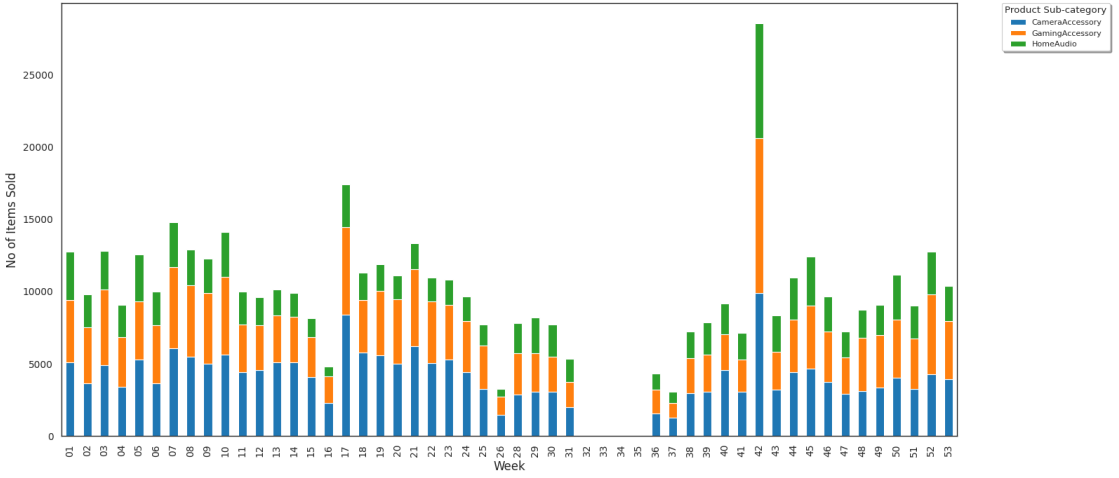
Exploratory Data Analysis



No. of items(Premium Product/Mass Market) sold per 3 product subcategories



Total items sold per 3 product subcategories per month

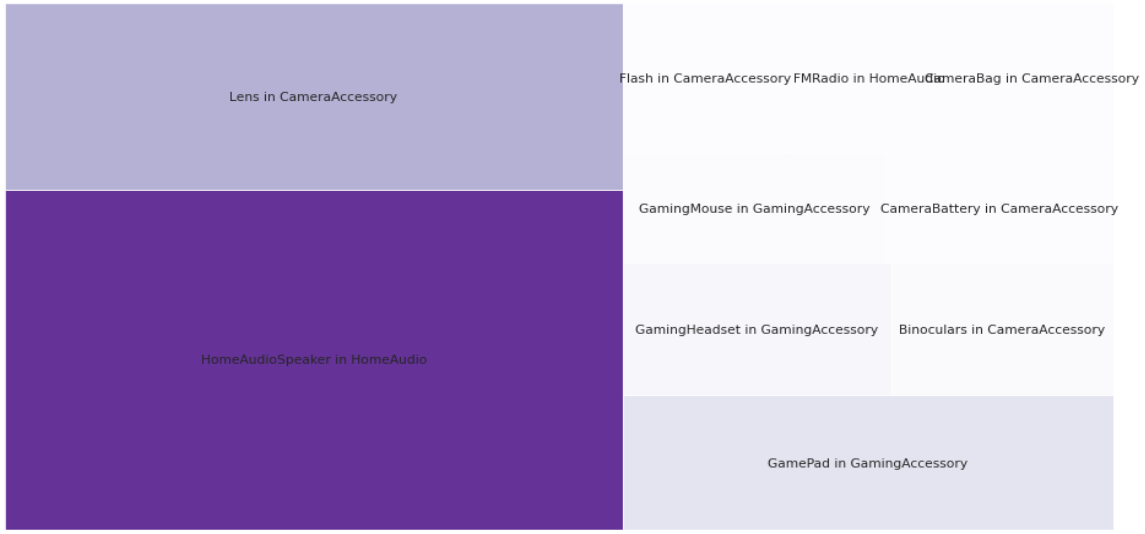


Total items sold per 3 product subcategories per week

Sale on 42nd week is maximum. Overall, October has seen most no. of items being sold.

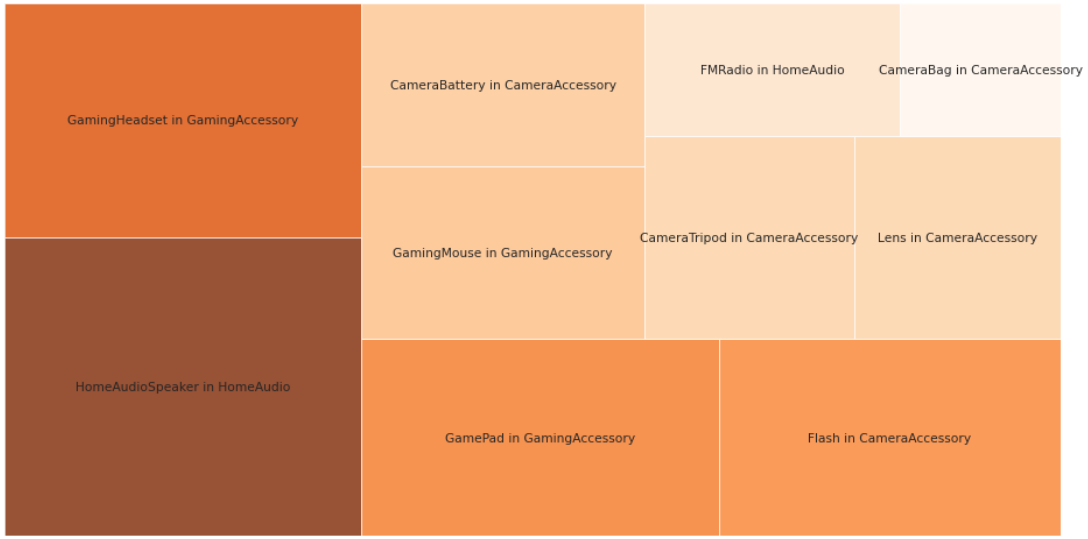
Exploratory Data Analysis

Top 10 Product Verticals which brought the Maximum Revenue for 3 Product subcategories



Home audio speaker in home audio brought the largest revenue followed by lens in camera accessory & gamepad in gaming accessory

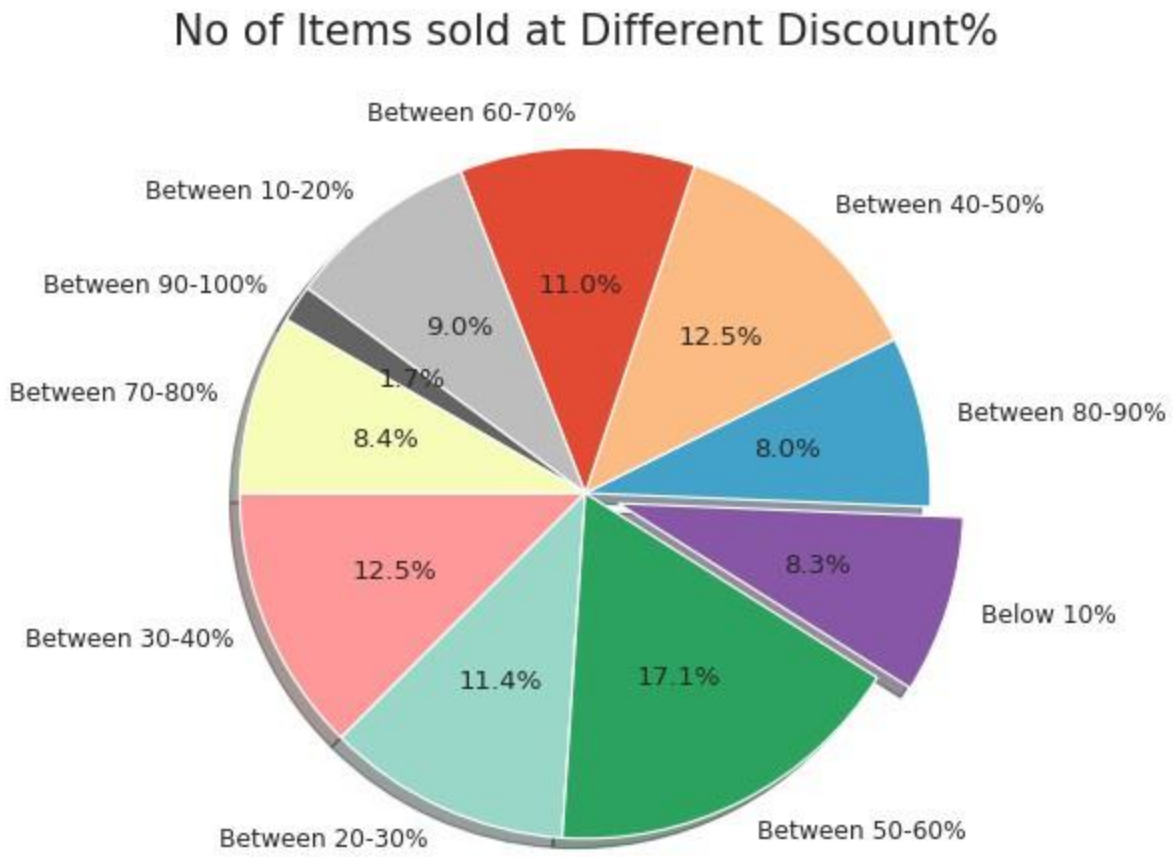
Top 10 Product Verticals with most no. of sales for 3 Product subcategories



Home audio speaker in home audio had the most no. of sales followed by gaming headset & gamepad in gaming accessory

Exploratory Data Analysis

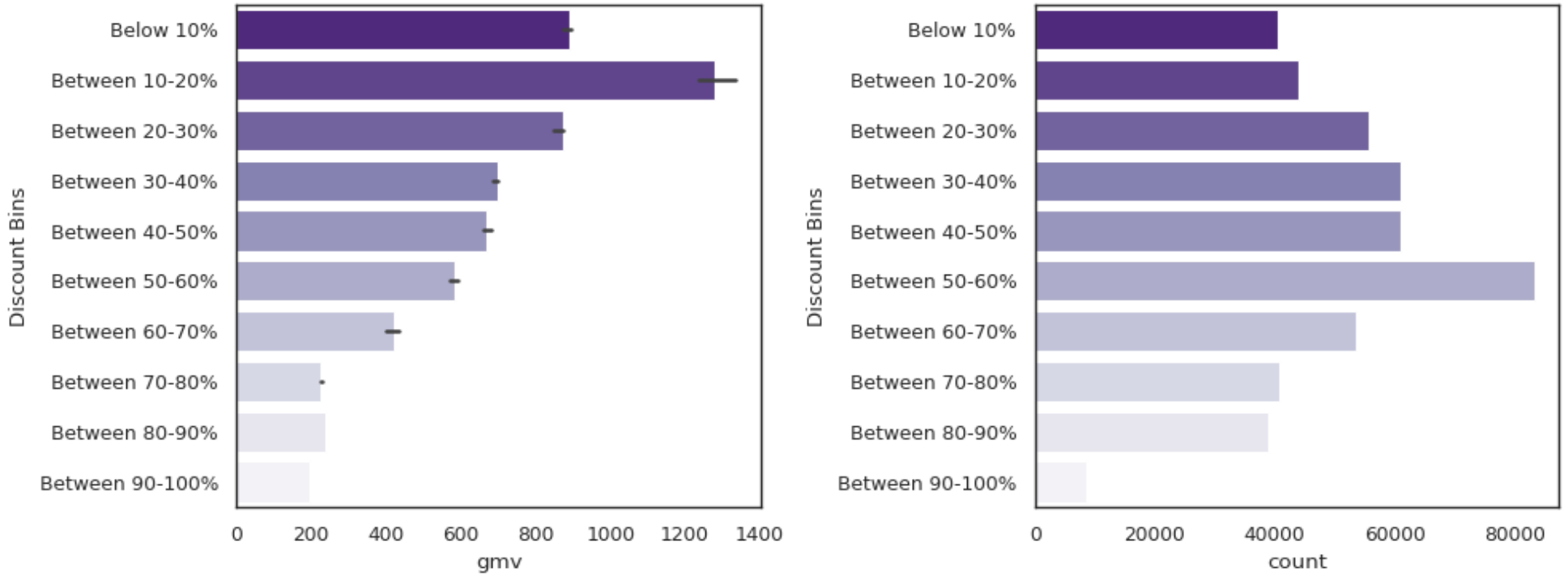
Percentage of items sold at different Discount% segments



Most Sales takes place when Discount% is between 50 – 60%

Exploratory Data Analysis

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



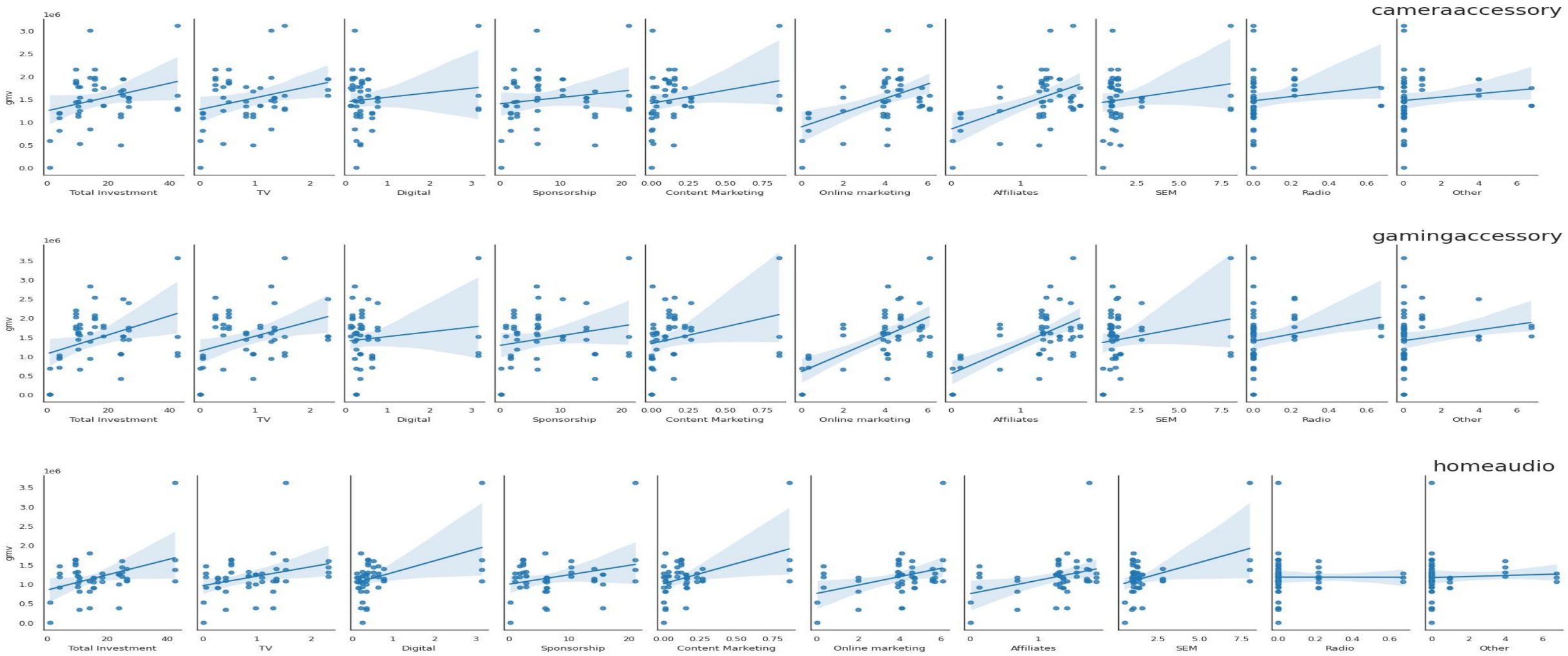
Median Revenue is maximum when Average discount% is between 10-20%. But beyond that, average revenue slowly starts to decline.

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.

Exploratory Data Analysis

Relationship between Revenue and Advertisement spends



Model Evaluation – Camera Accessories

Model	Top KPI	R-Square	Adj R-Square	MSE
Simple Linear Regression	sla, is_mass_market, special_sales, product_vertical_CameraBatteryCharger, product_vertical_FlashShoeAdapter, Sponsorship, Affiliates,Radio_SMA_3	0.962	0.949	0.00277
Multiplicative Model	Discount%, product_vertical_Filter, Affiliates_SMA_5	0.987	0.985	0.00336
Koyck Model	product_vertical_CameraBattery,product_vertical_CameraMicrophone, product_vertical_CameraTripod,product_vertical_Softbox, product_vertical_Telescope,Online marketing	0.859	0.827	0.010
Distributed Lag Additive	order_payment_type_Prepaid,product_vertical_CameraAccessory,product_vertical_CameraBatteryGrip,product_vertical_CameraTripod,product_vertical_Flash,product_vertical_Lens,Sponsorship_lag1,Cool Deg Days_lag1	0.958	0.944	0.00379
Distributed Lag Multiplicative	Discount%,product_vertical_CameraRemoteControl,Online marketing	0.964	0.960	0.00961

Model Evaluation – Gaming Accessories

Model	Top KPI	R-Square	Adj R-Square	MSE
Simple Linear Regression	product_vertical_GameControlMount, special_sales, product_vertical_GamingKeyboard,Content Marketing_SMA_5, Affiliates_SMA_5	0.773	0.734	0.0167
Multiplicative Model	sla, product_mrp, product_vertical_GamingAccessoryKit	0.971	0.968	0.00166
Koyck Model	product_vertical_GameControlMount,product_mrp,special_sales,TV_S MA_5, Content Marketing_SMA_5,Other_SMA_3,Total Snow (cm)	0.706	0.630	0.0225
Distributed Lag Additive	order_payment_type_Prepaid,product_mrp,special_sales,product_vertica l_GamingAccessoryKit	0.884	0.865	0.0075
Distributed Lag Multiplicative	product_mrp,product_vertical_GamingKeyboard,Affiliates	0.901	0.892	0.0078

Model Evaluation – Home Audio

Model	Top KPI	R- Square	Adj R-Square	MSE
Simple Linear Regression	sla, product_vertical_HomeAudioSpeake, product_vertical_VoiceRecorder, Digital, Sponsorship, Sponsorship_SMA_5, Radio_EMA_8	0.948	0.933	0.00535
Multiplicative Model	product_vertical_FMRadio,product_vertical_HomeAudioSpeake r,Radio_Ad_Stock	0.999	0.998	0.00035
Koyck Model	product_vertical_FMRadio,product_vertical_HomeAudioSpeake r,product_vertical_VoiceRecorder,Online marketing	0.988	0.987	0.00060
Distributed Lag Additive	sla_lag3,product_procurement_sla_lag1,product_vertical_FMRadio,product_vertical_HiFiSystem_lag1,product_vertical_HomeAudioSpeaker,product_vertical_VoiceR ecorder,NPS_SMA_5	0.990	0.987	0.00069
Distributed Lag Multiplicative	order_payment_type_Prepaid,product_mrp,Radio_lag3	0.899	0.888	0.2364

Recommendations & Results

Camera Accessory

- **Results**: Multiplicative Model leads with Adjusted R-square of 0.985 and MSE of 0.00336.
- **Recommendations:** Focus on Camera Filters, Accessories, and Flash via Affiliate marketing to boost visibility and sales.

Gaming Accessory

- **Results:** Multiplicative Model scores an R-square of 0.968 and MSE of 0.00166.
- **Recommendations:** Promote Gaming Accessory Kits with Prepaid orders and ensure SLA adherence for sustained growth.

Home Audio

- **Results:** Multiplicative Model dominates with R-square of 0.999 and MSE of 0.00035.
- **Recommendations:** Promote Home Audio Speakers & FM Radios to maximize revenue. Expanding the product lineup could further boost GMV.

Thank you