TEAM 2025108

INDEX

1. INTRODUCTION

1.1 Problem Statement:

ElectroMart, a high-growth e-commerce retailer, seeks to improve its marketing ROI through a data-driven strategy. Despite significant spending on advertising, digital marketing, and promotions, concerns remain about revenue growth contribution. By analyzing revenue drivers, measuring marketing ROI, and identifying consumption trends, ElectroMart addresses key challenges like inefficient budget allocation and ineffective promotions, as well as the absence of integrated insights on external drivers such as seasonality, holidays, and economic conditions. The initiative employs predictive modeling, KPI analysis, and dashboard visualizations to optimize resource allocation, enhance marketing efficiency, and foster sustainable, data-driven growth

1.2 Data Description:

Our analysis required working on multiple datasets such as customer orders, media expenditure, promotional calendars, NPS ratings, and external data like weather and holidays to gauge marketing effectiveness. Media investment data monitored ad spending for ROI assessment, while customer orders indicated buyers' behavior. Promotional events were monitored against sales, and NPS ratings were used as a proxy for customer satisfaction. Weather and holiday data helped in analyzing external influences on sales. In addition, a dataset covering over 12+ months was created with primary key variables like ad spend, order value, and discounts. Statistical techniques assessed marketing impact, feeding predictive models for budget optimization and revenue forecasting.

1.3. Objectives:

1. To identify key performance indicators (KPIs) that drive revenue by analyzing customer data, marketing spending, and external factors like promotions and customer satisfaction.
2. To assess the impact of marketing investments using A/B testing, regression modeling, and ROI analysis to optimize advertising and promotional strategies.
3. To develop a data-driven budget allocation strategy by identifying high-impact marketing channels and leveraging predictive analytics for optimization.

2. OBJECTIVE 1: PERFORMANCE DRIVER ANALYSIS

2.1. Identification of Key Performance Indicators (KPIs)

2.1.1. Data Preprocessing

Common preprocessing that took place across all datasets includes Date & Time Processing in which we converted ‘order_date’ and ‘sla’ to datetime format and filtered data for the period July’23 to June’24. Furthermore, we dropped rows where product_mrp or gmv is 0 and dropped rows where gmv is NA

1. Customer Order:

Negative numbers in "pin codes" and "cust_id" were taken as their absolute values to preserve data accuracy. Missing values in columns like “Deliverybdays,” “Deliverycdays,” “SLA,” and “Payment Type” were located and handled, and the "\N" values were replaced with 0 for uniformity. Also, duplicate records combined with “Order ID and Order Item ID” were identified and removed to ensure data integrity.

2. Media data-Sale Calendar-NPS Scores:

Media Investment: Standardized date format of year and month to maintain consistency.

Sales Calendar: Picked out revenue data especially during sale seasons to determine the effect of promotions and discount campaigns.

Payday & Holiday Data: Confirmed payday dates (1st & 15th of every month) and substantiated public holidays to ensure accurate data. Corresponded revenue data between sheets to monitor purchasing around key dates precisely and consistently.

NPS & Stock Index Data: Concurrent Net Promoter Scores (NPS), stock index directions, and spending data on marketing. This conjunction gave uniformity for proper correlation analysis, which facilitated the determination of the impact of brand attitude, market trend, and investment in marketing on sales performance.

Canada Holiday Data Processing:
Pulled official public holidays and added a Holiday_Week column to mark orders placed in the week having same week number as any holiday. This examined how holidays affect consumer buying habits and revenue patterns.

Product Hierarchy and Revenue Analysis:
Structured product information to enable category-level revenue analysis, creating a distinct hierarchy of product performance. This gave more insights into category-wise sales trends, supporting strategic inventory management and marketing decisions.

3. Weather Data:

Provided legend was transformed into a dictionary. Columns with more than 380 null values were removed and rows containing null values were dropped. Missing values in temperature columns were imputed using a rolling average interpolation method. 'Heat Degree Day' and 'Cool Degree Day' were calculated as per their definitions. Weather data from the years 2023 and 2024 were concatenated.

2.1.2 Exploratory Data Analysis (EDA) and Feature Engineering

We performed EDA on all the datasheets given to refine our dataset for better analysis and predictions. The basic EDA which took place over all the datasheets:

Now, the Specific EDA that we performed over the specific datasheets are mentioned sheets:

1. Customer Order:

Summary statistics were obtained for numerical columns, and categorical variables such as order payment type were checked for distinct values. We checked the distribution of revenue (GMV), units sold, and payment modes (Prepaid vs. COD) as shown in Fig 2.

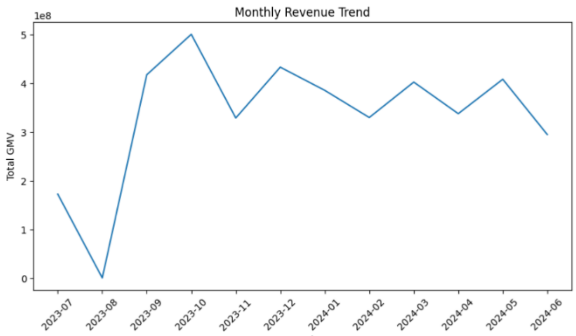


Fig 1.

(Trend Analysis)

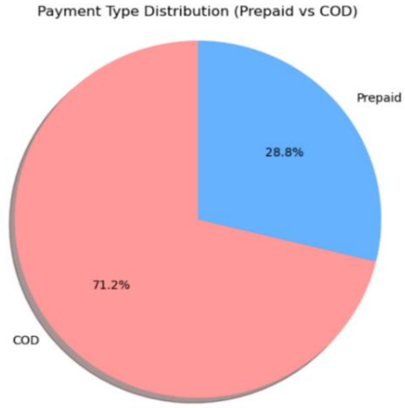


Fig 2.

Filtering & Outlier Handling:

Rows where "product_procurement_sla" and "SLA" were >= 0 alone were kept to maintain the validity of service-level agreement measures. Outliers in "total", "deliverycdays", and "deliverybdays" were detected to make the dataset more precise and refine delivery performance measures.

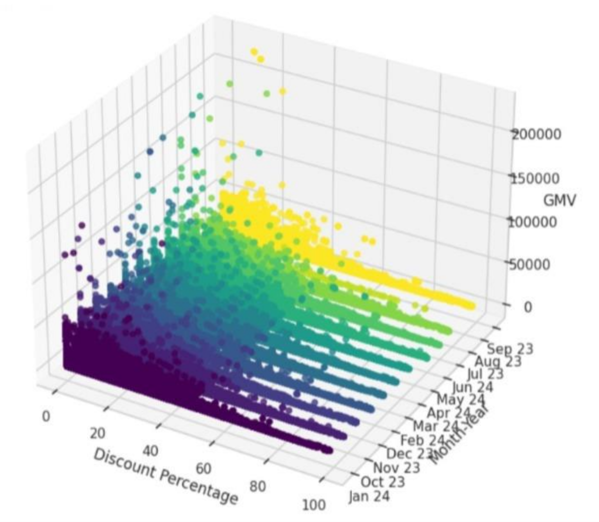
Feature Engineering which we performed on this sheet:

The formulas which are applicable are:

$$\text{Discount} = \frac{(\text{product\_mrp} - \text{list\_price})}{\text{product\_mrp}} \times 100$$

$$\text{Delivery Time (Total)} = \text{Deliverybdays} + \text{Deliverycdays}$$

The Product Type was categorized based on GMV, where products with a GMV above the 80th percentile were classified as Luxury, while all others were labeled as Mass-market. Customer Purchase Frequency was determined by calculating the total number of orders placed by each customer, providing insights into buying behavior.



The 3D scatter plot shows the relationship between Discount Percentage, GMV, and Month, where higher discounts generally lead to higher GMV. Notable spikes in June 2024 and July 2023 suggest the impact of seasonal or promotional sales.

2. Media Data - Sale Calendar - NPS Scores

Media Investment Analysis:

Radio and Other Investment missing values were replaced with 0. A column, "Total_Investment", was added as a sum of all media investments by month.

Revenue Integration & Categorization:

"Monthly revenue" was merged with "media investment" data. An investment was classified as "high" or "low" through a "Spending_Category* column, and "Marketing Efficiency Score" was calculated as:

$$Marketing\ Efficiency\ Score = \frac{Monthly\ Revenue - Total\ Investment}{Total\ Investment}$$

Special Sale Calendar Analysis:

Sale dates were dated, and "Start_Date" and "End_Date" columns were created. A function verified whether a date was within the sale period, and a "Sale_Period_Flag" indicated "sale days (1)" and "non-sale days (0)".

Sales & Discount Analysis:

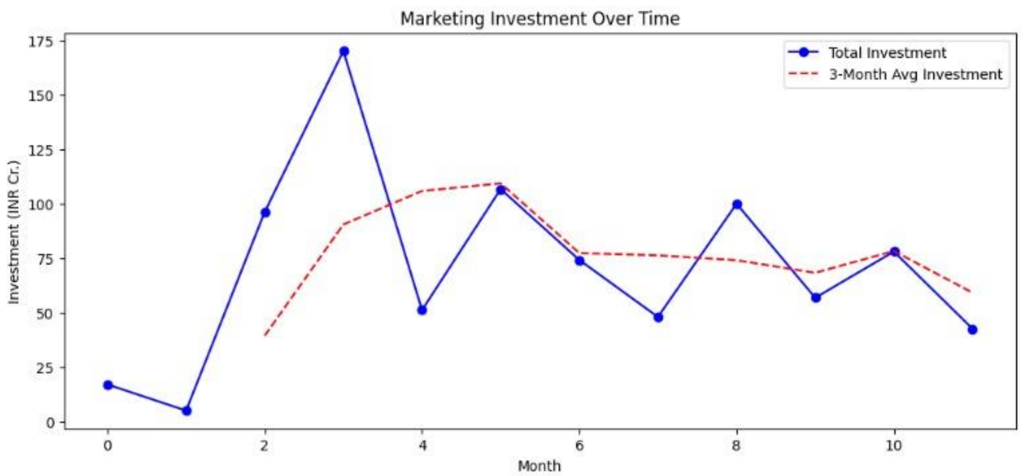
Sales were categorized by “Sale_Period_Flag” in order to compare sale vs. non-sale days. Discounts were explored in order to calculate Discount Boost, and Revenue Uplift was calculated as the difference in revenue between these periods.

Investment Lag & ROI Calculation:

Investment Lag was quantified through a 3-month rolling average. A linear regression model examined the correspondence between Monthly Revenue and Total Investment. ROI was calculated as:

$$\boldsymbol{ROI} = \frac{GMV - Total\ Investment}{Total\ Investment}$$

With all negative values throughout the months. The graph below depicts the same.



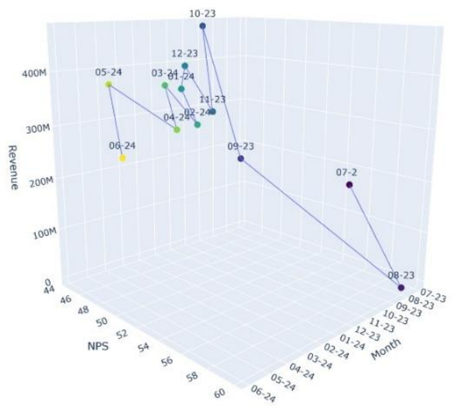


Fig 5.

This 3D scatter plot represents the relationship between NPS (Net Promoter Score), Revenue, and Month over time. Each point corresponds to a specific month, showing customer satisfaction and financial performance fluctuations. For instance, July 2023 (07-23) and August 2023 (08-23) show a decline in both revenue and NPS. May 2024 (05-24) and June 2024 (06-24) exhibit relatively higher revenue. The connected points highlight the trend, suggesting periods of both growth and decline.

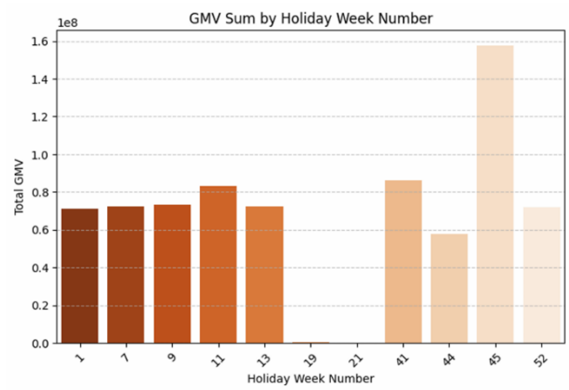
NPS:

NPS data was restructured into Month, NPS, and Stock_Index. NPS Growth Rate was calculated as:

$$NPS\ Growth\ Rate = \frac{Current\ Month\ NPS - Previous\ Month\ NPS}{Previous\ Month\ NPS} \times 100$$

Revenue_Change and NPS_Change measured percentage shifts, while Stock Price Volatility was calculated using a two-month rolling standard deviation of Stock_Index.

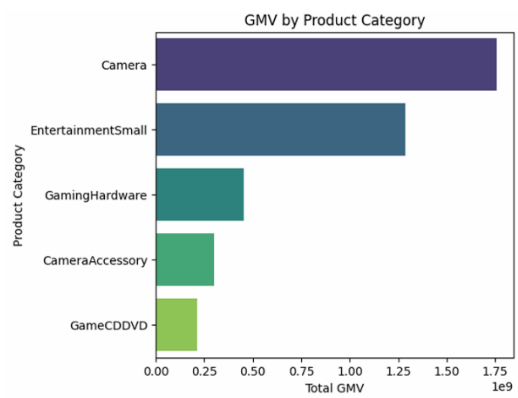
3. Holiday Sales Analysis:
A "Holiday_Week" column was introduced to the dataset, indicating orders in holiday weeks. Weeks 11, 41, 45, and 52 had greater GMV, with week 45 peaking because of significant sales events as in given figure.



Revenue & Sales Trends:
Total revenue, average order volume, and holiday vs. non-holiday sales were contrasted. Top-selling categories were determined in order to align marketing spend with holiday trends.

Marketing & Statistical Insights:
The effect of holiday-driven sales and additional marketing expenditures was determined. T-tests were used to evaluate sales uplift, and NPS scores were measured to determine customer sentiment.

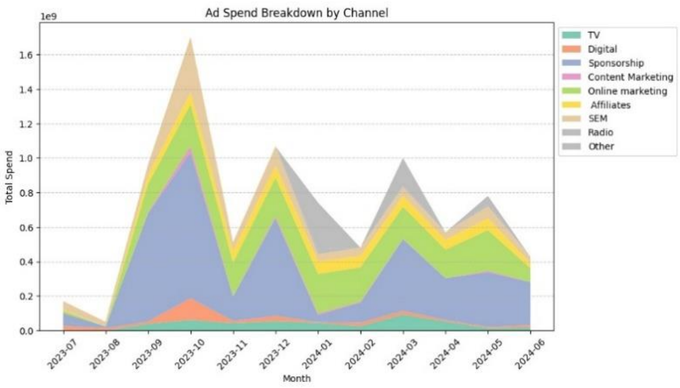
4. Product Hierarchy Analysis:
Organized the product hierarchy and reviewed category-wise revenue. Compared luxury vs. mass-market sales and evaluated the impact of discounts by categories.
Constructed a Product Category Performance Score as



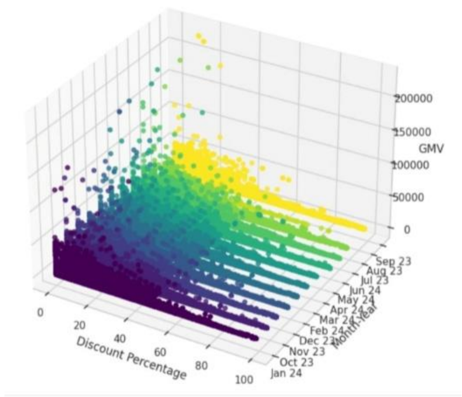
Product Category Performance Score $= \frac{Total\ Revenue}{Total\ Units\ Sold}$. Identified the Top 10% Best-Selling Products by revenue percentile and constructed a High-Discount Product Flag (1 if discount > 50%, else 0). It's depicted in the figure.

Here are some more important results of the EDA we performed:

The chart shows advertising spend by channel from July 2023 to June 2024, with peaks in October 2023 and January 2024, likely due to seasonal campaigns. TV and Digital dominate, while Online and Content Marketing also contribute. Smaller but steady allocations go to Sponsorships, SEM, and Affiliates. Spikes and declines suggest cyclical spending aligned with major sales events.



The 3D scatter plot shows the relationship between discount percentage, GMV, and time (Oct 2023–June 2024). Higher discounts generally increase GMV, but most data points remain in the lower GMV range, indicating diminishing returns. Seasonal effects and promotions influence the trend, highlighting the need for optimal discount strategies to maximize revenue.

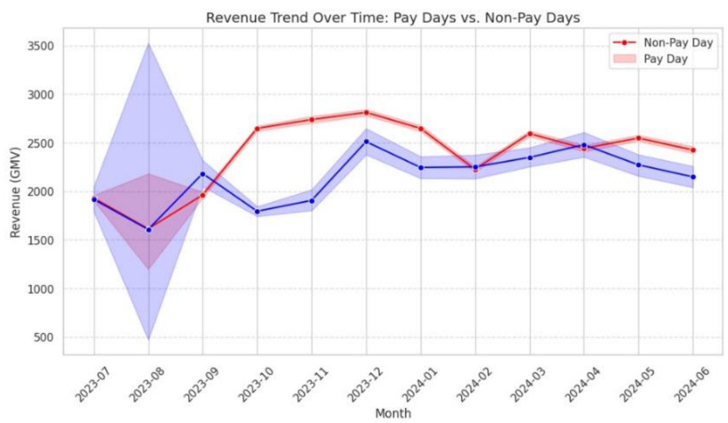


2.1.3. Correction Analysis Between KPIs and Revenue Plots

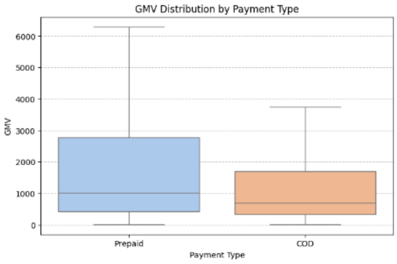
This is a correlation heatmap that visualizes the relationships between different numerical variables: GMV (Gross Merchandise Value), Units, Total Delivery Time, and Discount Percentage.

Observations: GMV has weak correlations: positive with Units (0.051), negative with Total Delivery Time (-0.019), and moderately negative with Discount Percentage (-0.2), suggesting higher discounts may reduce GMV. The color gradient indicates correlation strength.

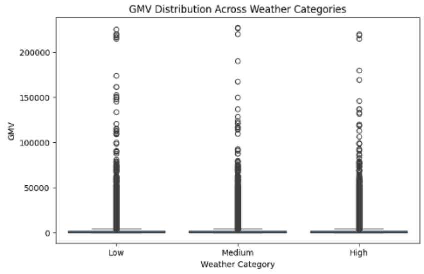
The line chart shows Pay Day vs. Non-Pay Day GMV trends (July 2023–June 2024). Non-Pay Days consistently outperform Pay Days, with a December 2023 spike likely due to holiday sales.



Observations: Darker shades indicate higher order volumes, while lighter shades signify lower activity.



In 2023, order volumes peaked on specific days, particularly in October, likely due to major sales events or promotions. Moderate activity is observed from July to December, while the first half of the year (January to June) shows little to no data. For 2024, moderate order volumes are present in the early months (January to April), with a significant peak in April, suggesting a special event or sales campaign. No data is visible for the latter half of the year, possibly due to incomplete data collection.



Order spikes in October 2023 and April 2024 suggest a connection to major sales events, holidays, or promotions. Comparing trends from both years can provide valuable insights for predicting future demand and refining sales strategies.

2.2. Impact of Weather on Revenue

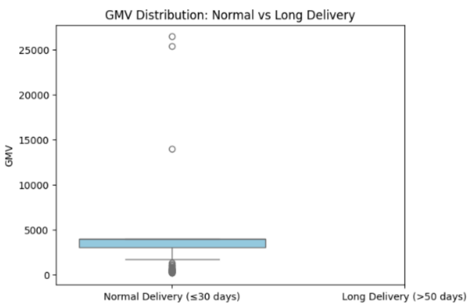
Hypothesis Testing:

Impact of Delivery Time on GMV:

Orders with long delivery times (>50 days) have a significantly higher GMV (₹3522.89) than normal delivery orders (₹2469.22, $p < 0.0001$). However, these long-delivery orders make up only 518 out of 14 lakhs, limiting their overall impact. The small to moderate effect size (Cohen's d = 0.254) suggests that while high-value orders tend to have longer delivery times, the influence on total GMV remains minimal.

Prepaid Orders Drive Higher GMV:

Prepaid orders generate significantly higher GMV than COD orders ($p < 0.0001$), despite COD being more frequent (1.03M vs. 396K). Higher spending in prepaid transactions suggests that targeting incentives for prepaid payments could boost revenue and improve operational efficiency.



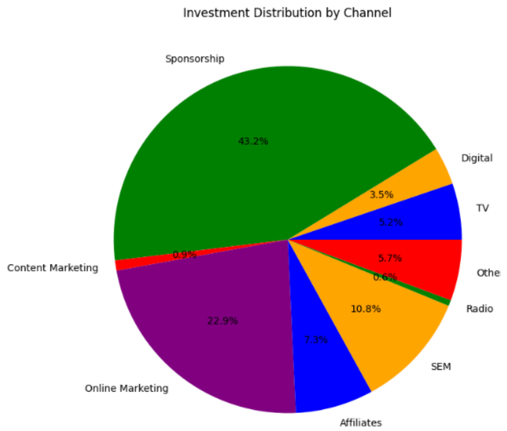
Extreme Weather Conditions Drive Higher GMV:

Extreme weather conditions significantly impact GMV ($p < 0.001$), with higher spending observed in both high and low temperatures (₹2592.60 & ₹2437.47) compared to moderate weather (₹2367.83). This suggests increased demand for specific products during extreme conditions.

Final Merge

The data was transformed into day-wise, merging customer, holiday, and weather information. "fsn_id" & "order_item_id" were assigned to each day's units sold, with "order payment types" (COD & Prepaid) and "product types" (Luxury & Mass-Market) being divided into unique columns. “product_mrp”and “list_price” were derived as the aggregate MRP of products sold per day. A “sale_period_flag” was assigned to 1 if a sale was made.

Time Series Modeling



Data Preparation:

GMV and revenue data were resampled daily, weekly, and monthly. The dataset was split 80-20 for training and testing, with weather-related features added as exogenous variables for better forecasting.

Forecasting Models for GMV:

Various models were used to predict GMV, evaluated using MSE, MAE, MAPE, and RMSE.

- LSTM: Captured temporal dependencies using all features, achieving a MAPE of 64.8% after hyperparameter tuning.
- Prophet: Modeled seasonality and holiday effects, integrating Canadian holidays. Optimized hyperparameters reduced MAPE to 30%, outperforming traditional models.
- ARIMA: Selected for its ability to capture temporal patterns. Manual tuning resulted in p=9, d=1, q=3, achieving a MAPE of 16.09%.
- SARIMA: Used to model seasonality with optimal parameters SARIMA (2,0,1)(0,1,1,12). Rolling forecasts were evaluated using MSE.
- Informer: Implemented Informer Transformer model with optimized hyperparameters using WandB. Outperformed ARIMA and SARIMA.
- Autoformer: Leveraged Auto-Correlation Transformers for long-term forecasting. Tuning improved accuracy, and Autoformer was used for final one-year predictions.

3. OBJECTIVE 2: IMPACT ANALYSIS ON MARKETING ROI

3.1. Attribution of Marketing Spend to Revenue

3.1.1. Media Spend Effectiveness Analysis

The analysis reveals that Sponsorships (43.2%), Content Marketing (22.9%), and Online Marketing (7.9%) received the highest investment, emphasizing a digital-focused strategy. Traditional channels like TV and Radio had minimal allocation. The distribution suggests a preference for content-driven and online marketing.

3.1.2. Evaluating the Performance of Different Advertising Channels

Categorization of Spending Groups:

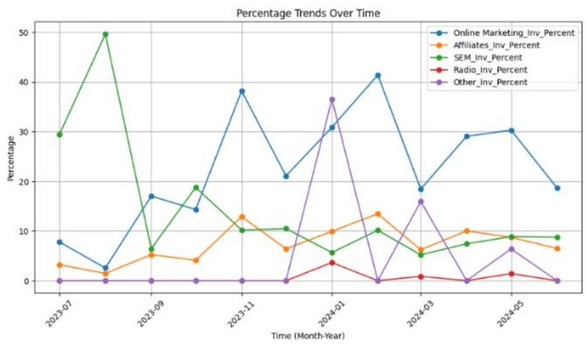
The dataset was segmented into high and low-spending groups based on the median total investment, revealing the impact of marketing spending on revenue.

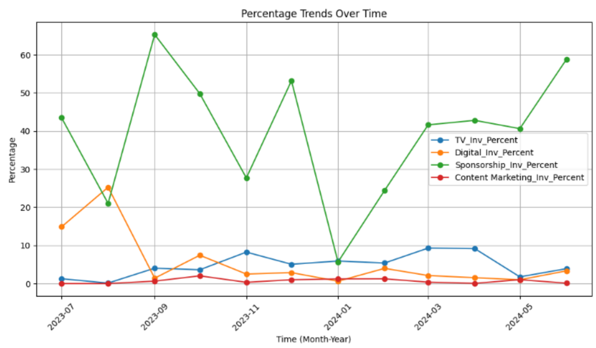
Revenue Comparison:

The analysis compares revenue between High and Low Spending groups. High Spending months show a higher mean revenue (36.26) vs. Low Spending months (22.67). The distribution also confirms that higher investments generally yield better returns, with significantly higher maximum and 75th percentile revenue in High Spending months.

Investment Percentage Calculation:

The dataset was filtered to include key media investment channels like TV, Digital, Sponsorship, Content Marketing, SEM, and Radio. The percentage contribution of each channel to the total investment was then calculated for analysis.





Higher spending consistently drives higher revenue, emphasizing the impact of marketing investments. Sponsorship and online marketing dominate but fluctuate, while digital and content marketing shows steady growth, signaling a shift to digital strategies.

3.1.3 Measuring the ROI of Promotional Campaigns

$$\text{ROI }\% = \frac{GMV - Total\ Investment}{Total\ Investment} \times 100$$

The ROI analysis compared total investment and monthly revenue, revealing consistently negative returns, indicating revenue fell short of investment across all months. July 2023 had the least negative ROI (-5.64%), while August 2023 performed the worst (-99.59%). High-investment months like October and December 2023 failed to yield proportional returns, suggesting that spending alone does not drive profitability. Meanwhile, months with lower losses highlight potential strategies for better efficiency.

4. OBJECTIVE 3: OPTIMIZING MARKETING SPENDING

4.1. Budget Reallocation Strategy

4.1.1 Identifying High-Impact Marketing Channels

Marketing channels and external factors were defined, with fixed mean values for external factors during optimization. Adstock transformation (decay = 0.3) was applied to model carryover effects. Revenue was modeled using PyMC, combining base revenue, channel contributions, and external effects. Priors: base revenue ~ Normal (40% GMV, 5% σ);

channel effects ~ Truncated Normal (0.4, 0.1, [0.1,1.0]); external effects ~ Normal (0, 0.01). Observed revenue ~ Normal (expected revenue, 10% σ of GMV). Bayesian inference used NUTS sampling with 3000 draws, 1000 tuning steps, and 4 chains.

4.1.2 Allocating Resources Based on Data-Driven Insights

Bayesian inference was used to estimate ROI for each marketing channel. Predicted revenue was derived from ROI estimates and initial spend, with a ±5% constraint imposed to ensure revenue stability during optimization. The objective function aimed to maximize profit while maintaining revenue within this constraint, applying penalties for any violations. Budget allocations were optimized using Bayesian optimization (gp_minimize), allowing spend adjustments within 80%–120% of the original allocation to achieve maximum profit without significant revenue fluctuation.

4.1.3 Scenario Planning and Simulation for Budget Optimization

Post-optimization analysis involved comparing optimized versus original spend, revenue, and profit to evaluate the effectiveness of reallocation. Budget shifts across individual channels were visualized to highlight changes in investment. The total optimized budget was then redistributed proportionally based on each channel's ROI, ensuring that channels with higher ROI received a more significant share of the budget.

ROI Comparison

Channel	Raw Budget (₹ Cr)	Raw ROI (%)	Optimized Budget (₹ Cr)	Optimized ROI (%)	ROI Improvement (%)
TV	44.37	-93.52	148.86	-65.89	27.62
Content Marketing	8.03	-99.06	155.66	-73.25	25.81
Affiliates	61.37	-93.35	161.8	-72.81	20.53
SEM	91.2	-90.3	131.4	-75.59	14.71
Radio	4.67	-99.35	45.46	-87.48	11.87
Other	48.02	-93.67	49.16	-86.53	7.14
Digital	29.66	-95.64	27.06	-91.39	4.25
Sponsorship	365.37	**-78.57**	124.56	-81.8	-3.23
Online Marketing	193.64	-84.41	2.37	-99.45	-15.04

Channel-wise Revenue Analysis:

Channel	Budget (₹ Cr)	Estimated Revenue (₹ Cr)
Online Marketing	193.64	61.45
SEM	91.2	27.73
Affiliates	61.37	18.61
Content Marketing	8.03	2.39
Sponsorship	365.37	105.82
Other	48.02	10.63
TV	44.37	7.99
Radio	4.67	0.83
Digital	29.66	5.223

4.2. Budget Reduction Strategy using Marketing Mix Modelling

The Bayesian MMM model estimates base revenue, channel effects, and external factor impact using probabilistic priors and MCMC (Markov Chain Monte Carlo) sampling. It applies adstock transformation to account for delayed marketing effects. The total predicted revenue is computed as the sum of base revenue, channel contributions, and external factor impact, enabling a detailed multi-channel ROI analysis and optimizing budget allocation.

The optimized budget is calculated using Bayesian Optimization, which iteratively adjusts the allocation across marketing channels to maximize predicted GMV. It normalizes the budget to ₹846.6 Cr, evaluates different allocations using a trained Random Forest model, and refines the best-performing ones.

The current year's budget allocation was determined by calculating each channel's contribution to total yearly revenue, using the formula:

$$\textbf{Channel} - \textbf{wise}\ \% = \frac{\textbf{channel revenue}}{\textbf{total revenue}} \times \textbf{100}$$

To allocate next year's budget, a **Bayesian Marketing Mix Model (MMM)** was implemented using PyMC. This model estimated each channel's influence on revenue through learned coefficients, which were used to predict future allocations. A constrained optimization problem was then solved using Differential Evolution to maximize total revenue. The optimization process was subject to budget constraints, and the objective function incorporated a penalty term to discourage violations. This ensured a balanced approach prioritizing revenue maximization while keeping budget allocations within acceptable limits, yielding an efficient and feasible future budget plan.

From these dashboards, we can state that GMV shows strong growth from 2023 to 2024, with luxury products dominating unit sales. Holiday events, particularly Boxing Day and

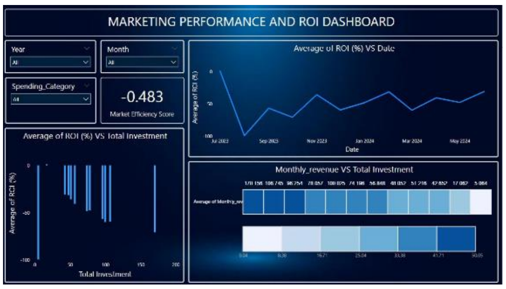
These are the final results:

Channel	Past budget allocation(Cr)	Future budget allocation(Cr)
Online Marketing	61.36	15.65
SEM	4.67	4.47
Affiliates	91.2	13.51
Content Marketing	193.64	39.35
Sponsorship	8.03	7.99
Other	353.59	105.44
TV	29.66	3.22
Radio	48.02	6.76
Digital	365.37	183.45

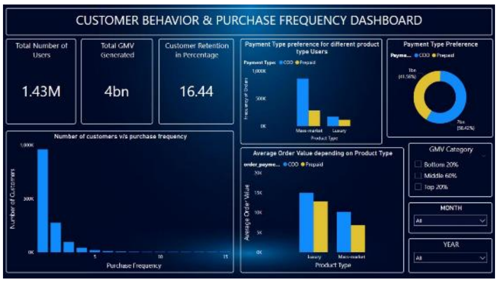
4.2 Implementation and Decision Support

4.2.1 Development of a Marketing Performance Dashboard

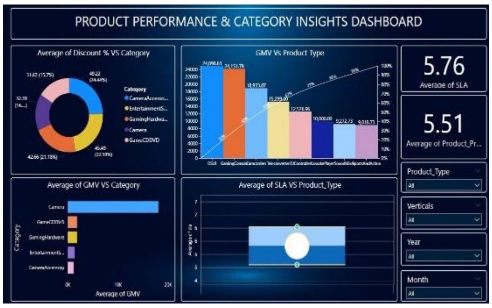
1.

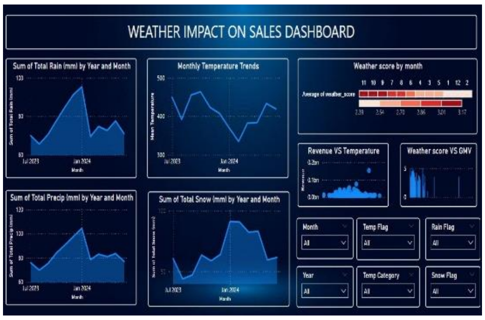


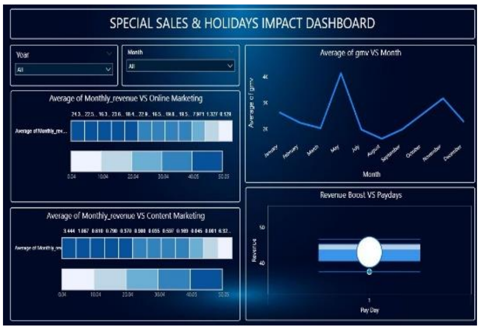
2.

3. 

4. 

5. 

6. 

7. 

1. Marketing Performance and ROI Dashboard
2. Sales and Revenue Dashboard
3. Customer Behaviour Dashboard
4. Discount and Pricing Impact Dashboard
5. Product performance and Category insights Dashboard
6. Weather impact Dashboard
7. Special Sale and Holiday Impact Dashboard

From these dashboards, we can state that GMV shows strong growth from 2023 to 2024, with luxury products dominating unit sales. Holiday events, particularly Boxing Day and Christmas, drive GMV spikes. May and April are high-selling months, contributing significantly to total units sold. Discounts average 43.97%, but high discounts don’t always drive GMV growth, requiring optimization for profitability. Marketing efficiency is inconsistent, with higher investments not always yielding better returns. Online marketing drives higher revenue than content marketing, and payday effects boost revenue. Total users stand at 1.43M, with 16.44% retention and 4bn GMV.

Most buyers are one-time purchasers, and COD (58.42%) is the preferred payment method, especially for mass-market users. GMV sees peaks in May, November, and December. Weather influences sales, with higher temperatures correlating positively with revenue. Among product categories, cameras lead in GMV, followed by GameCD/DVD. DSLRs and gaming consoles generate the highest GMV. Discounts vary across categories, some exceeding 45%. SLA remains stable at an average of 5.76 days.

5. OBJECTIVE 4: Product and Marketing Strategy for Upcoming Campaigns

5.1. Target Product Categories for Upcoming Campaigns and Their Rationale

Product Category Prioritization

- **EntertainmentSmall** has the highest **revenue and units sold**, making it the top priority for marketing.
- Campaigns should focus on this category for maximum impact.

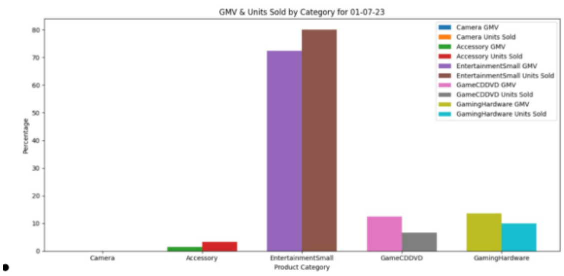
Visuals:

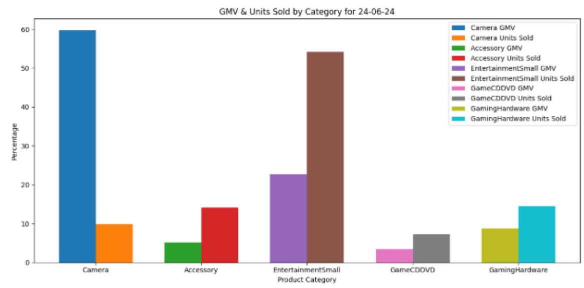
Campaign Topics for each holidays:

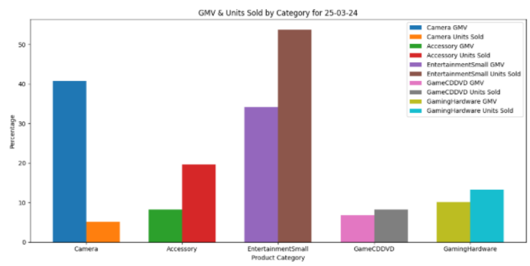
Holidays	Campaign Topics
New Year's Day,Valentine's Day St. Patrick's Day, Good Friday, Father's Day	Entertainment Small & Camera
Canada Day, Civic Holiday, Labour Day, Thanksgiving, Halloween,Remembrance Day, Christmas Day, Boxing Day, Islander Day & Louis Riel Day & Heritage Day & Family Day (15th Feb 2024),Easter Monday, Mother's Day, Victoria Day	Entertainment Small
Aboriginal Day, Leap Day, St. Jean Baptiste Day	Camera

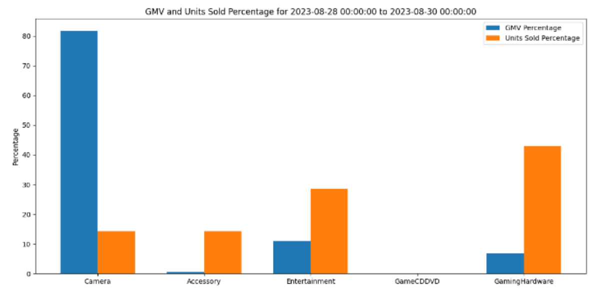
Campaign Topic on Sales Day of Every Month

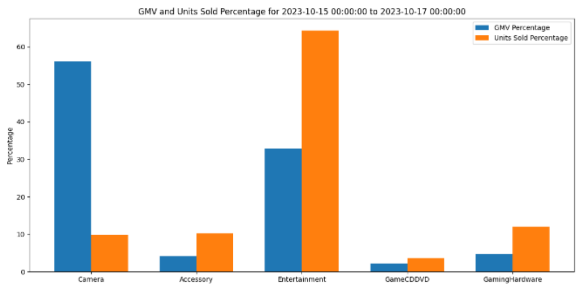
Sales Day	Campaign Topic
(18-19th July), (15-17th Aug), (25 - 31th) Dec, (1st - 3rd) Jan, (1-2 Feb)	Entertainment Small
(28-30th Aug), (7-14th Nov), (20-22) Jan, (20-21 Feb), (7-9 Mar), (25-27 May)	Camera
(15-17th Oct), (14-15 Feb)	EntertainmentSmall & Camera

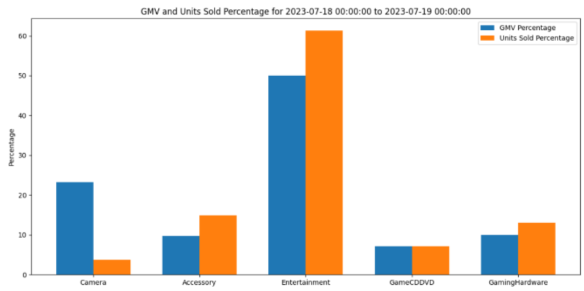












Logic for Campaign Selection:

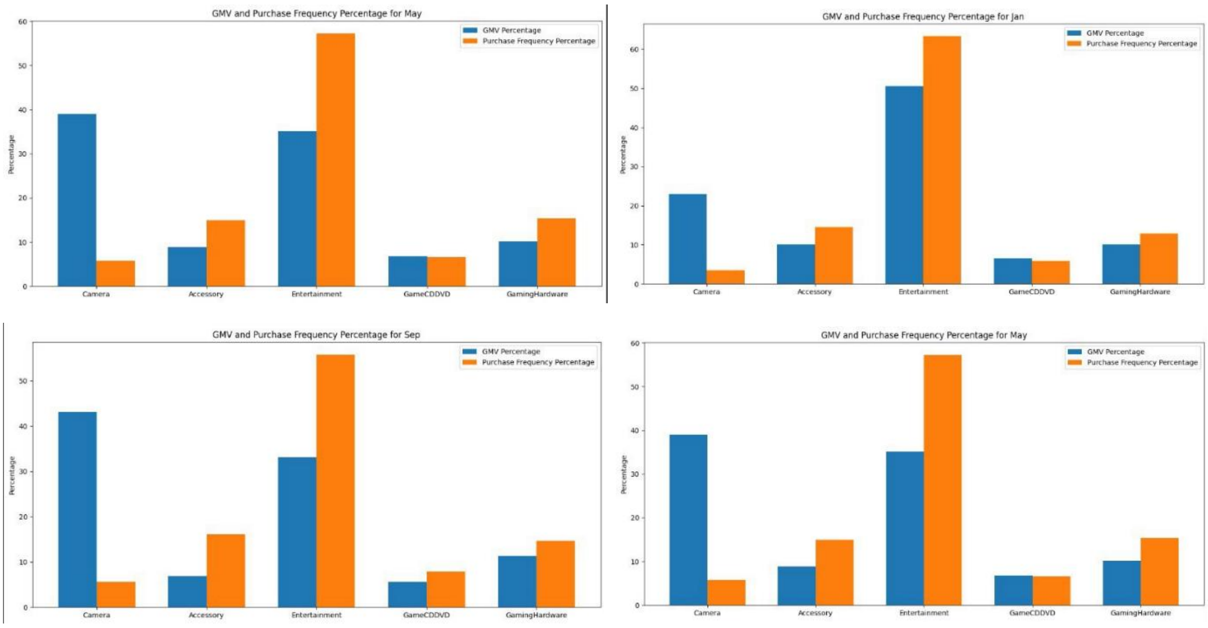
Product categories are selected based on revenue and sales performance to optimize campaign effectiveness. When EntertainmentSmall is the dominant category, revenue and units sold are high, making it the campaign's primary focus. However, when both EntertainmentSmall and Camera contribute significantly, Camera generates higher revenue, necessitating a balanced approach to ensure both categories receive adequate attention. In cases where Camera is the dominant category, its revenue remains significantly lower despite EntertainmentSmall having higher units sold. As a result, the campaign should prioritize Camera, as it serves as the major revenue driver.

5.2. Optimal Marketing Channel Allocation for Each Product Category and Its Rationale

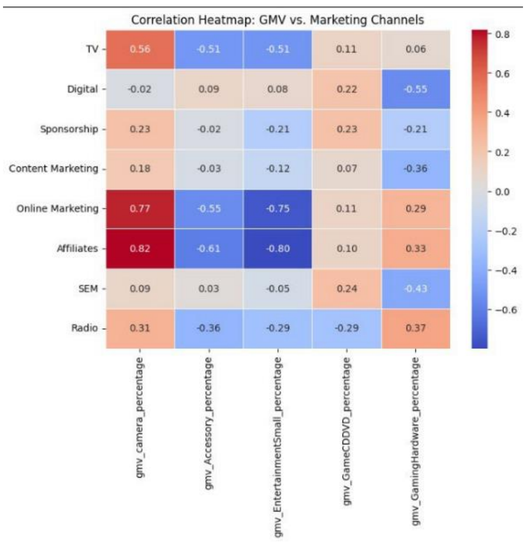
Customer order integration and product analytics data involved enriching the processed customer orders dataset with three columns: Product_Analytic_Category, Product_Analytic_Sub_Category, and Product_Analytic_Vertical. The columns were mapped based on the available FSN ID details in the SKU dataset to categorize products. Second, an aggregated dataframe was created for examination, with each row representing a month and each column indicating the overall Gross Merchandise Value (GMV) and the total units sold of a specific type of product within that month. This structured dataset provided a transparent view of the patterns of sales by month across different product groups, enabling effective marketing planning and budget distribution choices.We observed a consistent trend throughout the year by analyzing the monthly revenue and purchase frequency across different product types.

Key Insights:

- The dominance of entertainment products in revenue suggests **higher-priced items** or **higher sales volume** in this category.
- The shift in purchase frequency from entertainment to cameras after April may indicate **seasonal demand**, **marketing influences**, or **product launches**.
- Further investigation into **average order value**, **seasonal trends**, and **external market factors** can provide deeper insights into consumer behavior.



The correlation heatmap provides insights into each product type's most effective marketing channel by identifying the channel with the highest correlation.



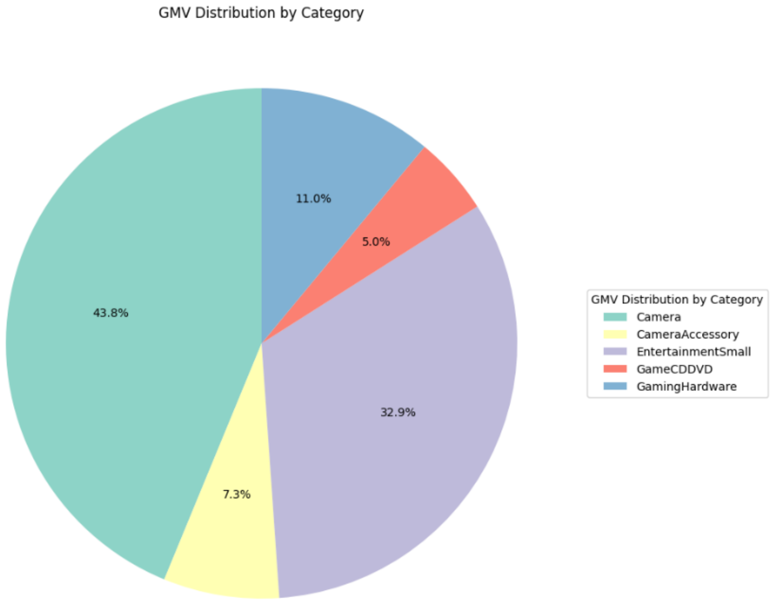
5.3 Recommendations for Future Marketing Strategies

Evaluate revenue contributions from different sales channels to calculate channel-wise ROI accurately.

1. Classify customers based on demographics (if available) to gain insights into buying patterns and preferences.
2. If regional sales data is accessible, analyze trends across different regions to identify market opportunities and growth areas.
3. Collect data on competitor pricing and marketing strategies (if feasible) to refine pricing models and enhance market positioning.

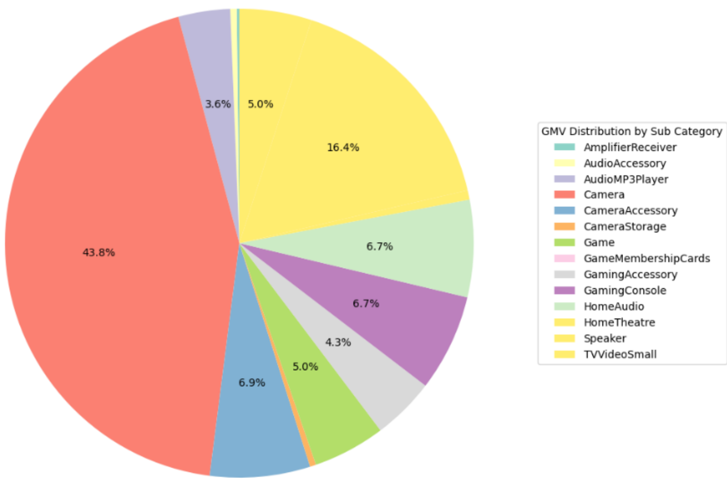
Annexure

1.



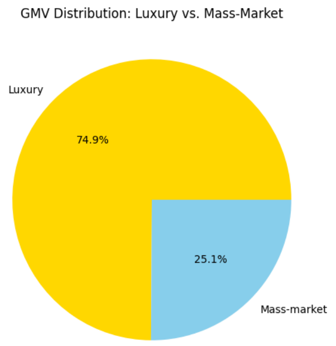
The pie chart represents the GMV (Gross Merchandise Value) distribution by category across five product categories: Camera, CameraAccessory, EntertainmentSmall, GameCDDVD, and GamingHardware. The Camera category holds the largest share at 43.8%, followed by GamingHardware at 32.9%, CameraAccessory at 7.3%, GameCDDVD at 11%, and EntertainmentSmall at 5%. This distribution highlights that the Camera and GamingHardware categories contribute the most to GMV, while EntertainmentSmall has the smallest share. The insights from this visualization can help businesses optimize inventory, marketing strategies, and budget allocation for different product categories.

2.



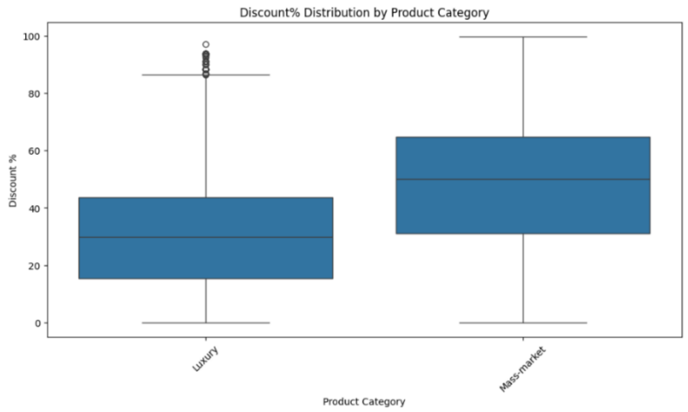
The pie chart illustrates the GMV (Gross Merchandise Value) distribution by sub-category, showcasing the contribution of various product sub-categories. The Camera sub-category dominates with 43.8% of the total GMV, making it the most significant contributor. Speaker follows at 16.4%, while other notable sub-categories include CameraAccessory (6.9%), CameraStorage (6.7%), and GamingConsole (6.7%). Smaller segments include HomeTheatre (5.0%), AudioMP3Player (5.0%), GameMembershipCards (4.3%), and GamingAccessory (3.6%). This distribution highlights the strong demand for Camera-related products and speakers, while gaming-related sub-categories contribute moderately. These insights can guide marketing, inventory, and strategic business decisions.

3.



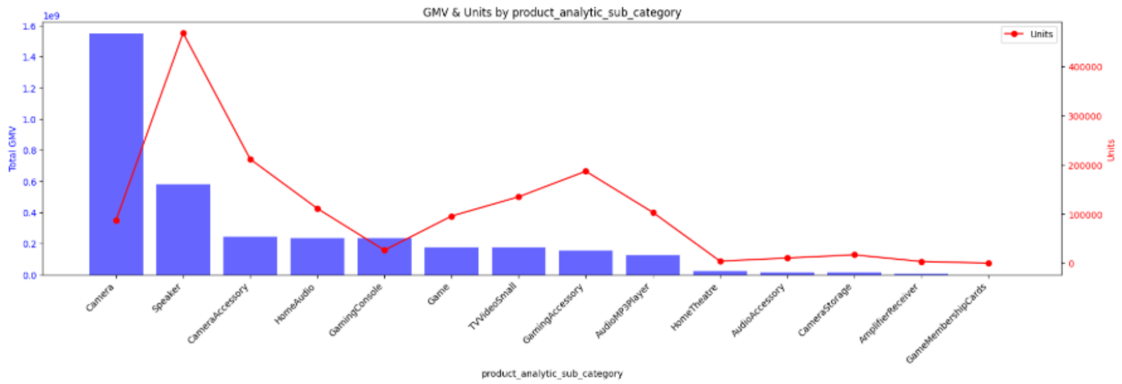
The pie chart represents the GMV distribution between Luxury and Mass-Market segments. Luxury products dominate with 74.9% of the total GMV, while Mass-Market products contribute 25.1%. This indicates a strong preference for high-end products, suggesting that the luxury segment significantly drives revenue. These insights can help in refining pricing strategies, marketing focus, and product assortment decisions.

4.



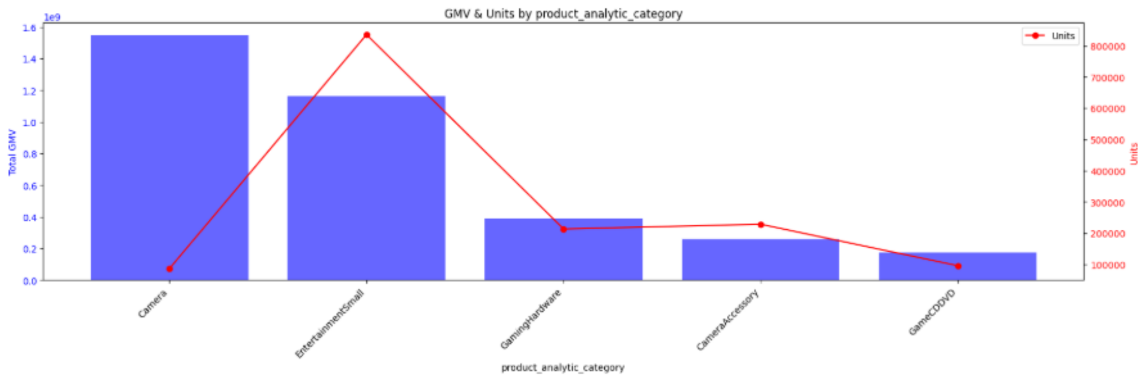
The box plot showcases the distribution of discount percentages across Luxury and Mass-Market product categories. Luxury products generally receive lower discounts, with a median discount of around 25-30%, though discounts can reach as high as 90%, with several outliers at the upper end. In contrast, Mass-Market products exhibit a higher median discount of approximately 50%, with a wider range extending up to 100%. This indicates that Mass-Market products are frequently discounted at higher rates, whereas Luxury products follow a more controlled discounting strategy, with occasional extreme markdowns. These insights can be leveraged to refine pricing and promotional strategies for both segments.

5.



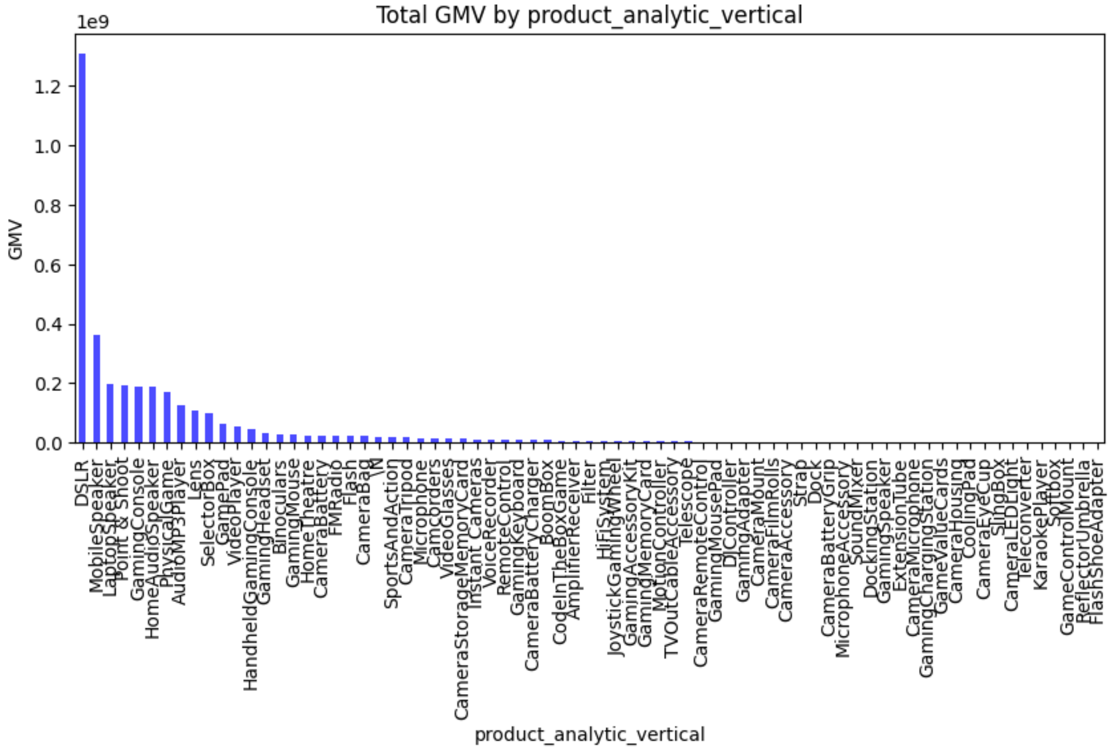
The bar and line chart illustrates GMV (Gross Merchandise Value) and Units sold across different product sub-categories. Cameras and Speakers generate the highest GMV, with Cameras leading significantly, followed by Speakers. However, the number of units sold varies, with Speakers showing a peak in unit sales. Other categories like Camera Accessories, Home Audio, and Gaming Consoles contribute moderately to GMV, while categories such as Amplifier Receivers, Game Membership Cards, and Audio Accessories show lower GMV and unit sales. The chart highlights a contrast between high-GMV but low-unit products (like Cameras) and high-unit but relatively lower-GMV products (like Speakers), indicating differences in pricing and market demand across sub-categories.

6.



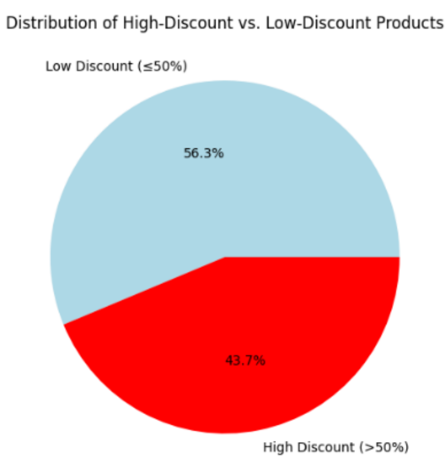
The chart displays GMV (Gross Merchandise Value) and Units sold across different product analytic categories. Cameras contribute the highest GMV, followed by the Entertainment & Small category. However, the highest number of units sold belongs to Entertainment & Small, suggesting a high-volume, lower-priced product segment. Gaming Hardware, Camera Accessories, and Game CDs/DVDs show comparatively lower GMV and unit sales. The data suggests that while Cameras generate significant revenue with relatively lower unit sales, Entertainment & Small products rely on volume sales for revenue generation.

7.



The bar chart illustrates the total GMV (Gross Merchandise Value) distribution across different product verticals, revealing a highly skewed pattern. DSLR cameras lead by a significant margin, generating the highest GMV, followed by mobile speakers, laptops, gaming consoles, and home audio products, which also contribute substantially but remain far behind DSLR cameras. The remaining product categories exhibit a sharp decline in GMV, with numerous verticals contributing only marginally to overall revenue. This trend highlights a concentration of revenue in a few high-performing product categories, emphasizing the dominance of premium and high-demand electronic items. Additionally, the long tail of low-GMV products suggests either niche demand or limited sales volume in those segments. Understanding this distribution can help businesses optimize product offerings, focus on high-revenue categories, and re-evaluate strategies for low-performing ones.

8.



The pie chart illustrates the proportion of products categorized based on their discount levels, distinguishing between high-discount products (those with more than 50% discount) and low-discount products (those with discounts of 50% or less). The majority, comprising 56.3% of the products, fall into the low-discount category, indicating that most items are sold with moderate or minimal price reductions. On the other hand, 43.7% of the products receive substantial discounts exceeding 50%, which suggests the presence of aggressive pricing strategies, likely aimed at clearing inventory, increasing customer engagement, or boosting sales volumes. The nearly balanced distribution highlights a competitive pricing approach, where a significant portion of products are offered at deep discounts, possibly during promotional events or clearance sales, while the rest maintain relatively lower price reductions to sustain profit margins.