Tab 1

TEAM 2025107

Index

Abstract

In today's competitive e-commerce landscape, data-driven decision-making is essential for optimizing marketing strategies. This study focuses on ElectroMart, a Canada-based electronics retailer, which seeks to improve its marketing budget allocation for the upcoming year. The primary challenge is that previous marketing expenditures did not yield expected revenue growth, prompting the need for a data-driven reassessment.

To address this, we conducted a comprehensive data analytics investigation using multiple datasets, including customer orders, media spend, promotional events, customer satisfaction scores (NPS), and weather data. Through exploratory data analysis, mathematical modeling, and statistical insights, we assessed the impact of different marketing levers on revenue. Our findings reveal the most influential KPIs affecting revenue, the effectiveness of various advertising channels, and seasonal trends influencing sales performance.

Based on these insights, we propose a **SHAP-based Multi-Channel Optimization Algorithm** to enhance marketing effectiveness while optimizing costs. Our approach is projected to drive a **12.56% increase in revenue** by ensuring more efficient allocation of the investment budget.

Data Processing

Preprocessing and Cleanup

Missing values are identified and addressed, particularly in columns with a high proportion of null values. Noisy columns that contribute excessive noise or lack meaningful insights, such as *Deliverybdays* and *Deliverycdays*, are removed. Duplicate rows are detected and eliminated to maintain data integrity. Date columns are identified and converted to the appropriate datetime format. Additionally, inconsistent data points, such as instances where *product_mrp* or *gmv* equals zero, are filtered out to prevent data entry errors. Missing ***gmv*** values are estimated using ***fsn_id*** and ***Sales*** as reference points. After imputation, rows where ***gmv*** remains zero are removed to maintain consistency. Datasets are merged based on common identifiers like ***FSN ID, Order Date, and Customer ID*** to enable a comprehensive understanding of sales trends and customer behavior.

Feature Engineering

Feature engineering is then performed to enhance the dataset with additional insights. A ***Luxury Indicator*** is introduced to categorize products based on price points, determining whether they fall within the luxury segment. ***Holiday Flag*** is created to indicate whether a product was sold on a holiday or a specific day of the week. A ***Discount Calculation*** feature is added to capture the impact of price reductions on sales.

Tech Stack Used:

Python, Matplotlib, Seaborn, Plotly : For efficiently querying and visualization

Power BI : For Dashboard

Key Performance Indicators (KPIs)

1. Overview

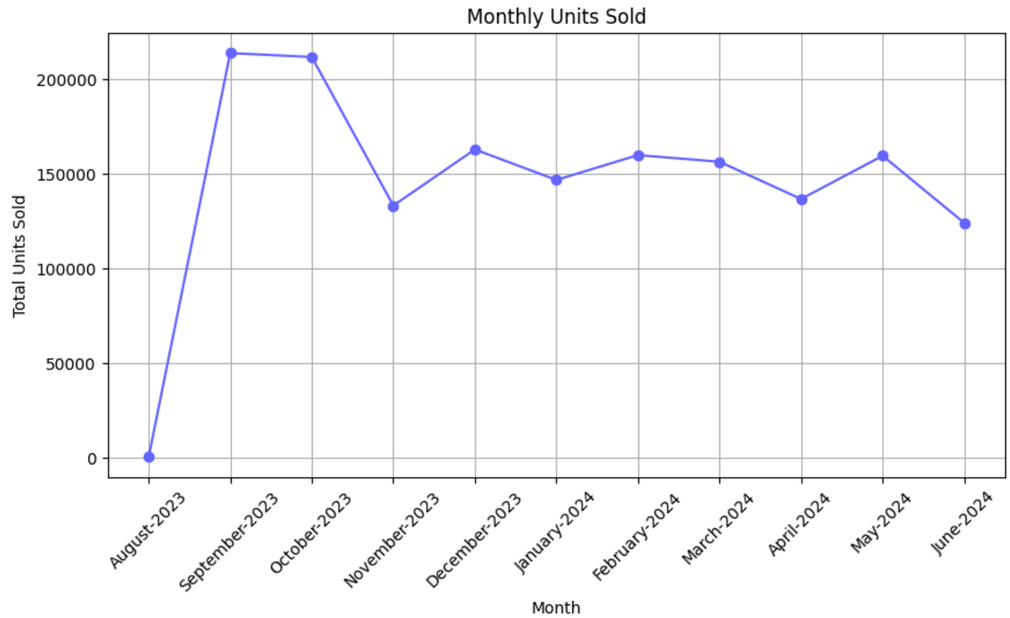
Key Performance Indicators (KPIs) are measurable metrics that assess business performance, guiding data-driven decisions to optimize strategies, improve efficiency, and drive growth.

2. KPIs Analysis

a. **Monthly Units Sold:** Sales surged in September 2023, then stabilized with peaks in February and May. Fluctuations suggest seasonal trends or campaigns, while June saw lower demand.

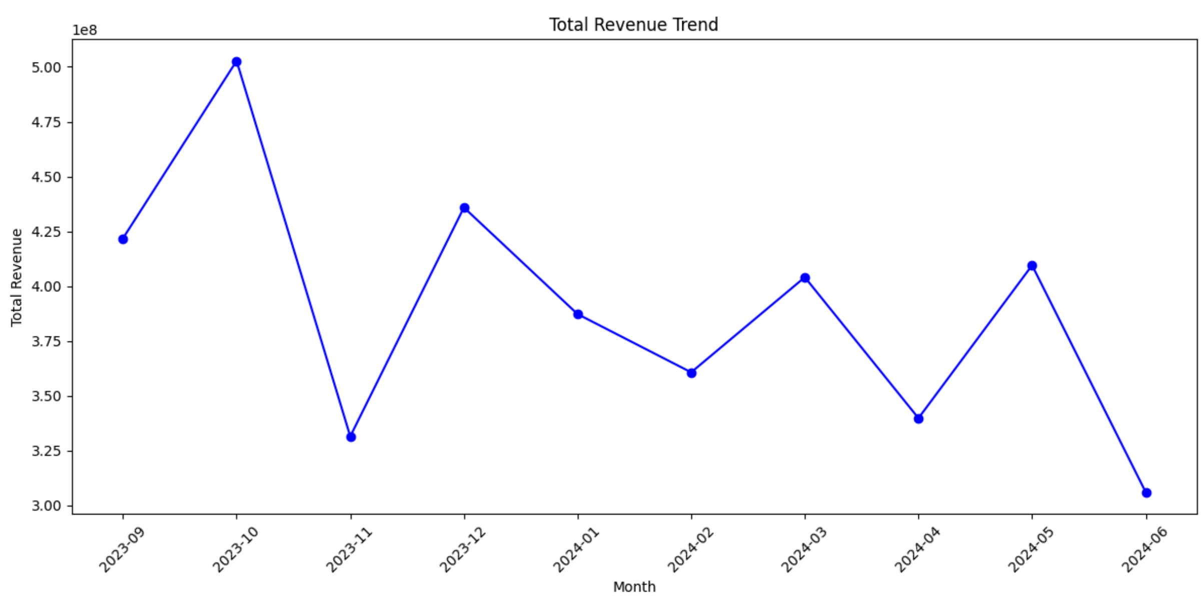
b. **Annual Gross Merchandise Value (GMV):** The GMV is ₹4 billion.

c. **Annual number of Orders:** Total number of orders was 1413102.

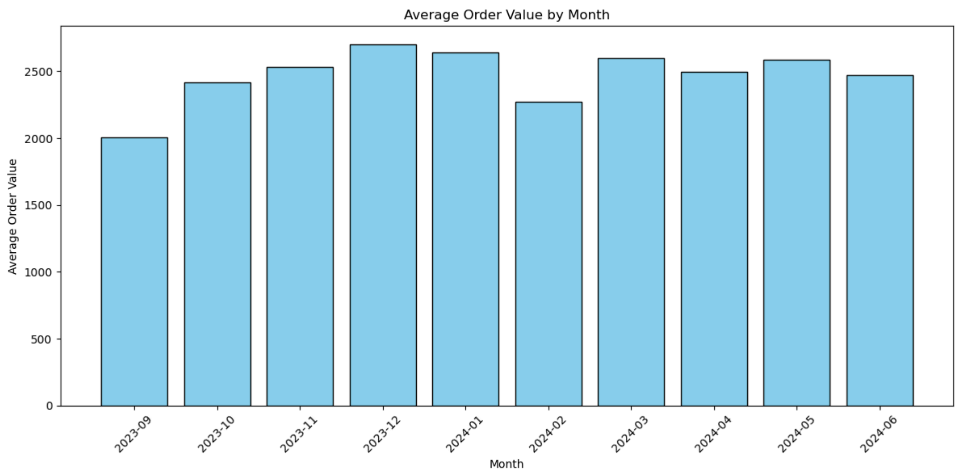


d. **Net Promoter Score (NPS):** It measures customer loyalty by subtracting the percentage of detractors (0–6) from promoters (9–10) based on their likelihood to recommend a business.

e. **Revenue trend line(total revenue vs time):** The graph shows the total revenue trend over time, with fluctuations indicating periods of increase and decline in revenue.



f. **No. of active customers per day:** The number of active customers per day is crucial for tracking user engagement, measuring retention, identifying trends, and optimizing business strategies to drive growth and revenue.



g. **Avg order value per month:** Average Order Value (AOV) per month helps businesses assess revenue efficiency, optimize pricing strategies, and improve marketing efforts by understanding customer spending behaviour.

h. **Total units sold by product category:** This metric enables businesses to identify high-demand products, streamline inventory management, refine marketing strategies, and enhance profitability by prioritizing top-selling categories.

i. **Customer acquisition cost:** Customer Acquisition Cost (CAC) is crucial for evaluating the efficiency of marketing efforts, ensuring sustainable profitability, optimizing budget allocation, and balancing customer lifetime value (LTV) with acquisition expenses.

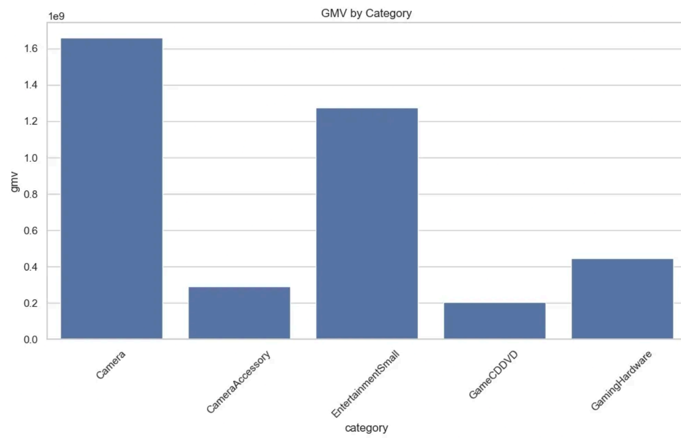
Sales and SKU Performance Analysis

Understanding sales performance across product categories and verticals is crucial for optimizing revenue streams and inventory planning. This analysis explores key trends in Gross Merchandise Value (GMV), fulfillment efficiency, procurement challenges, and sales distribution. The study identifies high-performing categories, inefficiencies in supply chain processes, and strategic growth opportunities to enhance market positioning.

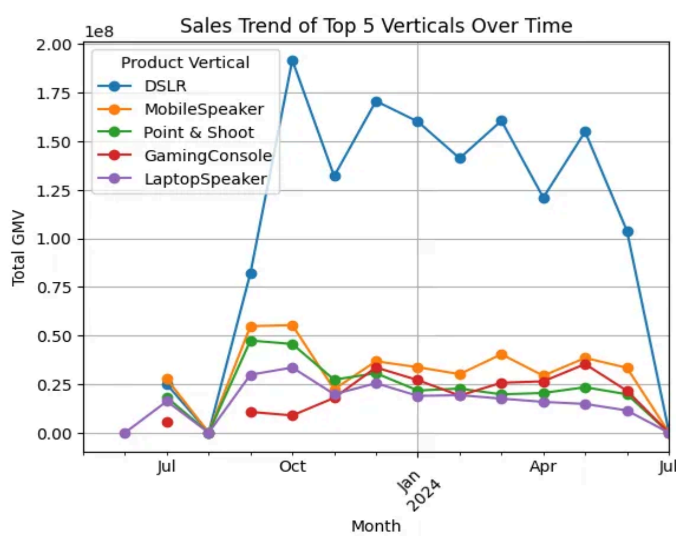
1. Top-Selling SKUs and Market Trends

The sales distribution across categories highlights a dominant revenue contribution from **Cameras**, followed by **EntertainmentSmall**, indicating strong consumer demand for high-value electronics. **Camera Accessories** and **Gaming Hardware** show moderate growth, presenting opportunities for bundling and upselling. In contrast, **Game CD/DVD** has the lowest GMV, reflecting an industry shift toward digital consumption and streaming services.

- **DSLRs lead in GMV**, posing a risk of over-reliance on a single product segment.
- Other verticals contribute marginally, indicating weak diversification.
- **Peripheral accessories**, such as camera add-ons and gaming peripherals, maintain steady demand, reinforcing their role in enhancing core product ecosystems.

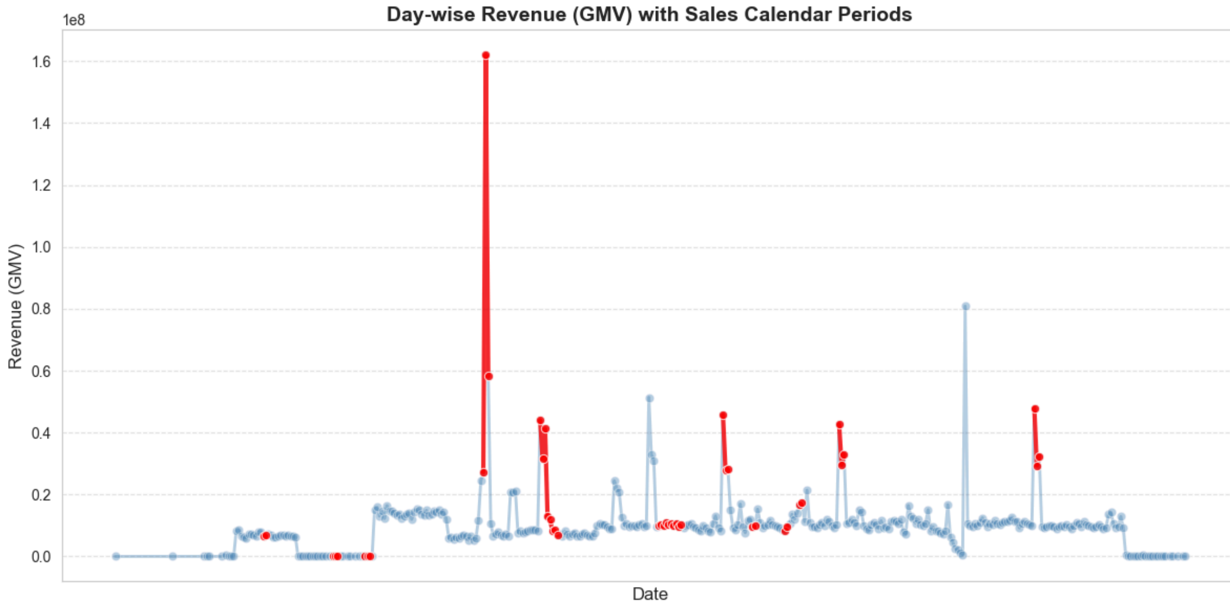


To mitigate the risks of market saturation and reliance on specific categories, diversifying the product portfolio is essential. Expanding into gaming peripherals, premium audio devices, and smart home entertainment can generate additional revenue streams. Bundling strategies - such as offering Camera Accessories with DSLRs or Gaming Hardware with console purchases - can enhance cross-selling potential. Additionally, underperforming categories should be reassessed for repositioning or discontinuation to optimize resource allocation.



Sales spiked sharply in mid-October, driven by seasonal promotions, before stabilizing at a lower level with fluctuations. DSLR dominated the trend with the highest peak and volatility, while other categories showed steadier performance.

2. GMV Trends and Key Observations



2.1 Peak GMV Periods

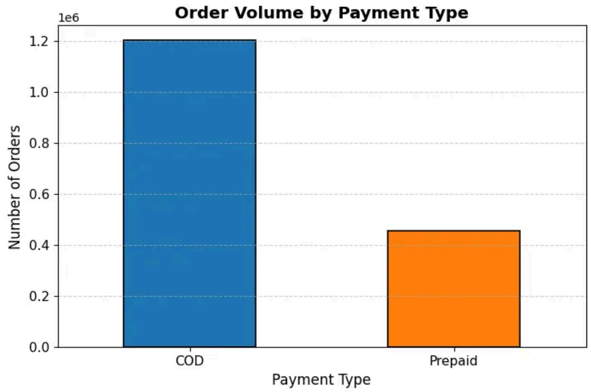
The mid-October 2023 revenue spike coincided with seasonal shopping events, driven by aggressive promotions such as discounts, bundled deals, and exclusive launches. The sharp GMV increase indicates strong consumer responsiveness to time-sensitive offers. However, sales declined significantly post-event, reflecting the temporary nature of such surges without sustained demand-generation strategies.

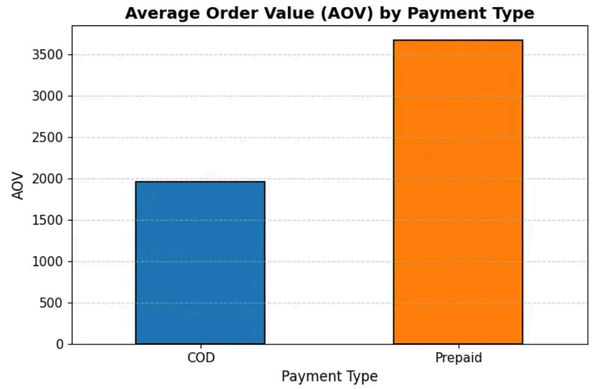
2.2 Revenue Normalization Post-Peak

After the peak, GMV stabilized with minor fluctuations from organic sales and smaller promotions. This pattern highlights the challenge of sustaining revenue beyond major sales events. Relying solely on periodic promotions increases revenue volatility, emphasizing the need for long-term demand-generation strategies to ensure consistent growth.

2.3 Payment Type Impact on AOV and Order Volume

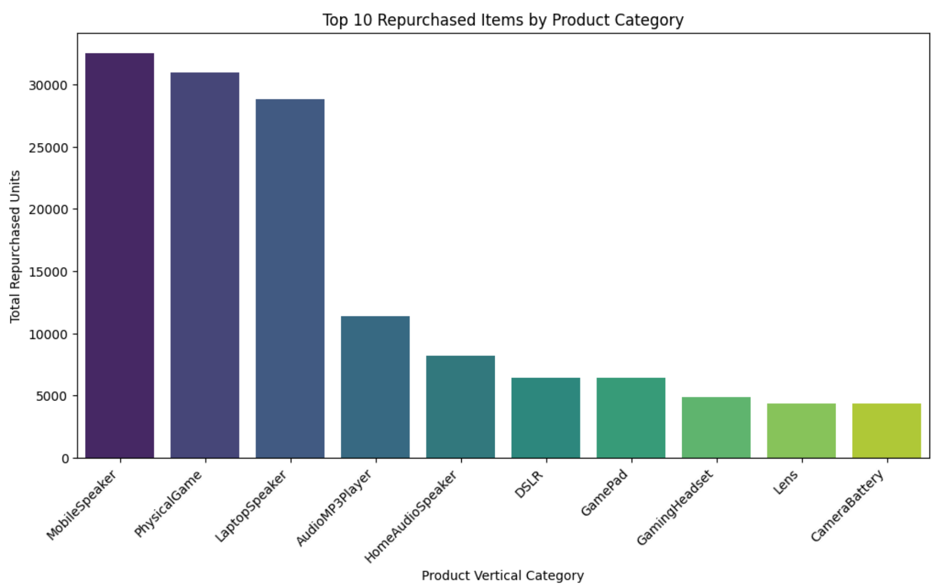
Prepaid transactions have a significantly higher **Average Order Value (AOV)** compared to **Cash on Delivery (COD)**, indicating that customers spend more per order when paying upfront. However, **order volume** is much higher for COD, suggesting that more customers prefer this payment method despite lower spending per order.
Understanding these trends helps businesses optimize payment strategies, manage risk, and tailor promotions. Encouraging prepaid transactions through incentives can increase revenue, while ensuring seamless COD experiences can maintain high order volumes and customer trust.





2.4 Repurchased Items

Mobile Speakers, Physical Games, and Laptop Speakers are the most frequently repurchased items, indicating strong consumer demand for audio and gaming products. While high-value items like DSLRs and lenses also see repeat purchases, the dominance of smaller, more affordable electronics suggests that customers frequently upgrade or replace these products.



Customer Behaviour

Customer segmentation is a vital analytical framework that allows businesses to classify their customer base into distinct groups based on shopping behavior, spending patterns, discount sensitivity, and product preferences. By analyzing key transactional metrics such as order volume, frequency, and category engagement, businesses can enhance their marketing strategies, optimize promotional campaigns, and increase customer lifetime value.

Customer Behavioral Segmentation

The segmentation model categorizes customers into four groups, each exhibiting distinct purchasing behaviors:

1. **Casual Shoppers (78.3%)**
   They purchase infrequently, averaging 0.8 orders per customer, but their average order value (AOV) is high at around $2500. Despite this, their sale ratio is low at 0.15, indicating a high risk of churn.

2. **Category Enthusiasts (17.2%)**
   They tend to focus on specific product categories and place about 2.5 orders on average. Their AOV is also around $2500, and they have the highest sale ratio at 0.65, making them highly responsive to targeted recommendations.

3. **Discount-Driven Shoppers (2.3%)**
   They place around 2.2 orders per customer, primarily during promotions. Their AOV is the lowest at $1800, and their sale ratio is just 0.05, reflecting a strong dependency on discounts.

4. **High-Value Loyal Customers (2.2%)**
   High-Value Loyal Customers make up just 2.2% of the base but are major contributors to revenue. They order frequently, with an average of 5 purchases per customer, and have the highest AOV at approximately $2700. However, their sale ratio is still low at 0.10, suggesting room for improved engagement.

Strategic Insights and Business Implications

The majority of customers are **Casual Shoppers** and **Discount-Driven Shoppers**, highlighting strong price sensitivity. However, **High-Value Loyal Customers**, though fewer, generate the most revenue due to their high purchase frequency and AOV, making them a priority for retention strategies. Strengthening loyalty programs with exclusive rewards and premium services can enhance long-term engagement.

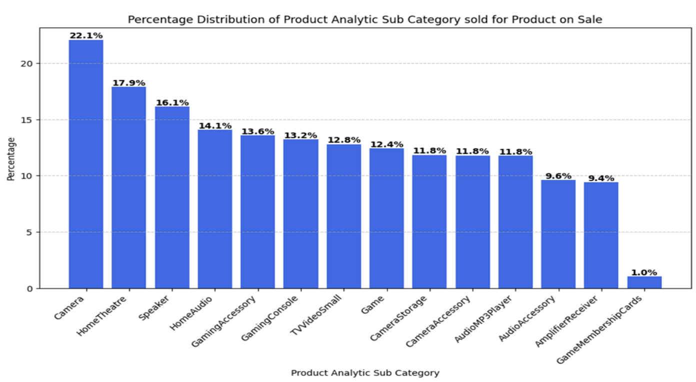
Discount-Driven Shoppers require a refined promotional strategy to balance frequent purchases with profitability. Personalized, time-sensitive discounts can improve conversions without excessive margin loss. **Category Enthusiasts**, who purchase 65% of the times in sale, respond well to AI-driven recommendations, making personalized product suggestions a key driver of repeat purchases.

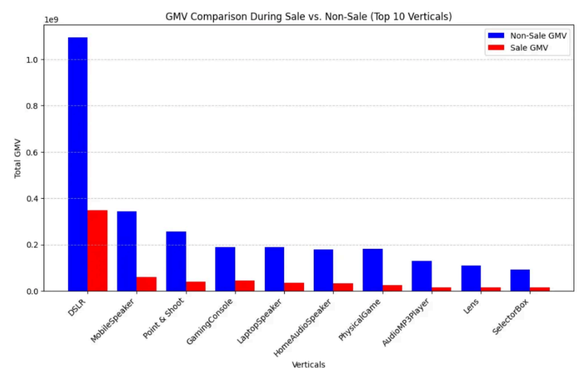
Casual Shoppers, the largest but least engaged group, present a reactivation opportunity. With a low sale ratio (**0.15**), targeted campaigns, personalized promotions, and strategic discounts can encourage repeat purchases. Aligning marketing efforts with these segment-specific behaviors will optimize engagement, boost profitability, and strengthen long-term customer retention.

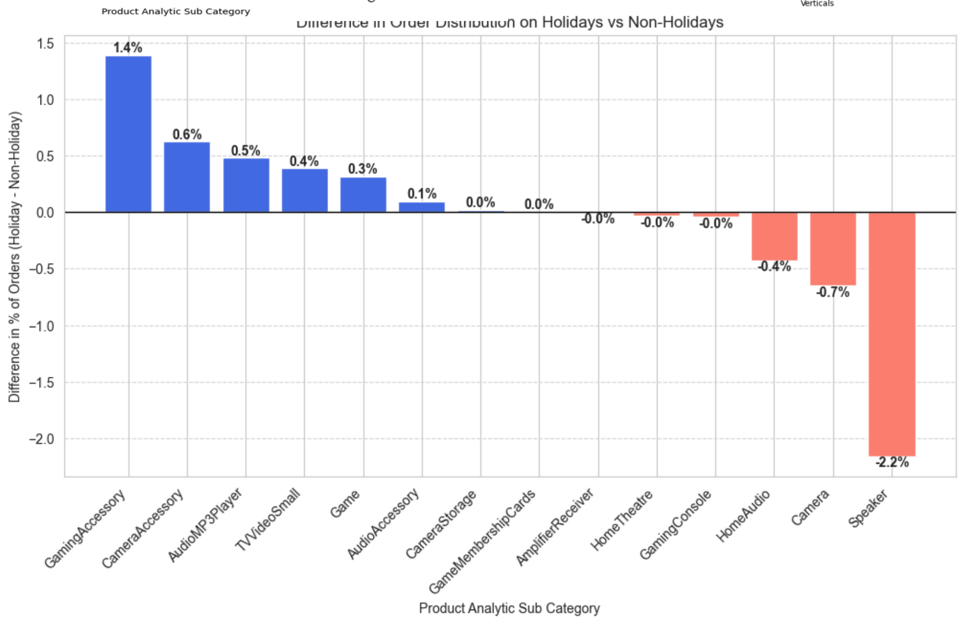
Understanding customer behavior is essential for improving engagement, personalization, and overall sales performance. This analysis explores key trends in purchase patterns, occasional demand shifts and buying frequency. Additionally, it highlights opportunities for enhancing customer experience, optimizing marketing strategies, and improving retention through data-driven insights.

3.1 Impact of Sales

Cameras, Home Theatre, and Speakers account for the highest percentage of products sold during sales events. While Cameras and Home Theatre systems are high-value items, the strong







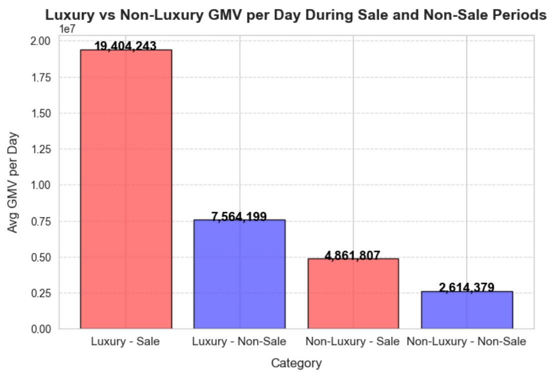
sales of Speakers suggest that consumers are also inclined to purchase relatively affordable electronics during promotions.

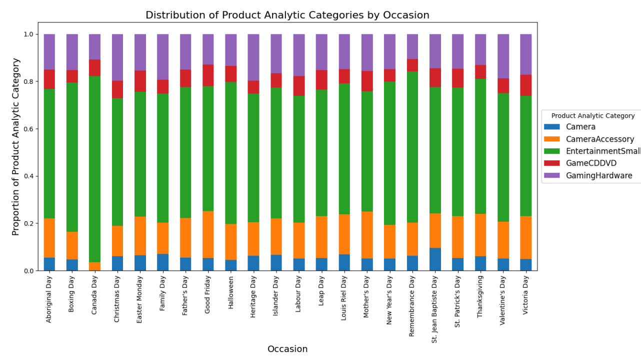
3.2 Impact of Holidays

Gaming and camera accessories see increased sales on holidays, while high-value items like speakers and cameras experience a drop. Consumers may delay expensive purchases, highlighting the need for discounts, financing options, and bundled offers to boost holiday sales.

3.3 Impact Of Sales and Occasion

EntertainmentSmall is the top-selling category across all occasions, with Camera Accessories and Gaming Hardware contributing steadily. Holiday events like Christmas and New Year's show a slight rise in GameCD/DVD purchases, reflecting seasonal gifting trends.

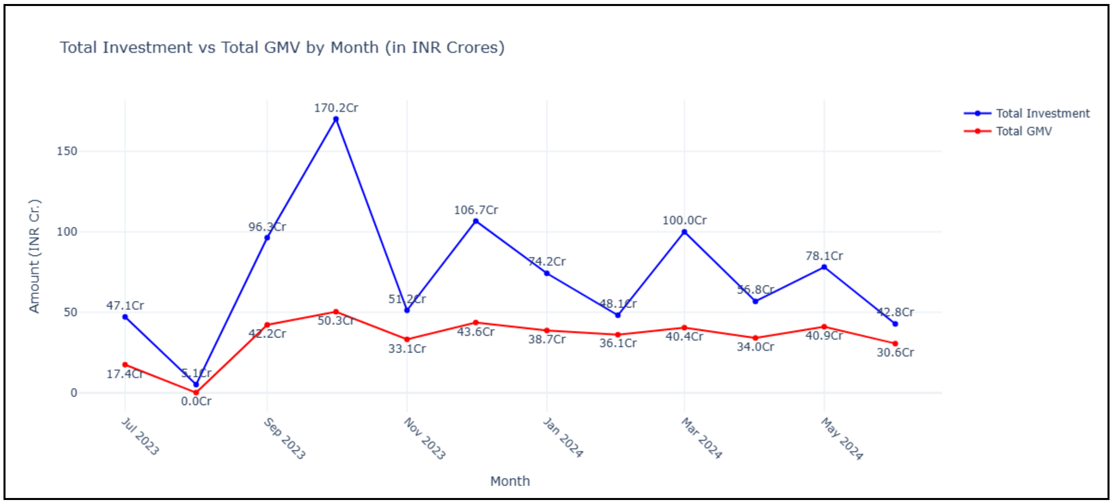




Luxury products see a massive 162% increase in GMV during sales. Non-luxury products experience a 90% increase in GMV during sales. Luxury products are more dependent on sales, as their increase is nearly 1.8x higher than non-luxury products.

Marketing Efficiency

Upon analyzing the investment and GMV trends in relation to the sales calendar, we identified several key inefficiencies that warrant attention. Firstly, we observed instances of disproportionate spending during sales periods which did not really make an impact given the investment delta. For example, during *October 2023*, investment was 170.2 Cr for a 3-day sale, yet the resulting GMV was only 8.1 Cr more than the previous month. Similarly, in *March 2024*, investment spiked to 100 Cr, nearly twice the amount in February for a similar 3-day sale, but the GMV only saw a 4 Cr addition. Interestingly, both *September 2023* and *March 2024* had nearly identical investment levels (around 96.3 Cr and 100 Cr), yet the GMV returns were similarly underwhelming, in fact lesser GMV was reported in September.



We observed a pattern of GMV stability despite fluctuating investments, suggesting potential inefficiencies in marketing strategy. For instance, in February 2023 and April 2024, investment dropped significantly, yet GMV remained steady, indicating diminishing returns or ineffective targeting. Notably, in January 2024, GMV was comparable to higher-investment months despite a lower budget, likely due to effective cross-channel allocation or improved conversion rates. However, these strategies were not consistently scaled, leading to irregular performance in subsequent months.

Additionally, we noticed that investment-driven GMV spikes often fail to deliver sustained impact, highlighting a potential gap in customer retention strategies. For instance, after the high GMV month of October 2023, where significant investment was made, there was no noticeable carryover effect on subsequent months. This suggests that the campaigns may have been too focused on short-term sales boosts rather than fostering long-term customer relationships. This can be cross-verified by the fact that repeat customers account for only around 24% of the overall customer base. This low retention rate indicates that while the campaigns may effectively attract new buyers, they struggle to convert them into loyal customers.

Budget Optimization

Data-driven framework for optimizing marketing budget allocation using explainable machine learning outputs and temporal effect modeling. Leveraging SHAP (SHapley Additive exPlanations) values to capture the marginal impact of marketing channels on Gross Merchandise Value (GMV), and lag coefficients to account for time-dependent effects, the approach identifies optimal investment strategies that maximize predicted returns. A global optimization technique, Differential Evolution, ensures robustness in navigating the non-convex search space while maintaining the total investment constraint. The resulting budget allocation framework supports interpretable, ROI-focused marketing decisions.

Channel Impact Attribution via SHAP Values

The model begins with SHAP values derived from a trained XGBoost Model quantifying the marginal contribution of each marketing channel to predicted GMV. These values offer an interpretable estimate of each feature's influence, serving as the basis for assessing the responsiveness of GMV to changes in budget allocation.

SHAP (SHapley Additive exPlanations) values quantify the marginal contribution of each marketing channel to predicted GMV by measuring its impact across different feature combinations. Derived from an XGBoost model, these values provide an interpretable estimate of each channel's influence, enabling data-driven budget optimization.

Lag coefficients are introduced to model temporal persistence in channel performance. These coefficients, estimated via time series or regression models, reflect how prior period investments continue to impact future GMV. Each SHAP value is adjusted using a geometric decay model to integrate lag with SHAP-based attribution. This results in an effective SHAP metric:

$$\text{Effective SHAP}_i = \text{SHAP}_i \times \left(\frac{1}{1-\lambda_i}\right)$$

where λ_i denotes the normalized lag coefficient for channel i, capturing cumulative delayed effects.

The optimization objective is to maximize the predicted GMV under new budget allocations. The change in GMV is modeled multiplicatively:

$$\widehat{GMV} = GMV_{\text{initial}} \times \prod_i \left(1 + \Delta B_i \cdot \frac{\text{Effective SHAP}_i}{B_{i,\text{initial}}}\right)$$

Where ΔB_i represents the change in budget for channel i. This formulation assumes independent and compounding effects of each budget change on GMV.

The optimization is subject to a strict budget constraint that keeps the total investment constant. The differential evolution algorithm is applied to solve the resulting constrained, non-linear optimization problem. This evolutionary strategy is particularly well-suited for high-dimensional, non-convex problems and efficiently explores the feasible space without requiring gradient information.

The optimized allocation provides a new distribution of marketing spend across channels expected to yield the highest GMV, considering both immediate effects (SHAP) and carryover impacts (lag). A comparative analysis between initial and optimized budgets and the corresponding GMV estimates enables marketing teams to make transparent, data-backed decisions.

Our objective is to allocate each marketing channel's fixed annual budget over 12 months in a way that maximizes revenue while accounting for seasonal sales variations. To achieve this, we combine two critical components:

(1) assessing monthly revenue efficiency

(2) adjusting historical channel spending

The in_sale column indicates people buying goods during sale periods, and since most sales occur at specific times of the year, incorporating this metric helps capture seasonal spikes in purchase activity.

1. Monthly Revenue Efficiency

We calculate a reverse efficiency measure for each month to determine which months deliver the best return on investment. For a given month m, the reverse efficiency is defined as:

$$E'_m = \frac{\mathrm{GMV}_m}{\text{Total Investment}_m}$$

Where GMV (Gross Merchandise Value) represents revenue, and Total Investment is the overall spending across all channels for that month. We then normalize these efficiencies across the 12 months to derive monthly weights:

$$W_m = \frac{E'_m}{\sum_{m=1}^{12} E'_m}$$

A higher W_m indicates a month where the investment is more effective in generating revenue, suggesting that a larger share of the overall budget should be allocated during that period.

2. Historical Channel Share

Historically, each channel has contributed a specific proportion of the monthly investment. For each channel j in month m, we calculate its historical spending share as:

$$r_{mj} = \frac{\text{Investment}_{mj}}{\text{Total Investment}_m}$$

This ratio provides a baseline allocation based on past spending patterns across channels.

To further refine our allocation, we adjust the historical channel shares by incorporating the in_sale metric, which indicates the level of purchase activity during sale periods. Because most sales occur seasonally, using the in_sale weightage helps ensure that our allocation reflects these consumer behavior spikes.

We compute a global correlation coefficient $\acute{C}_j$ for each channel between spending and the in_sale metric.

We then adjust the historical share by comparing each channel's correlation with the average correlation $\acute{C}_j$ using a scaling parameter α:

$$\tilde{r}_{mj} = r_{mj} \times \left[1 + \alpha\left(\tilde{C}_j - \bar{C}\right)\right]$$

Channels that exhibit above-average correlations with in_sale - meaning they are more effective at driving sales - receive a boost, whereas channels with lower correlations are scaled down. Finally, we normalize these adjusted shares for each month to obtain channel-specific weights:

$$W_{mj} = \frac{\tilde{r}_{mj}}{\sum_j \tilde{r}_{mj}}$$

3. Budget Allocation

Each marketing channel has a predetermined fixed annual budget. The final step is to distribute each channel's budget over the 12 months using the insights gained from monthly efficiency and the adjusted channel weights.

For each channel j and month m, we compute an unnormalized fraction of the channel's budget:

$$F_{mj} = W_m \times W_{mj}$$

These fractions are then normalized for each channel so that the sum over all months equals 1:

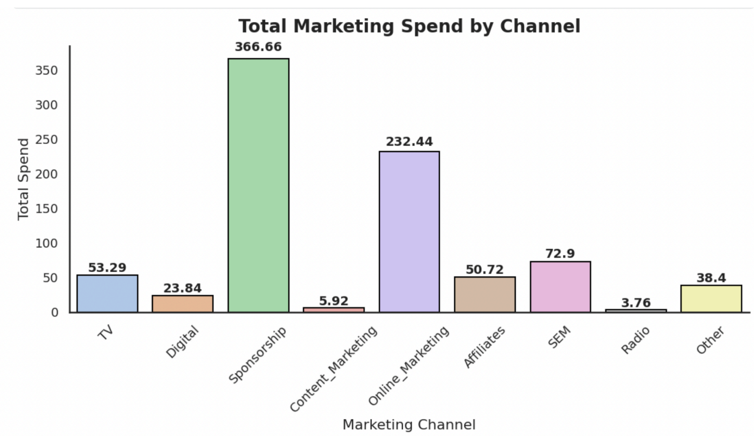
$$F'_{mj} = \frac{F_{mj}}{\sum_{m=1}^{12} F_{mj}}$$

Finally, given a fixed annual budget B_j for channel j, the budget allocated for month m is:

$$B_{mj} = B_j \times F'_{mj}$$

This method ensures that each channel's fixed budget is distributed over the months, reflecting both overall revenue efficiency and the seasonal impact on sales captured by the in_sale metric.

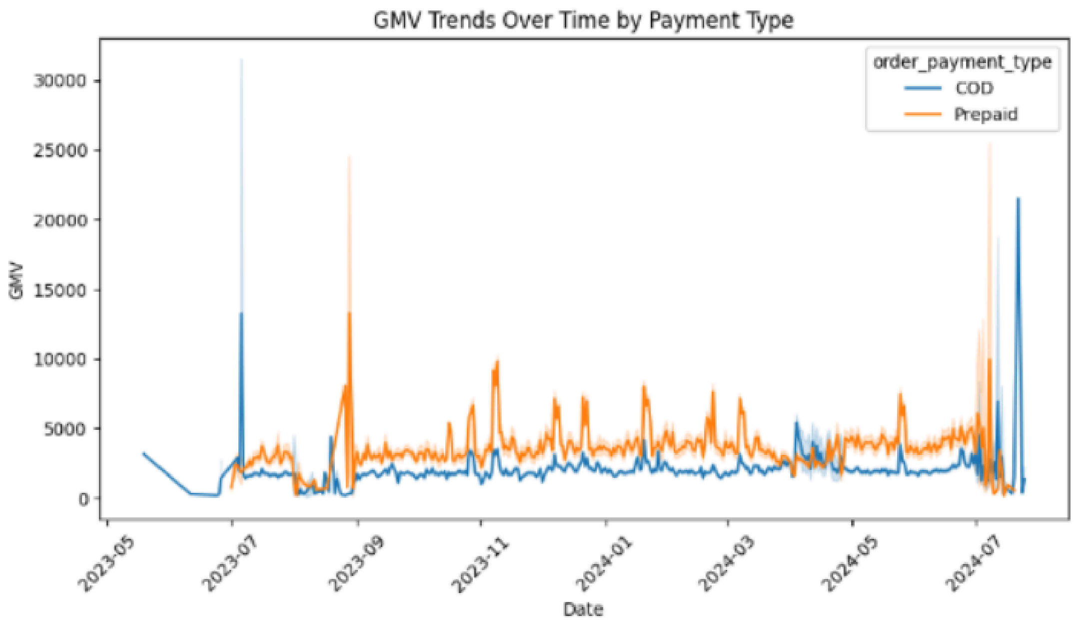
This balanced, data-driven approach optimizes the distribution of fixed channel budgets over time, aligning spending with periods of high revenue potential.



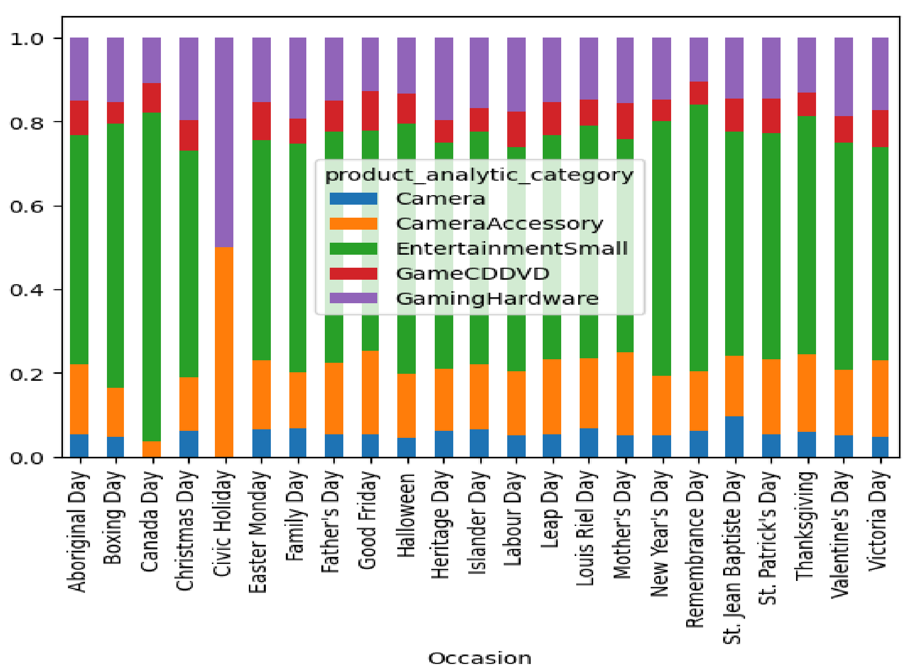
We forecast a 12.56% increase in the annual GMV in the next year with our optimized multi-channel allocation which is given as follows.

YM	TV	Digital	Sponsorship	Content_Marketing	Online_Marketing	Affiliates	SEM	Radio	Other
2024-7	2.52	4.74	38.57	2.06	9.4	2.67	12.86	0	0
2024-8	9.9	2.78	36.85	0.69	43	14.33	11.95	0	0
2024-9	2.95	1.54	34.1	0.88	11.44	4.43	5.96	0	0
2024-10	1.27	14.38	51.57	0	7.66	2.92	30.85	0	0
2024-11	0	0.29	0.29	0	0.02	0.02	0.59	0	0
2024-12	3.73	1.22	66.14	0.84	14.61	4.41	5.78	0	0
2025-1	5.69	0.59	5.98	1.71	27.56	8.81	5.29	3.68	25.72
2025-2	5.76	3.83	28.54	1.95	41.05	13.27	10.58	0	25.86
2025-3	7.91	1.62	38.97	0.75	14.57	4.86	4.31	0.81	0
2025-4	8.53	1.34	43.87	0	25.19	8.61	6.71	0	6.02
2025-5	1.4	0.73	34.88	0	22.05	6.26	6.72	1.16	0
2025-6	3.63	2.71	58.71	0	15.89	5.5	7.69	0	0

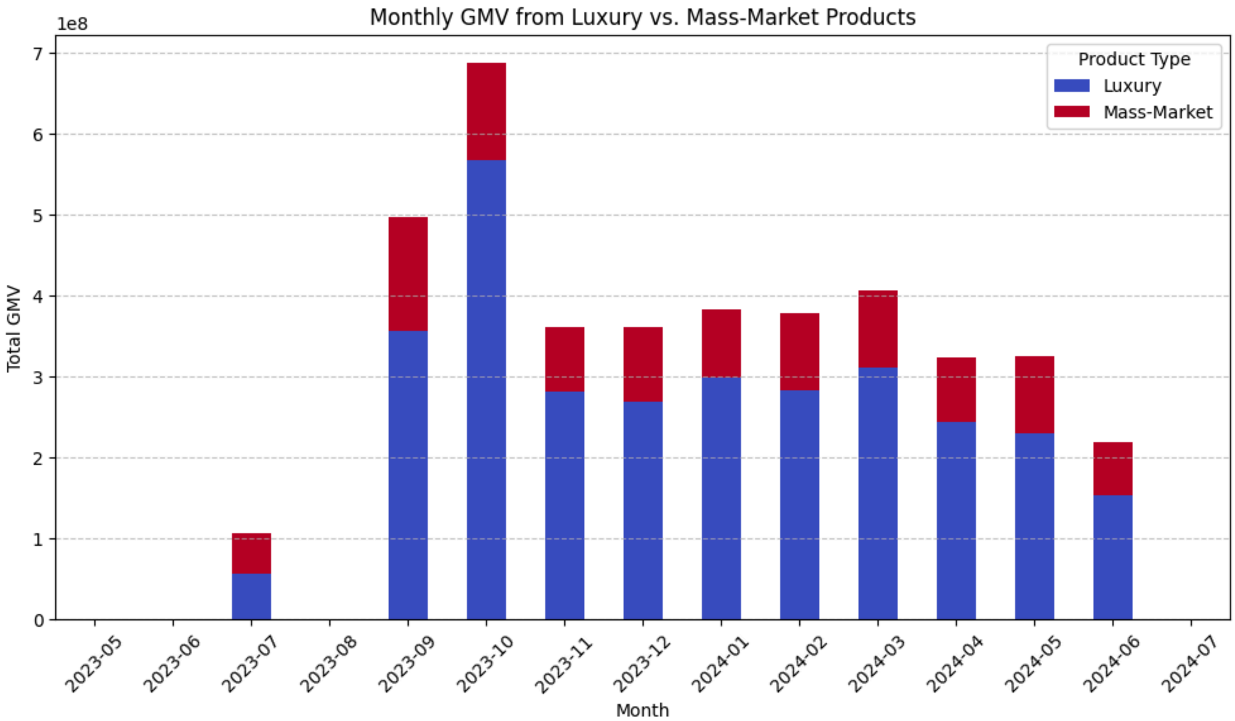
Annexure



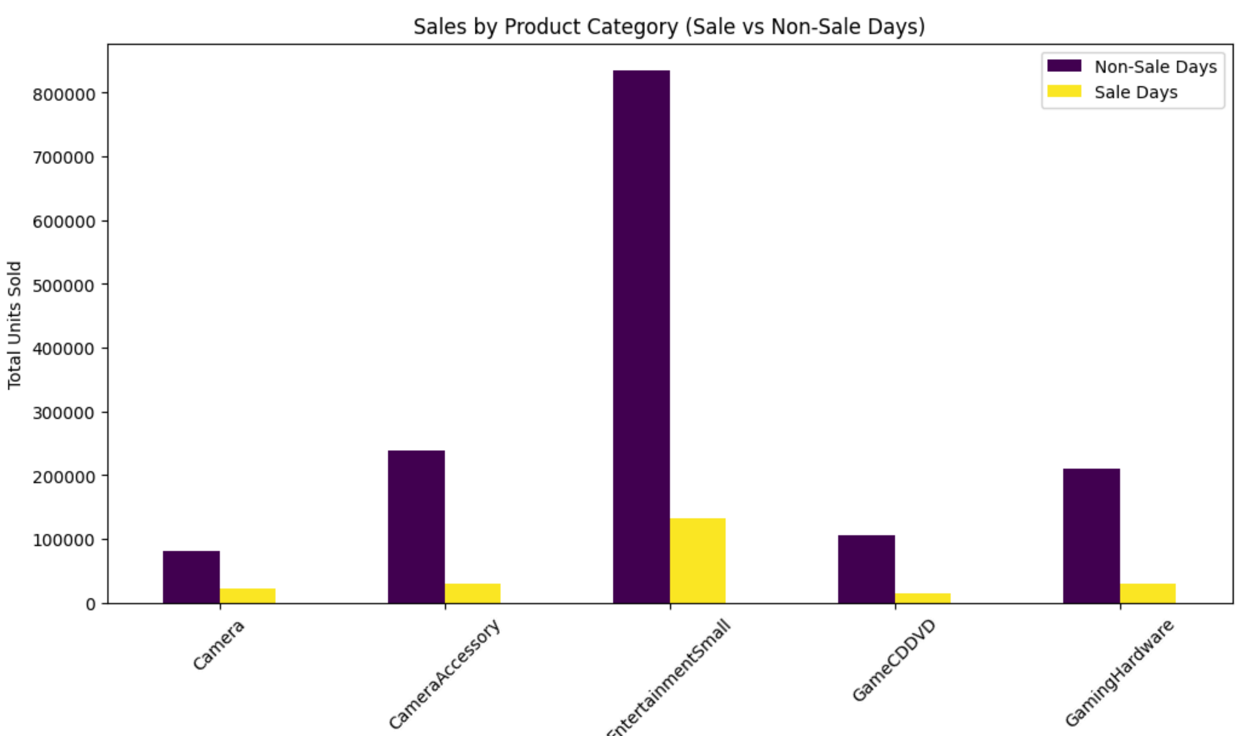
The graph shows GMV trends, where prepaid payments have higher volatility with frequent spikes, while COD remains more stable with occasional peaks, indicating different purchasing behaviors.



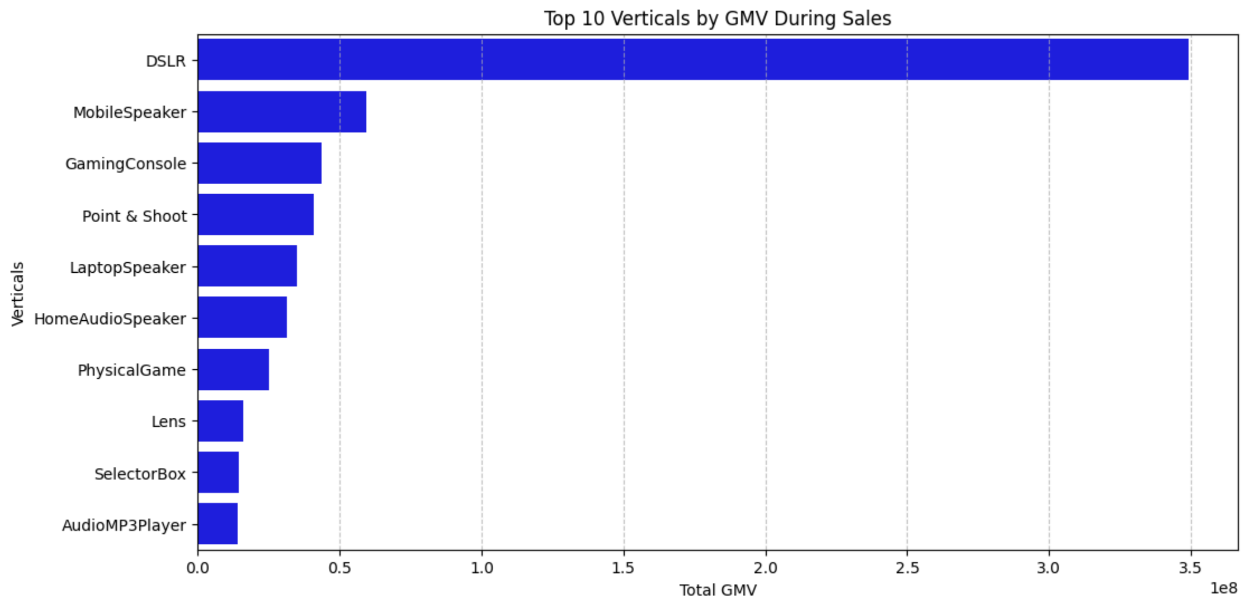
The graph shows product category sales distribution across different occasions, with EntertainmentSmall and GamingHardware dominating.



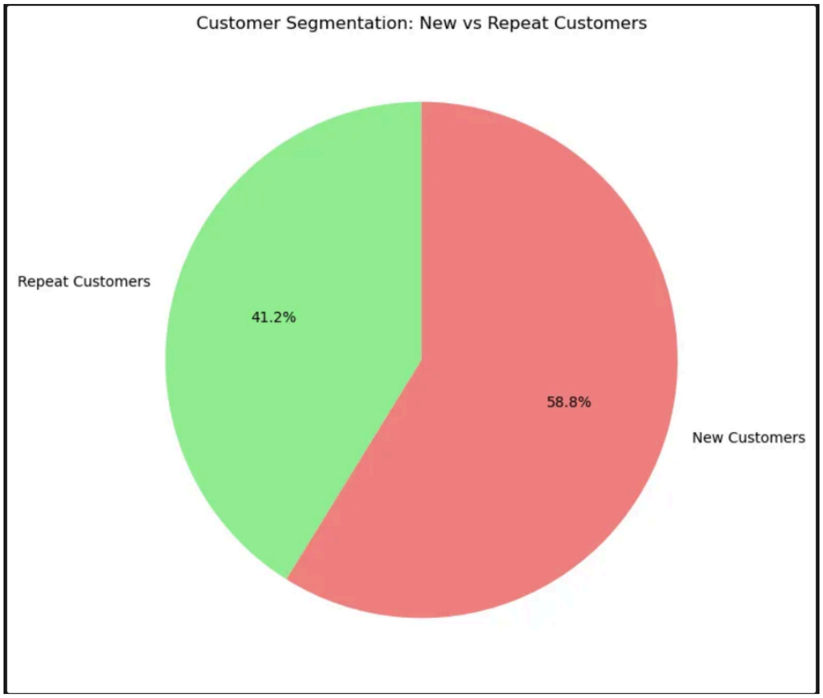
The graph illustrates monthly GMV trends for Luxury and Mass-Market products. Luxury products consistently dominate GMV, peaking around September-October 2023, while Mass-Market products contribute a smaller but steady share over time.



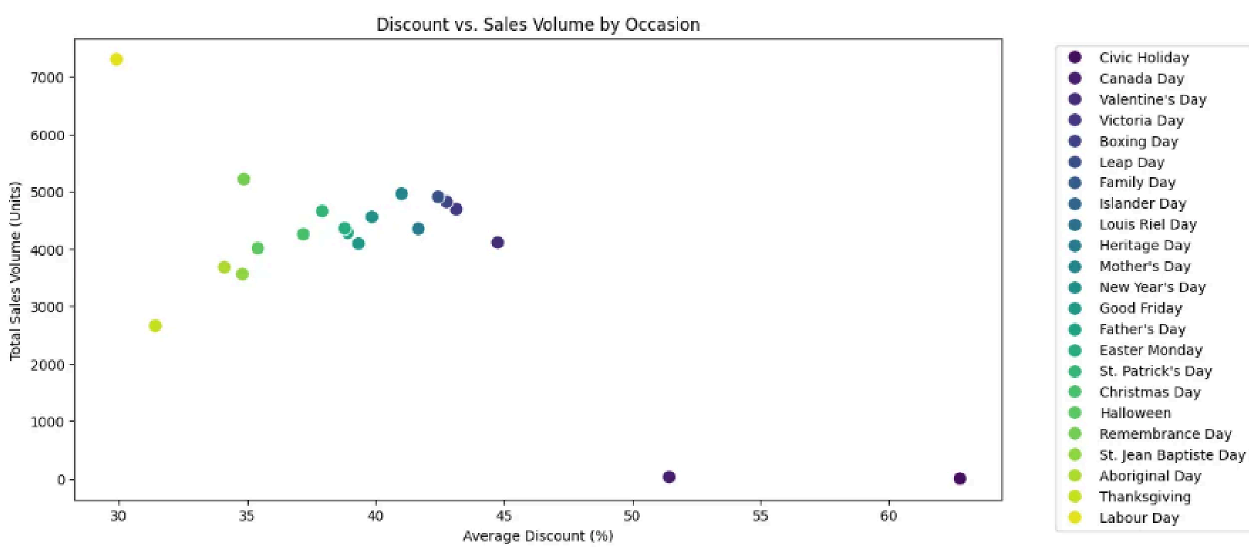
The chart compares product category sales on sale vs. non-sale days. Non-sale days dominate in all categories, especially "EntertainmentSmall," which has the highest total sales. Sales increase on sale days but remain significantly lower than non-sale days.



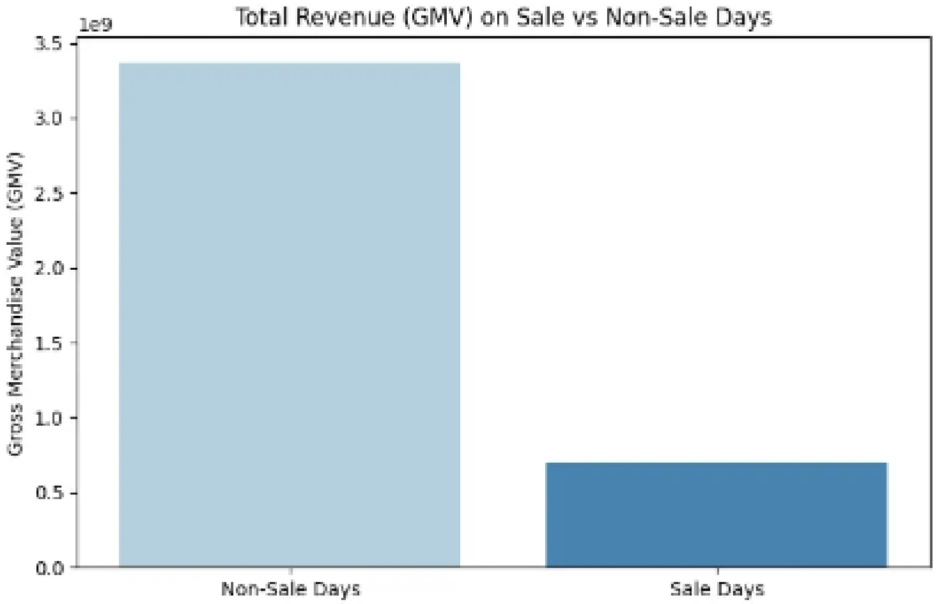
The bar chart shows the top 10 product verticals by GMV during sales. **DSLRs** lead by a huge margin, followed by **Mobile Speakers and Gaming Consoles**. Other categories like **Point & Shoot cameras, Laptop Speakers, and Home Audio Speakers** also contribute significantly.



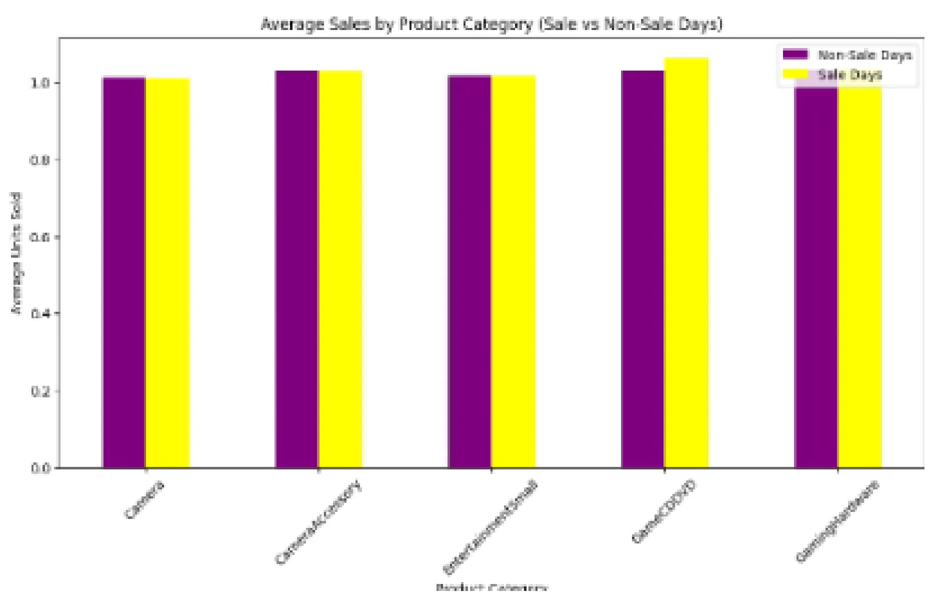
The pie chart represents customer segmentation, with **58.8% new customers** and **41.2% repeat customers**. This indicates that while new customers form the majority, a significant portion of the business comes from repeat buyers, highlighting customer retention.



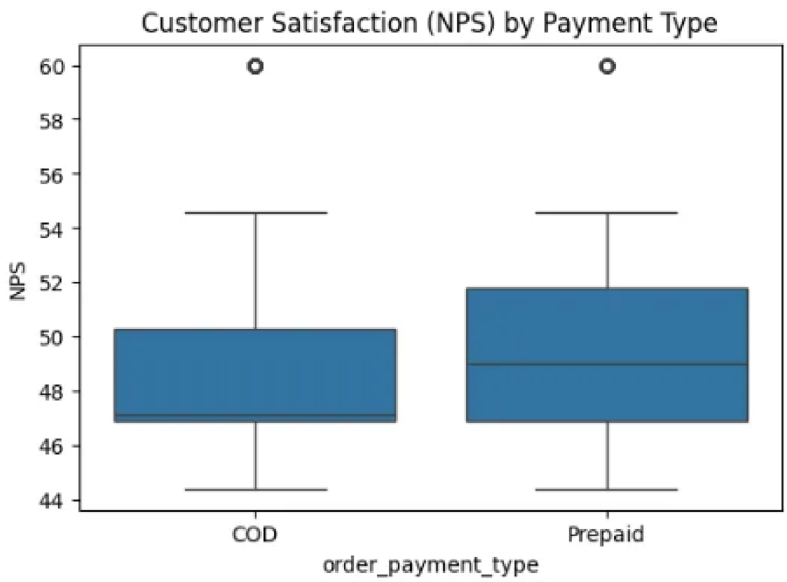
The scatter plot shows **sales volume vs. discount percentage** across different occasions. **Higher sales** occur mostly at **30-45% discounts**, while extreme discounts (**50%+**) see lower sales. Certain holidays, like **Labour Day and Aboriginal Day,** have notably high sales.



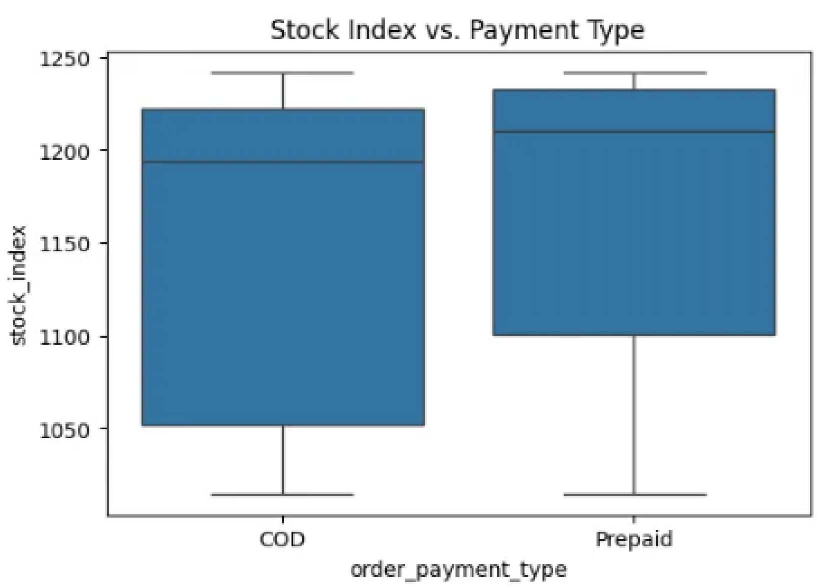
The chart shows that **Non-Sale Days generate much higher GMV** than Sale Days. Despite discounts, **Sale Days contribute less to total revenue**. This suggests that regular days drive more stable and significant business revenue.



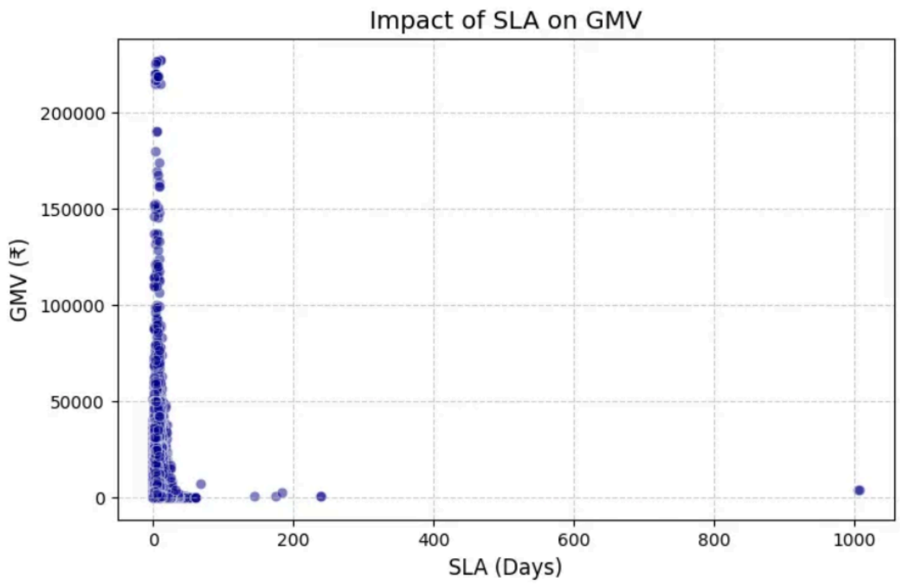
The bar chart compares **average sales across product categories on Sale vs. Non-Sale Days**. Sales remain nearly the same for most categories, with **Non-Sale Days (purple) slightly higher in some cases**. This suggests that sales events don't significantly boost average unit sales across categories.



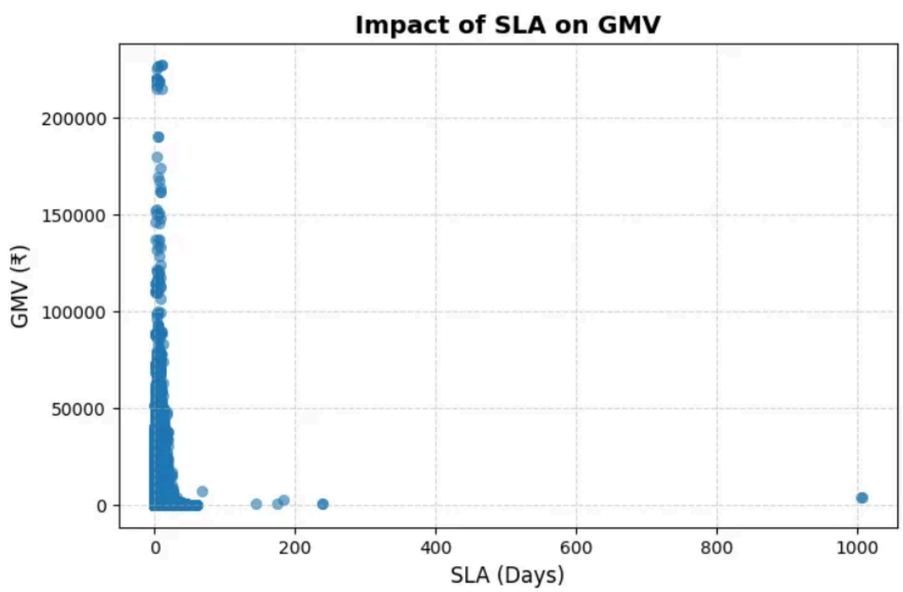
The box plot compares **Customer Satisfaction (NPS)** for **COD vs. Prepaid** payments. **Prepaid orders have a slightly higher median NPS**, indicating better satisfaction. However, both payment types show similar overall distributions with some outliers.



The box plot compares the **Stock Index** for **COD vs. Prepaid** payment types. Both have similar upper limits, but **COD shows a slightly lower median stock index** with more variation. Prepaid orders have a more consistent stock index distribution.



The scatter plot shows the **impact of SLA (Service Level Agreement) on GMV (Gross Merchandise Value)**. **Shorter SLAs correspond to higher GMV**, while longer SLAs see a sharp decline in sales. This suggests that faster delivery improves revenue.



The scatter plot illustrates the **impact of SLA (Service Level Agreement) on GMV (Gross Merchandise Value)**. **Lower SLA durations correlate with higher GMV**, while longer SLAs result in significantly lower sales. This highlights the importance of fast delivery in driving revenue.