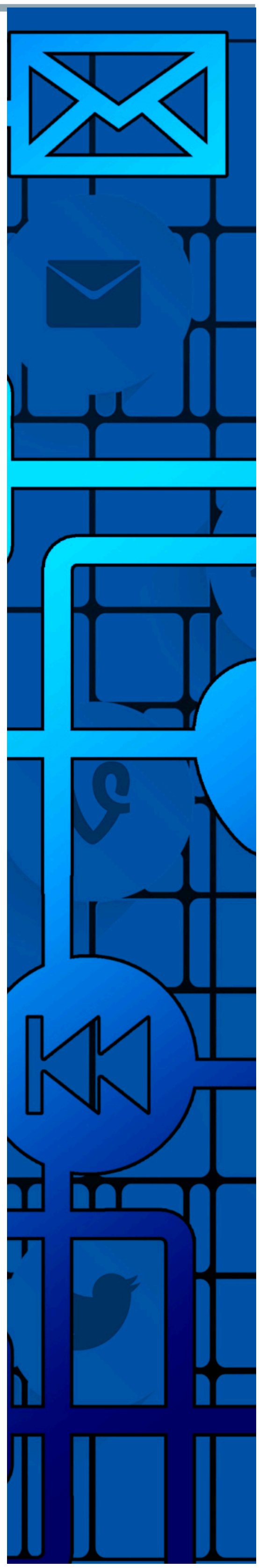
Team 2025102

# INDEX



# ABSTRACT


**ElectroMart**, an Ontario-based e-commerce company, is reviewing its marketing budget after the CFO expressed concerns about insufficient return on investments, prompting this analysis to support the marketing team in optimizing investment allocation.

## BUSINESS PROBLEM STATEMENT

The primary objective is to assess the effectiveness of past marketing expenditures and identify the most impactful marketing levers to enhance revenue. This involves analyzing key performance indicators (KPIs) that drive revenue, quantifying the return on investment (ROI) for each marketing channel, and determining the optimal budget allocation for the upcoming year. The key objectives are:

- **Performance Driver Analysis:** Identify KPIs that significantly influence revenue.
- **Marketing ROI Impact Analysis:** Quantify each channel's effect to revenue.
- **Optimizing Marketing Spend:** Allocate the budget strategically to maximize revenue.

## DATASET DESCRIPTION

The datasets for ElectroMart include several key files that collectively support a comprehensive analysis of marketing effectiveness and budget optimization:

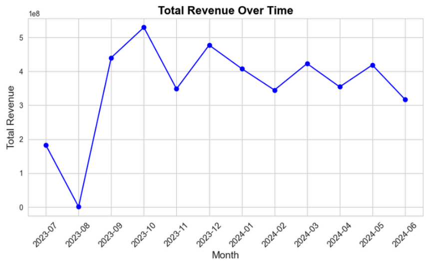
- **Customer Orders File**: Contains order details from July 2023 to June 2024, including order dates, FSN IDs, GMV, units sold, payment types, and delivery timelines.
- **Media Investment File**: Provides monthly spending data on various advertising channels, enabling an assessment of marketing ROI across different media channels.
- **Sale Calendar File**: Lists dates when special sales or promotions were offered, helping to analyse their impact on revenue.
- **NPS File**: Includes monthly Net Promoter Scores and company stock values, which can be correlated with marketing efforts.
- **Weather File**: Includes detailed weather reports for Ontario, Canada, from 2023 to 2024, which can help assess environmental influences on sales.
- **SKU Mapping** (Product Hierarchy Details): Offers insights into product categorization, including super categories, categories, sub-categories, and verticals.
- **Holidays List**: Provides a list of Canadian holidays during the data period, useful for understanding seasonal fluctuations in sales.

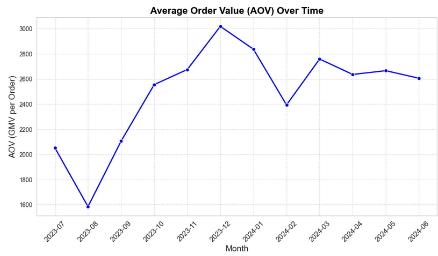
## SOLUTION OVERVIEW

Based on the data provided, we identified five **Key Resultant Areas** (KRAs) for in-depth analysis, conducting EDA to derive insights that shaped KPIs and KRIs for performance evaluation. Interactive dashboards integrate these findings for quick visualization and decision-making.

Using these insights, we engineer features impacting GMV and apply ML models to rank their importance. To optimize budget allocation, we assess marketing ROI and propose a Three-Tier Scenario Planning framework using the SLSQP algorithm. We also offer channel-level recommendations and future research directions.

# PRELIMINARY EXPLORATORY DATA ANALYSIS




## Total Revenue Over Time:

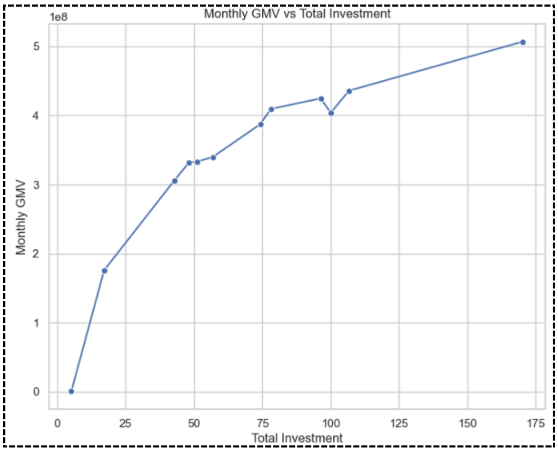
Revenue surged in September 2023, peaked in October, and showed fluctuations with a slight decline by May 2024.

## Average Order Value (AOV):

AOV increased steadily until December 2023, peaked, then declined and stabilized from February 2024.

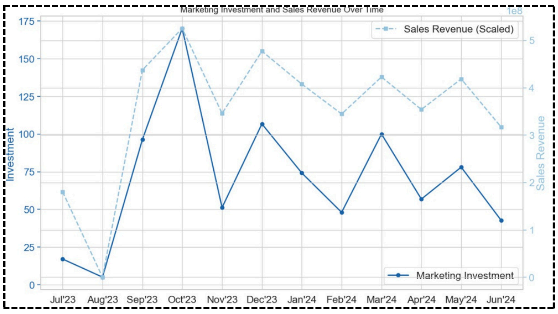
## Relationship between revenue & advertisement spends:

The plot shows a positive correlation between ad spending and revenue, but with diminishing returns. Revenue grows rapidly at lower ad spend levels, then slows down beyond a point. This suggests an optimal spending range where marketing is most effective, after which additional spending yields lower returns. Strategic budgeting is key for maximizing ROI.



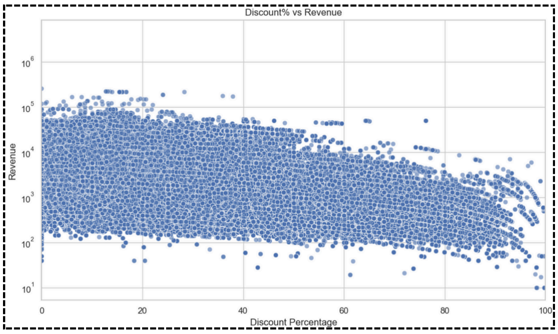

## Marketing Investment vs. Sales Revenue Trends:

When scaled to equal the maximum investment, revenue shows relative stability despite fluctuating investments, with visible lag effects.




## Effect of discounts on revenue:

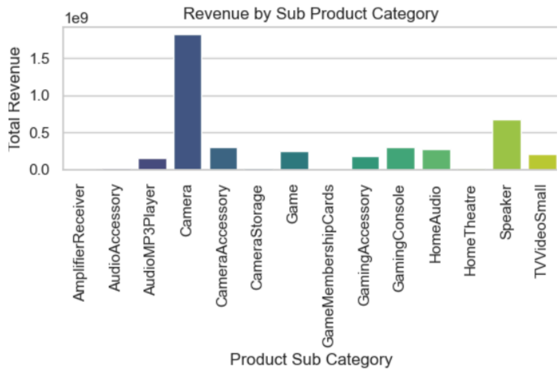
The plot shows the impact of discounts on revenue (GMV). It suggests that lower discounts (0-30%) correlate with higher revenue, while higher discounts reduce revenue potential. Excessive discounting may not always drive sales effectively and could lower overall revenue.

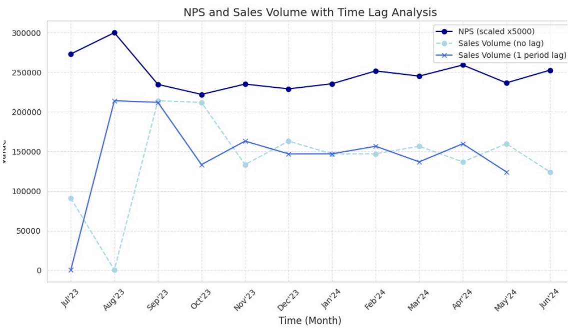


# PRELIMINARY EXPLORATORY DATA ANALYSIS

### Relationship between product categories and revenue:

The revenue by sub-category plot reveals clear differences across product segments. Notably, sub-categories like Camera and Camera Accessories (in particular DSLR) contribute disproportionately to total revenue.



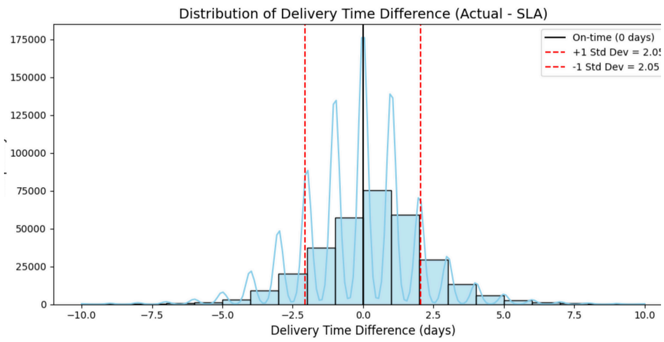



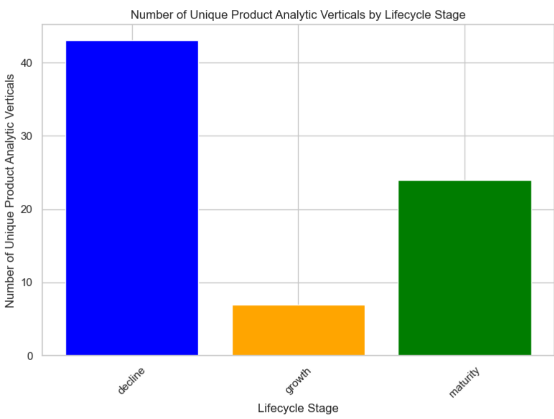

### Customer satisfaction scores driving future revenues:

NPS positively impacts future sales with a time lag. Higher NPS (Aug-Oct 2023) led to increased sales, while a decline (Sept-Oct 2023) resulted in lower sales. 1-period lag sales trend closely follows NPS, showing that satisfied customers drive revenue through repeat purchases and referrals.

### Delivery Time Variance :

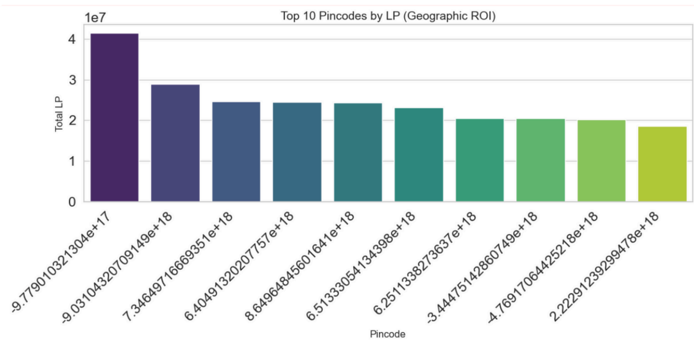
This histogram illustrates the variance between actual delivery time and the promised SLA (delivery_days – SLA). Most deliveries cluster around 0 days, indicating strong on-time performance. The ±1 standard deviation range (~±2 days) captures the majority of orders, reflecting relatively consistent delivery times.




### Analysis of number of products across different product lifecycle stages:

The product lifecycle analysis reveals a portfolio imbalance, with 43 verticals in decline, 24 in maturity, and only 7 in growth. This suggests over-reliance on aging products and a lack of innovation, possibly due to market saturation or evolving consumer preferences. The limited growth-stage presence highlights the urgent need for revitalization strategies to ensure long-term revenue sustainability.

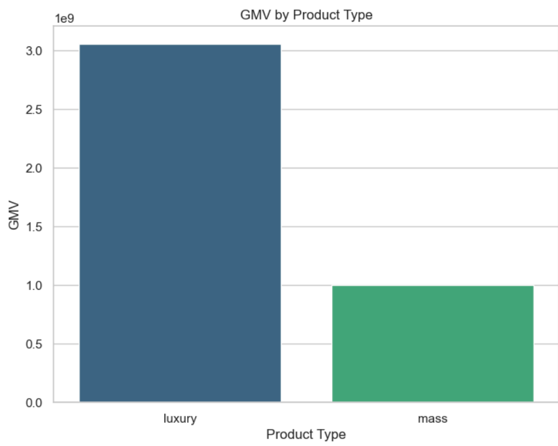



## Geographic ROI:

Identifying high-performing regions (or clusters of regions) allows us to focus marketing budgets and operational resources where we can generate the highest ROI.

## Revenue Concentration Analysis:

Luxury products, defined as those above the 80th percentile of GMV, generate approximately 75% of the total GMV (~3.0 billion) despite representing only 20% of products. This concentration exceeds the traditional 80/20 Pareto principle. Mass products contribute ~1.0 billion GMV, bringing the total to ~4.0 billion. The data highlights significant revenue concentration in high-value products, offering insights for optimizing inventory, marketing, and pricing strategies.




Apart from conducting a preliminary EDA, we identified 5 Key Resultant Areas (KRAs) for a comprehensive analysis. Additionally, we conduct a KRA-level analysis. For each KRA, further EDA was conducted, yielding meaningful insights that informed the development of KPIs and KRIs for performance driver analysis. The 5 KRAs are:

| S.No. | Key Resultant Area (KRA) | Description of the KRA |
|---|---|---|
| 1 | Product Portfolio Performance & Profitability Analysis | Evaluates performances of different products |
| 2 | Marketing Effectiveness Evaluation | Evaluates investment in different marketing levers |
| 3 | Customer Experience and Satisfaction Analysis | Analyzes customer trends |
| 4 | Revenue Growth and Sales Performance Analysis | Analyzes performance using sales trends |
| 5 | External Factors Analysis | Evaluates the effect of factors like weather and holidays |

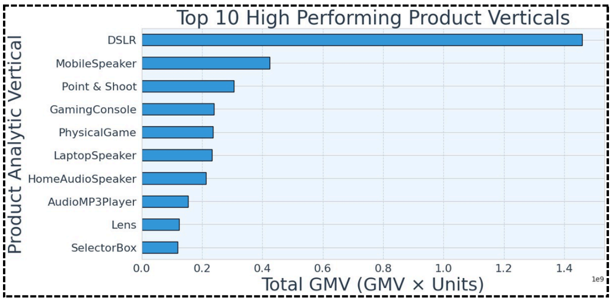
We report the additional EDA insights with KPIs and KRIs development for each of the 5 KRAs in the following pages.

# KRA 1: PRODUCT PORTFOLIO PERFORMANCE & PROFITABILITY

## EDA OF KRA 1

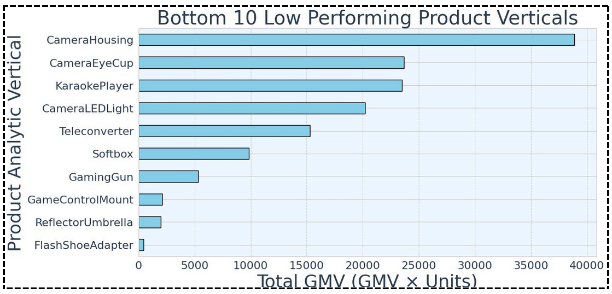
**High-Performing Product Verticals**

- DSLR leads in GMV and MobileSpeaker, Point & Shoot also perform well.
- Strong sales in gaming and audio suggest continued investment in these segments for growth.



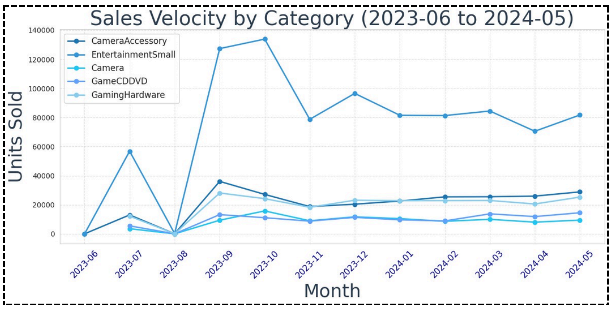

**Low-Performing Product Verticals:**

- FlashShoeAdapter has the lowest GMV among underperforming products.
- Camera accessories dominate the list, suggesting a need for bundling strategies to boost demand.

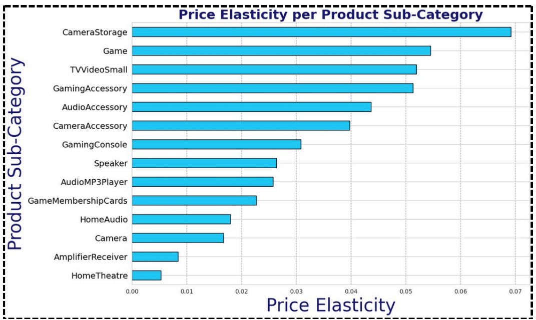


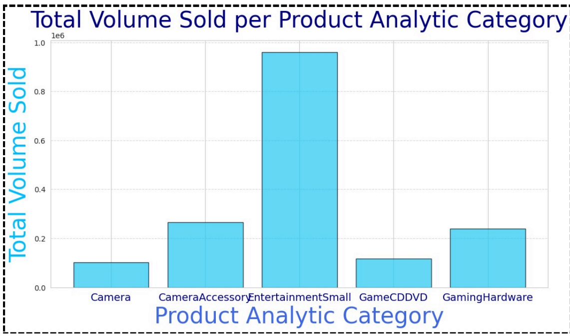

**Sales Trends by Category:**

- EntertainmentSmall saw a sharp spike in sales around September 2023 but later stabilized.
- Other categories showed gradual growth, with GamingHardware and CameraAccessory maintaining low steady sales volumes over time.




## UNDERSTANDING KPIs




**Sales Volume Distribution by Category:**

- **EntertainmentSmall** leads in total volume sold
- **CameraAccessory** and **GamingHardware** have moderate sales, while Camera and GameCDDVD show the lowest volume, indicating potential market limitations or lower demand.

**Price Sensitivity (Product Sub-Category):**

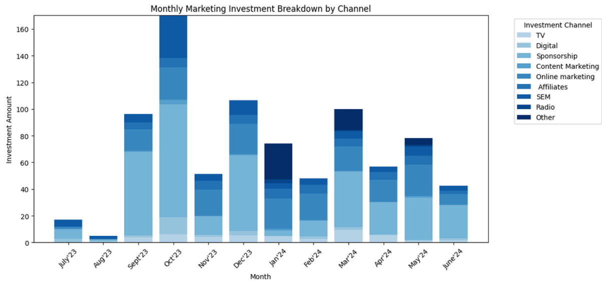
- **Camera Storage** and **Game** exhibit high price elasticity, indicating demand changes with price shifts
- In contrast, Home Theatre and Amplifier Receiver have low elasticity, suggesting stable demand despite price fluctuations.

# KRA 2: MARKETING EFFECTIVENESS EVALUATION

## EDA OF KRA 2

### Investment Allocation by Channel:

- October 2023 had the highest investment, mainly in TV and Sponsorship.
- Investments dropped in November, then peaked again in January with higher Affiliate spending.
- Online and Sponsorship dominate overall, while Content Marketing and SEM receive lower allocations.

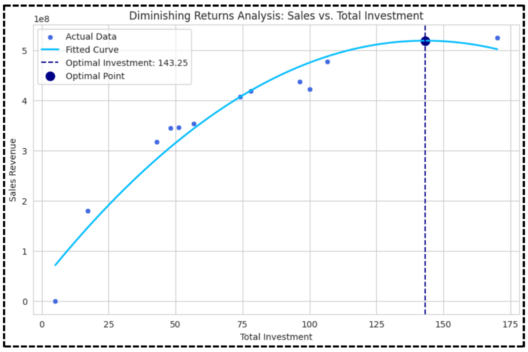


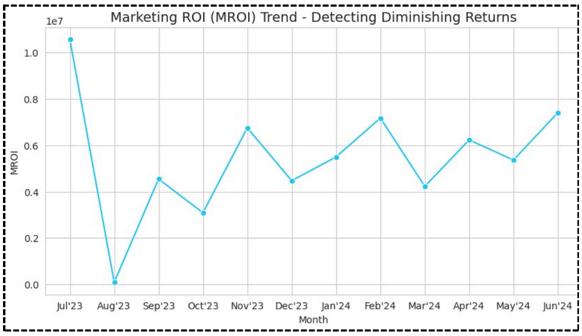


### Total Marketing Investment Trend:

- Marketing investment peaked in October 2023, before dropping sharply in November.
- Another peak occurred in December 2023 and March 2024, suggesting seasonal spending patterns.
- The overall trend shows fluctuations with a gradual decline after March 2024, possibly due to budget adjustments or strategy shifts.

## UNDERSTANDING KPIs




### Diminishing Returns Analysis

- Sales revenue increases with investment but plateau after 143.25, indicating diminishing returns.
- Beyond this point, additional spending yields minimal revenue growth.
- The optimal investment level is 143.25, maximizing revenue before efficiency drops.
- Spending beyond this leads to overspending without significant sales improvement.

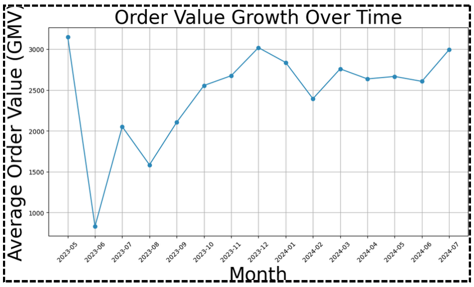
### Marketing Return On Investment:

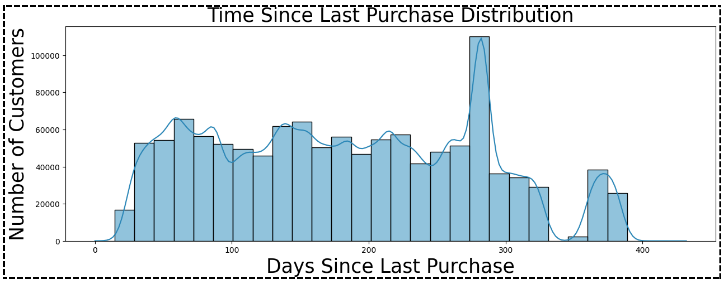
- **Formula:** (Sales Revenue - Total Investment)/Total Investment.
- MROI is declining, indicating diminishing returns on spend.
- A negative correlation (-0.30) between investment and MROI suggests inefficiency.
- Five months of decline highlight saturation risk.
- Increasing investment does not translate into proportional growth.

# KRA 3: CUSTOMER EXPERIENCE AND SATISFACTION ANALYSIS

## EDA OF KRA 3

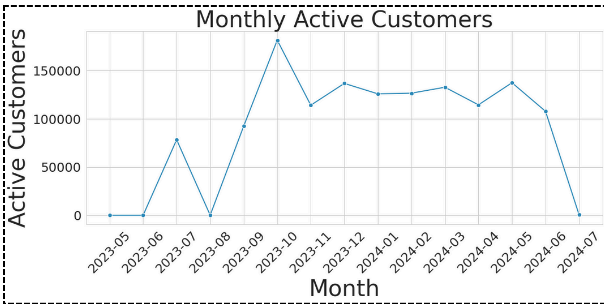
AOV peaked in May 2023, dropped in June, then fluctuated until October before rising steadily through December. After a February dip, it showed stable growth from March to July 2024.



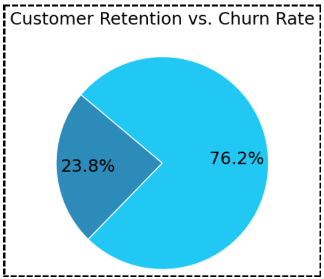



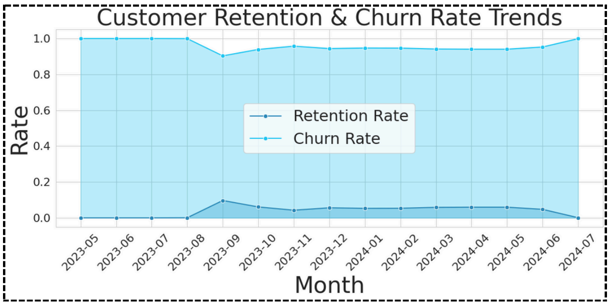

## UNDERSTANDING KPIs

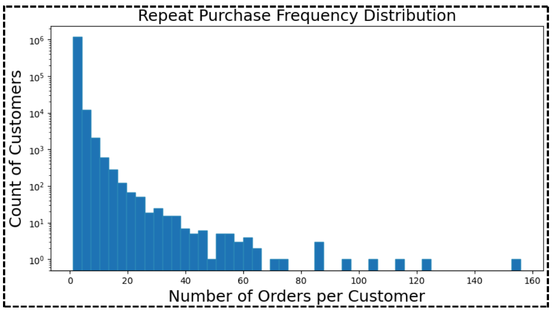
Customers showed increased engagement and purchasing activity from mid-2023, peaking in October, likely due to promotions or seasonal demand. However, the recent decline suggests potential issues with product appeal, pricing, or competition.




The low retention rate (23.8%) and high churn rate (76.2%) of last 180 days indicate poor customer loyalty, likely due to lack of engagement, competition, or dissatisfaction. To improve retention, ElectroMart should implement loyalty programs, personalized recommendations, and proactive customer engagement, while enhancing support to address issues quickly.



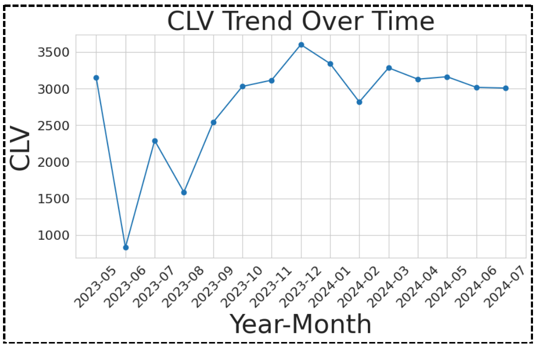



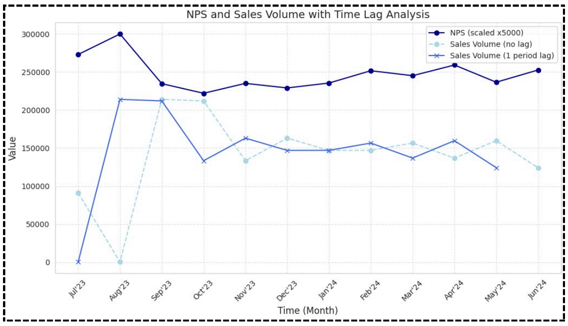

The histogram shows that most customers are one-time buyers, indicating a high influx of first-time shoppers who don't return. Repeat purchases decline exponentially, following a power-law distribution where a small group of loyal customers drive sales.

A few high-value customers place 60+ or even 100+ orders, making them key repeat buyers.

# KRA 3: CUSTOMER EXPERIENCE AND SATISFACTION ANALYSIS

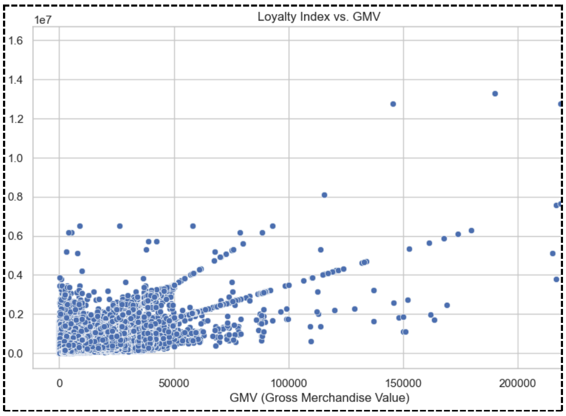
**Customer Lifetime Value (CLV):**
The CLV of last 180 days is around 4223. The recovery and stabilization from August 2023 onward, indicating improvements in customer engagement and retention strategies. The steady trend from early 2024 highlights consistent customer satisfaction and sustained loyalty.

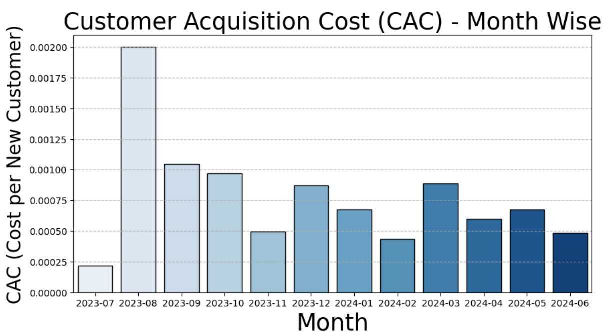



**Net Promoter Score (NPS):**
The chart shows that sales volume trends follow NPS with a slight lag, suggesting a direct impact of customer satisfaction on sales. As NPS remains strong, sales tend to stabilize, reinforcing the importance of maintaining high customer satisfaction for sustained growth.

**Customer Loyalty and GMV Relationship:**
The scatter plot shows a positive but non-linear correlation between loyalty and spending. Most customers have low GMV (<50,000) and loyalty indices (<0.4), while high-value outliers (GMV >150,000) exhibit exceptional loyalty (>1.2M). This suggests factors like recency, frequency, and AOV also drive loyalty, highlighting segmentation opportunities.




**Customer Aquisition Cost (CAC):**
The declining CAC suggests improved marketing efficiency, allowing more focus on customer experience. Lower costs indicate stronger organic engagement and higher satisfaction. The sharp drop after August 2023 may reflect better onboarding and engagement strategies.

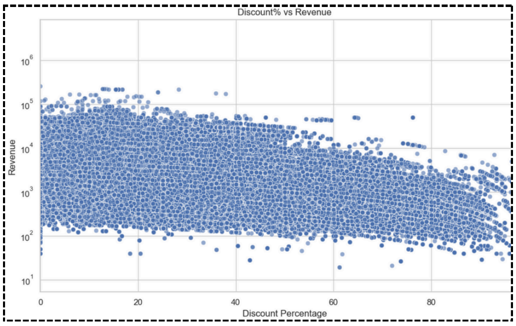
# KRA 4: REVENUE GROWTH AND SALES PERFORMANCE ANALYSIS

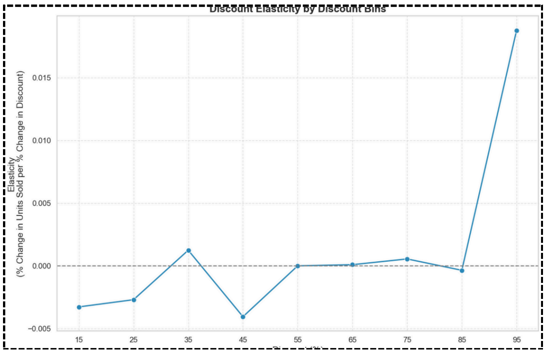
## UNDERSTANDING KPIs

### Discount Effectiveness :

**Formula:** Discount Effectiveness = (Increase in Sales Quantity) / (Discount Percentage)

Shows diminishing returns beyond a 40% discount. While initial discounts drive higher sales, deeper discounts yield lower proportional sales growth, making them less effective in maximizing revenue.

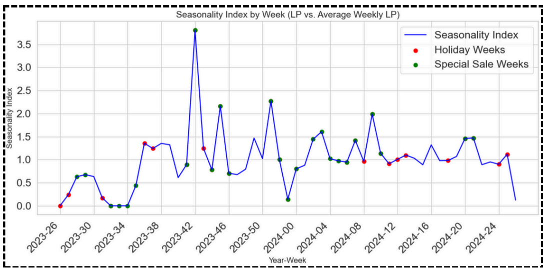



### Discount elasticity :

It is calculated as the **percentage change in units sold** divided by the **percentage change in discount**

Elasticity is minimal or negative at moderate discounts but **spikes sharply around 95%,** suggesting customers become highly sensitive at steep discount levels.




### Seasonality Index :

It is calculated as the **Weekly List Price** divided by the **Average Weekly List Price** across all weeks

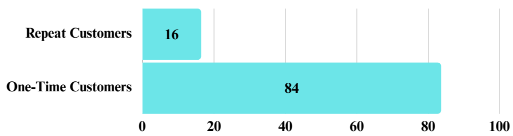
Peaks occur during holidays and special sales, while stable and low-index periods indicate off-season slumps

Overall, the data suggests cyclical sales behavior aligned with key events.

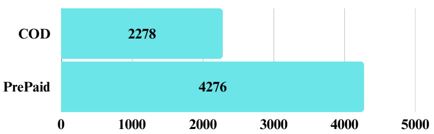
## INTERESTING INSIGHTS

**64% On-Time Deliveries**

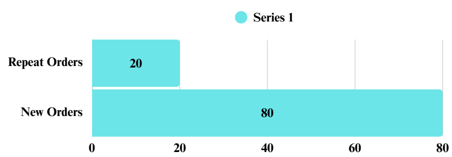
**On-Time Delivery Rate :** No. of Orders with deliverycdays<=sla divided by Total Orders*




**Repeat Customer Rate**




**Payment Method Efficiency :** Average List Price per order

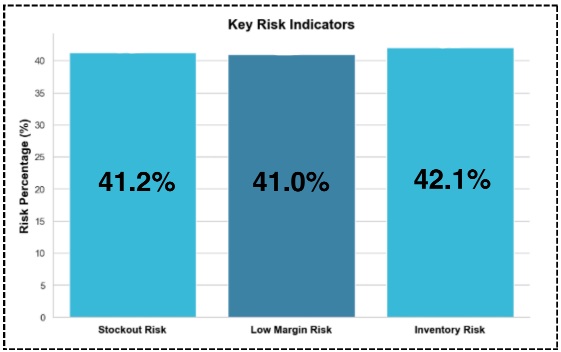



**Repeat Order Rate :** Percentage of orders placed by returning customers

*(Assumption: For further analysis, rows imputed with 0 delivery days were dropped since they hampered the analysis)

# KRA 4: REVENUE GROWTH AND SALES PERFORMANCE ANALYSIS

## UNDERSTANDING KRIs




**Low Margin Risk (%)**

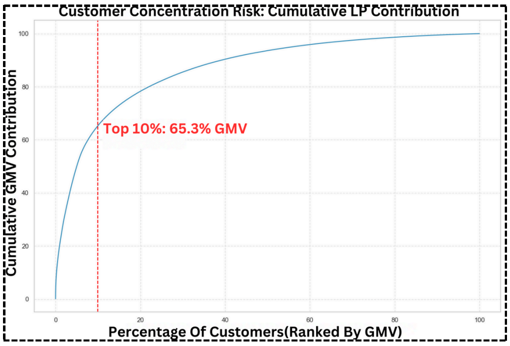
- Formula : Number of products with discount percentage > 50%) / Total number of products × 100
- Indicates the percentage of products being sold at high discounts, potentially reducing profit margins.

**Stockout Risk**

- S(i) for a product in the Camera or Speaker (they were the top two categories) category during special sale or holiday weeks is defined as:
- S(i) = 1, if SLA(i) > median_SLA(C)
- S(i) = 0, otherwise.
  - S(i) represents the stockout risk indicator (1 = high risk, 0 = low risk)
  - SLA(i) is the procurement SLA (Supply Lead Time) of product i
  - median_SLA(C) is the median proc. SLA for Camera & Speaker categories

**Inventory Risk (%)**

- Formula : (Number of products with procurement SLA above median) / Total
- This identifies the percentage of products with long procurement SLAs, increasing the risk of supply chain delays.

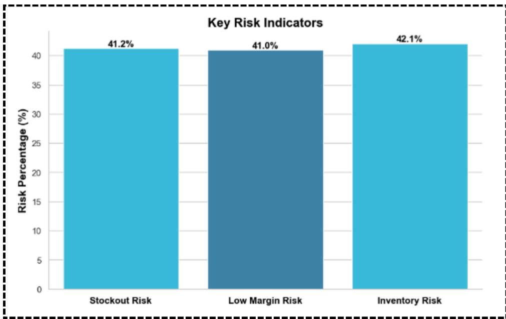



**Customer Concentration Risk**

- Roughly 65% of total List Price comes from the top 10% of customers, indicating a high revenue concentration among a small group.
- This imbalance suggests the business is vulnerable if these key customers reduce or stop purchasing.

**Low Margin Risk (%)**

- Formula : Number of products with discount percentage > 50%) / Total number of products × 100
- Indicates the percentage of products being sold at high discounts, potentially reducing profit margins.




# KRA 5: EXTERNAL FACTORS ANALYSIS

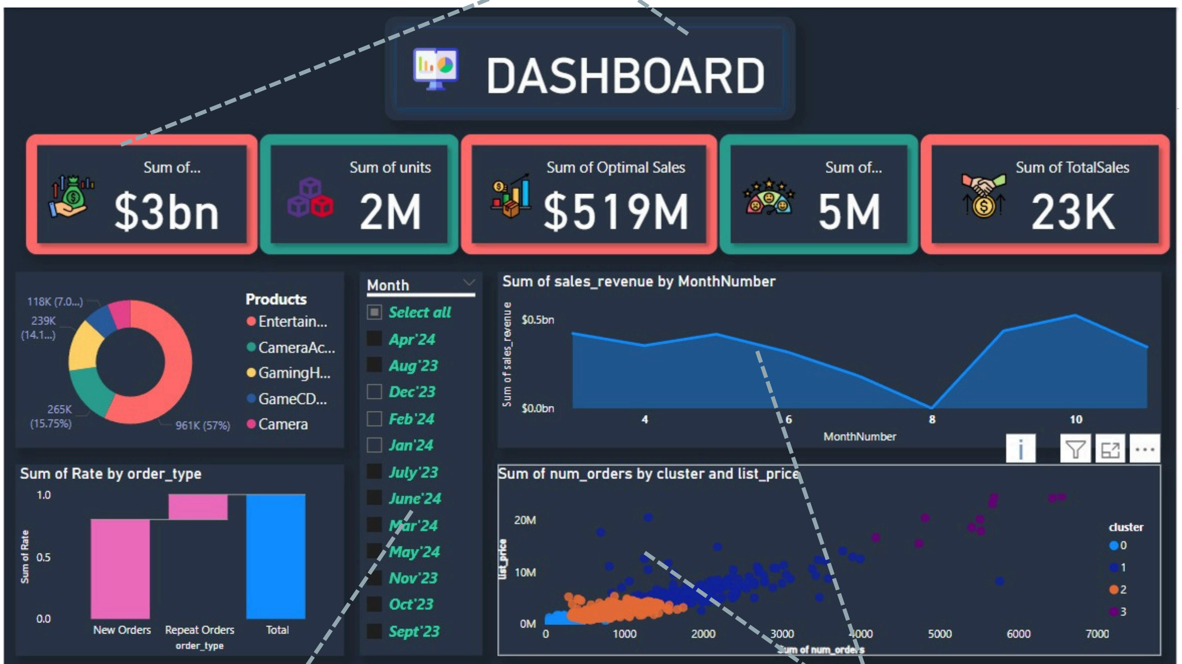
Refer annexure pg-20, 21 for detailed analysis of external factors like temperature, precipitation, holidays etc.

# VISUALISATION USING INTERACTIVE DASHBOARDS

### Why Use Power BI for Analysis?

Power BI enables interactive data visualization, real-time insights, seamless integration, automated reporting, and advanced analytics, enhancing decision-making efficiency and accuracy.

In Power BI, KPI Box (alternative for Card) highlights key metrics, while Header Label (alternative for Title) improves clarity. These enhance readability, maintain a structured layout, and support dynamic formatting. Customizable colors improve visualization, branding consistency, and user focus using themes and conditional formatting for better data insights.

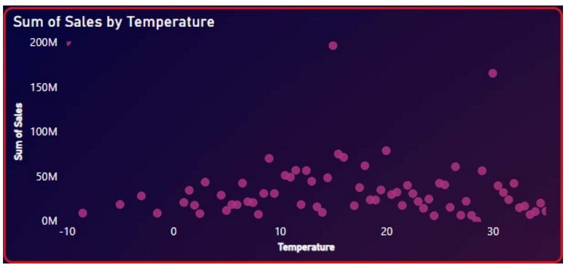



A Slicer in Power BI is an interactive filter that enables users to refine data, apply multi-selection, sync across pages, and enhance dashboard interactivity and usability.

Dynamic graphs in Power BI update automatically based on filters, slicers, and real-time data changes, providing interactive, visually adaptive insights for better decision-making and analysis.

### Benefits of Interactive Graphs

Interactive graphs help businesses by enabling real-time data exploration, trend analysis, quick decision-making, deeper insights, and improved storytelling, leading to better strategic planning and efficiency.



# FEATURE ENGINEERING

| Feature | Explaination |
|---|---|
| Discount% | 100*(product_mrp – list price) / product_mrp |
| Holiday Week Flag | Binary flag identifying orders within ±3 days of holidays |
| Month Extraction | Month from order date as numeric value (1-12) |
| Day of Week | Weekday from order date as numeric value (0-6, where 0=Monday) |
| Quarter | Fiscal quarter (1-4) derived from order date |
| Weekend Indicator | Binary flag (0/1) for weekend orders |
| Month Sine/Cosine Transformations | Cyclical month_sin/month_cos features preserving circular month relationships |
| Day of Week Sine/Cosine Transformations | Cyclical dow_sin/dow_cos features preserving circular weekday relationships |
| Monthly Category GMV | Monthly GMV aggregated by product category for time-series analysis |
| Channel-Category Elasticity | Responsiveness of category GMV to channel spend: (Δ%GMV)/(Δ%spend) |

| Feature | Explaination |
|---|---|
| Previous Average Units | Mean units/product in preceding 3-month period |
| Growth Rate | Percentage change between recent vs. previous 3-month average units sold |
| Lifecycle Stage Classification | Product categorization: growth (>20%), maturity (0-20%), decline (<0%) based on growth rate |
| Payday Week | If Payday falls within the week, then payday week =1, else 0 |
| Recency | $R_i = T_{max} - t_i^*$ (in days) |
| Frequency $F_i$ | Number of unique order_id |
| Monetary | $M_i = \sum_{j \in O_i} gmv_{ij}$ |
| Average Order Value | $A_i = \frac{M_i}{F_i}$ |
| Loyalty Index | $L_i = \left(0.35 \cdot F_i + 0.35 \cdot A_i + 0.30 \cdot \frac{1}{R_i + 1}\right) \times 100$ |
| Recent Average Units | Mean units/product in most recent 3-month period |
| Product Type | Binary classification: luxury (1) if GMV > 80th percentile, else mass-market (0) |

# MODELLING

We applied various models to the features obtained through feature engineering, among which XGBoost performed the best. The models that were applied

- Deep Neural Network (DNN)
- LightGBM
- AdaBoost, XGBoost, CatBoost
- Stacked Model: Base:[XGBoost+Catboost], Meta-model:[Linear Regeressor]

## XGBOOST

- XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework.
- We calculated the mean SHAP values, which refers to the average absolute contribution of each feature across all samples, for determining the relative importance of each feature in case of the XGBoost model applied.

$$\text{Mean SHAP}(j) = \frac{1}{N}\sum_{i=1}^{N} |SHAP_{i,j}|$$

*Refer annexure pg-19 for graph

## MODEL COMPARISIONS

We have done a detailed analysis of all the models applied on various parameters:

- RMSE (Root Mean Squared Error)
- SMAPE (Symmetric Mean Absolute Percentage Error)
- Median Absolute Error
- Mean Absolute Error (MAE)

*Refer annexure pg-19 for comparison table

# BUDGET OPTIMIZATION

## BUDGET OPTIMIZATION FOR MAXIMIZING REVENUE

**Objective:**

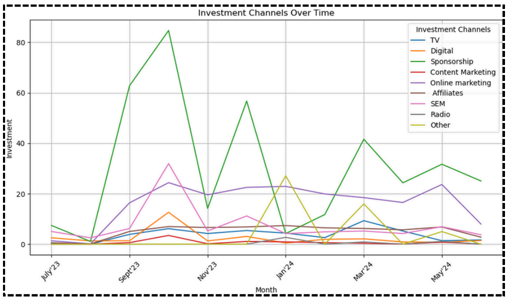
The goal of budget optimization is to strategically reallocate investments across different channels to maximize overall revenue while considering past information.

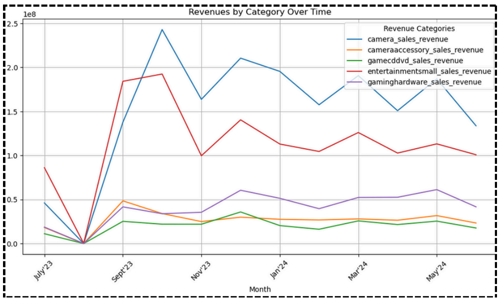
**Approach:**

To achieve this, we employ a Three-Tier Scenario Planning framework, which evaluates investment strategies for the period July 2024 – June 2025 across three scenarios:

- **Tier 1:** Conservative – Prioritizes proven high-ROI channels with minimal risk
- **Tier 2:** Balanced – Adjusts investments based on seasonal trends and past
- **Tier 3:** Growth – Allocates more funds to emerging and high-potential channels

Each scenario is analyzed using ROI-based optimization, ensuring the most effective allocation of funds while respecting budget constraints.



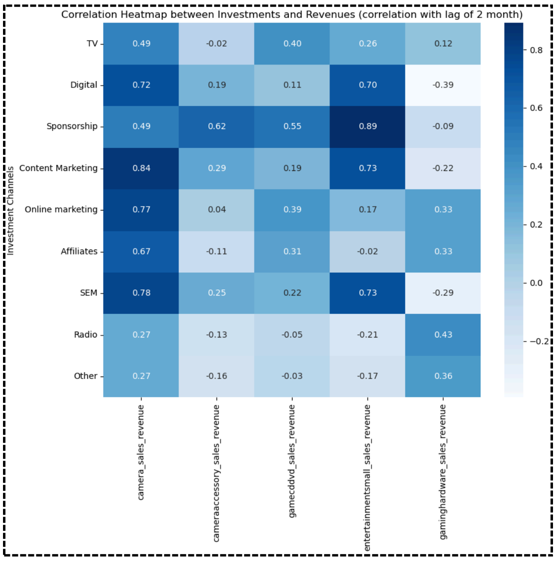

**ROI estimator:** As we can't directly calculate the ROI for every channel of investment, we take the weighted average of the effect of a channel of investment on the total revenue through every revenue catagory.

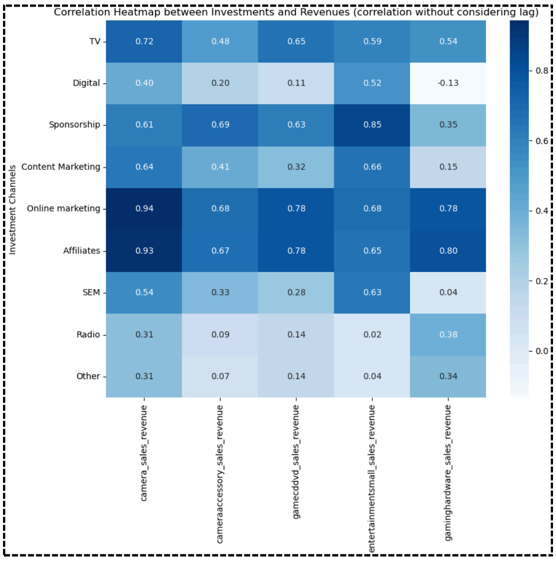
$$\text{ROI Estimator} = \sum_{j=1}^{M} R_j \cdot \frac{C_{ij}}{\sum_{k=1}^{N} C_{kj}}$$

**Lagged analysis**(Hypothesis testing): To consider the effect of the investment on the revenue after a lag of n months. The weighted investment, $I'_t = I_t + \gamma \cdot I_{t-1} + \gamma^2 \cdot I_{t-2}$

## CHANNEL RECOMMENDATIONS

- **TV**
  - Insight: Stable input and returns; Recommendation: Maintain stable investment unless experimentation is viable
- **Digital**
  - Insight: Lag analysis shows a 2-month correlation with Camera_sales and Entertainment_small sales; Recommendation: Stable investment for the long term (Tier 1); increase before predicted sales spikes (Tier 2)
- **Sponsorship**
  - Insight: Limited impact on Entertainment_small sales despite alternating investment; Recommendation: Maintain balanced investment at a lower level
- **Content Marketing**
  - Insight: Strong 1-2 month lagged impact on Camera sales and Entertainment small sales; Recommendation: Increase investment 2 months before predicted sales spikes (Tier 3)




- **Online Marketing**
  - Insight: Strongly drives Camera_sales with stable, high investment levels; reduced correlation with revenue in 1-2 month lag analysis
  - Recommendation: Maintain stable, high investment (Tier 1); reduce before predicted dips in Camera_sales
- **Affiliates**
  - Insight: Similar to Online Marketing but with lower investment levels; reduced correlation with revenue in 1-2 month lag analysis
  - Recommendation: Increase investment if surplus funds are available; reduce before predicted dips in Camera_sales
- **SEM**
  - Insight: Performs well with 2-month lag in Camera_sales and Entertainment_small sales; inconsistent otherwise
  - Recommendation: Stable, low investment for steady performance (Tier 1 & 3); increase before predicted sales spikes (Tier 2)
- **Radio and Other**
  - Insight: Limited data prevents strong conclusions
  - Recommendation: Maintain low investment; experiment if funds allow.

**Optimization of Investment Allocation for Maximum ROI**:

This optimization determines the best investment distribution to maximize ROI estimator. The objective function models ROI using revenue ratios and correlation factors. Constraints ensure allocations sum to 1 and stay within historical bounds. The problem is solved using SLSQP, which efficiently handles nonlinear constraints.

$$\max_{p} \sum_{i=1}^{N} p_i \cdot \sum_{j=1}^{M} R_j \cdot \frac{C_{ij}}{\sum_{k=1}^{N} C_{kj}} \quad \text{; subject to } \sum_{i=1}^{N} p_i = 1 \text{ , } L_i \leqslant p_i \leqslant U_i \; \forall i$$

N = Number of investment channels; M = Number of revenue categories
$p_i$ = Proportion of investment allocated to channel i
$R_j$= Mean revenue ratio for category j
$C_{ij}$ = Correlation between investment channel i and revenue category j
$L_i$ = Minimum observed historical proportion for channel i
$U_i$ = Maximum observed historical proportion for channel i

**Future Prospect**:
- Analysis of the diminishing returns

$x_i$ : Investment amount in channel i (for i = 1, 2, ...9),
$ROI_i(x_i)$ : Expected ROI estimator for channel i given investment $x_i$
B : Total budget
For each channel, we can model the expected ROI using diminishing returns function:
$ROI_i(x_i) = a_i \cdot \left(1 - e^{-b_i \cdot x_i}\right)$
To encorporate laf effects, we can modify this function to include weighted investments:

$$ROI_i(x_i, x_{i,t-1}, x_{i,t-2}) = a_i \cdot \left(1 - e^{-b_i \cdot (x_i + \gamma \cdot x_{i,t-1} + \gamma^2 \cdot x_{i,t-2})}\right)$$

where: $x_{i,t-1}$ , $x_{i,t-2}$ represent investments in previous periods & $\gamma$ is decay factor

**Optimization Problem:**
The optimization problem can be formulated as:

maximize: $ROI_{total} = \sum_{i=1}^{9} x_i \cdot ROI_i(x_i, x_{i,t-1}, x_{i,t-2})$

subject to: $\sum_{i=1}^{9} x_i$ = B(budget constraint)

$x_i \geqslant 0$; potentially; $x_i \leqslant max_i$(maximum allocation constraints) for all i
$x_i \geqslant min_i$ (minimum allocation constraints) for all i

## Results:

The Sequential Least Squares Quadratic Programming (SLSQP) method was used to determine the optimal investment proportions across various channels. Below are the results sorted in descending order of proportion:

- Online Marketing: 0.413976 ; Sponsorship: 0.300435; Affiliates: 0.134542; TV: 0.092977; SEM: 0.051923; Digital: 0.006146; Content Marketing: 0.000001; Other: 0.000000; Radio: 0.000000

The sum of all proportions is 1.0, ensuring a balanced allocation.

*example in annexure pg-25

## Dynamic Investment Planning and Forecasting:

The framework dynamically adjusts allocations using real-time revenue updates and past 12-month correlations to maximize ROI. With sufficient data, revenues are modeled as functions of investment channels, enabling predictive financial planning. This adaptive strategy ensures optimized returns and sustained growth.

# ANNEXURE

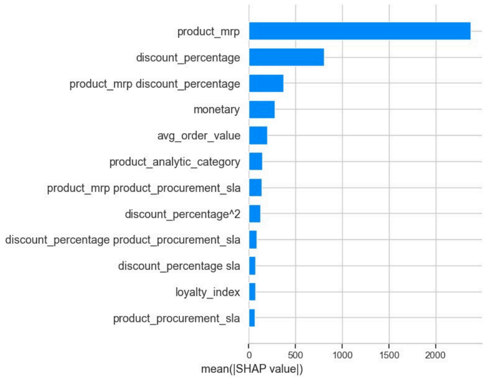
# MODELLING

We applied various models to the features obtained through feature engineering, among which XGBoost performed the best. The models that were applied are mentioned below:

- Deep Neural Network (DNN)
- LightGBM
- AdaBoost, XGBoost, CatBoost
- Stacked Model: Base:[XGBoost+Catboost], Meta-model:[Linear Regeressor]

# XGBoost

- XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework.
- We calculated the mean SHAP values, which refers to the average absolute contribution of each feature across all samples, for determining the relative importance of each feature in case of the XGBoost model applied.




$$\text{Mean SHAP}(j) = \frac{1}{N}\sum_{i=1}^{N} |SHAP_{i,j}|$$

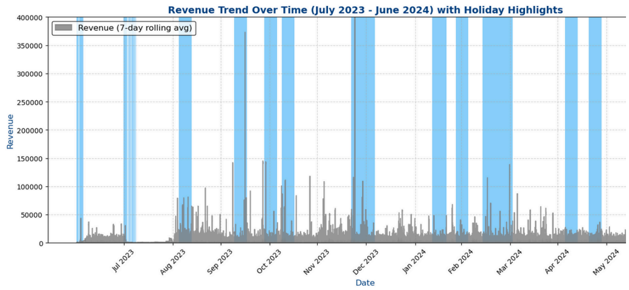
# MODEL COMPARISIONS

We have done a detailed analysis of all the models applied on various parameters such as:

- RMSE (Root Mean Squared Error)
- SMAPE (Symmetric Mean Absolute Percentage Error)
- Median Absolute Error
- Mean Absolute Error (MAE)

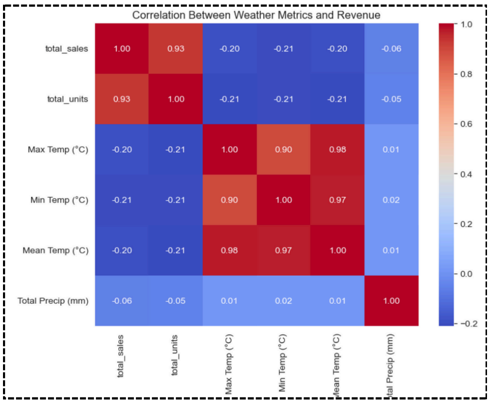
| S. No | Name of the Model Applied | RMSE | SMAPE | Mean absolute error | Median Absolute Error |
|---|---|---|---|---|---|
| 1. | XGBoost | 2839.78 | 2.82 | 81.66 | 3.08 |
| 2. | CatBoost | 2790.57 | 8.61 | 141.93 | 6.29 |
| 3. | Deep Neural Network (DNN) | 7294.64 | 125.64 | 2892.93 | 44.97 |
| 4. | AdaBoost | 4070.17 | 115.54 | 2956.09 | 55.06 |
| 5. | LightGBM | 4030.70 | 3.42 | 97.27 | 3.60 |
| 6. | Stacked Model: Base:[XGBoost+ Catboost], Meta-model:[Linear Regeressor] | 2915.36 | 8.76 | 127.44 | 6.06 |

# EDA OF KRA 5: EXTERNAL FACTORS ANALYSIS



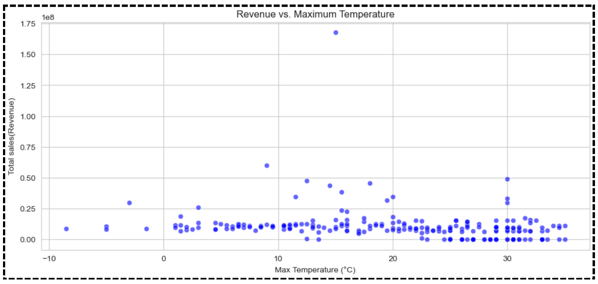

### Analysis of Demand Trends with Holidays:

GMV trends from July 2023 to June 2024 shows that holidays have little impact on sales, GMV increase is observed during holiday periods highlighted in blue-area.



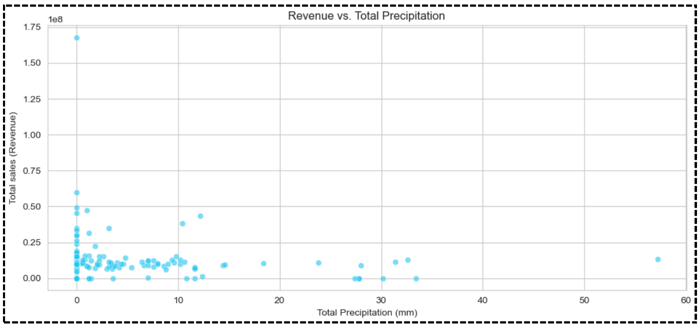

### Impact of Weather on Sales: Correlation Analysis:

The correlation heatmap analysis shows that temperature variables have a weak negative correlation with Sales and total units sold, suggesting minimal impact.




### Analysis of Temperature Impact on Sales

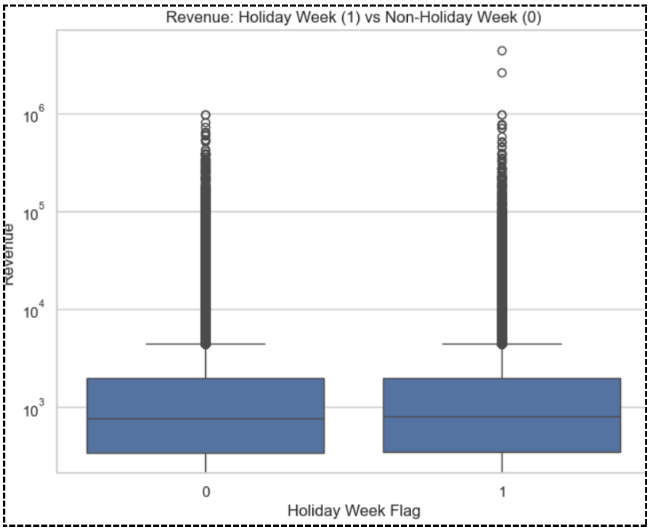
- GMV vs max temperature shows no strong correlation between the two variables.
- Revenue values are scattered across different temperature ranges, indicating that extreme heat or cold does not significantly impact sales.



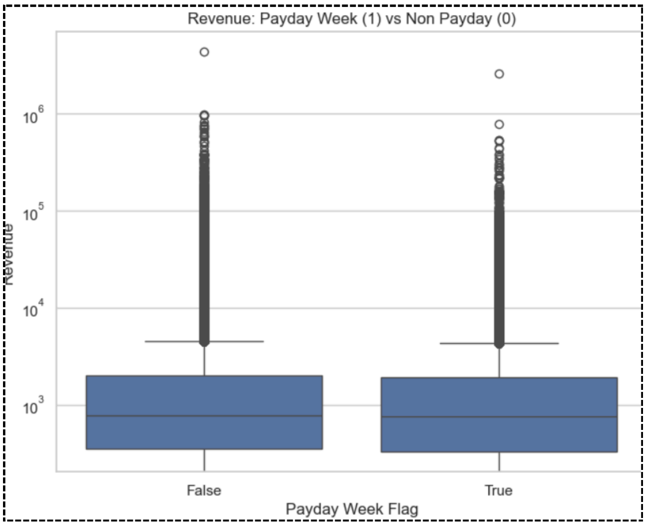

### Analysis of Precipitation Impact on Sales

- The analysis shows a negative correlation between precipitation and GMV, with higher rainfall linked to lower revenue.
- Wet conditions may reduce consumer spending, making drier days more favorable.

# EDA OF KRA 5: EXTERNAL FACTORS ANALYSIS



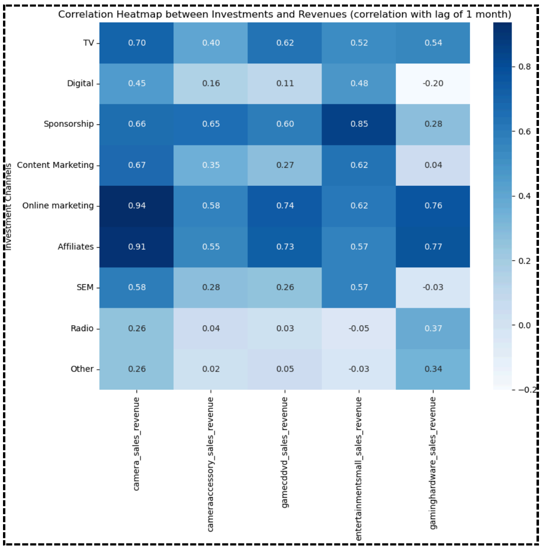

**Holiday vs. Non-Holiday Weeks:**
**T-statistic:** 1.64; **P-value:** 0.101 (above 0.05 threshold); **Inference:** No statistically significant difference in revenue, so we fail to reject the null hypothesis.

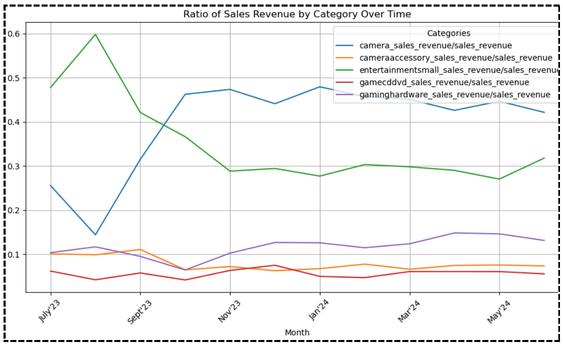


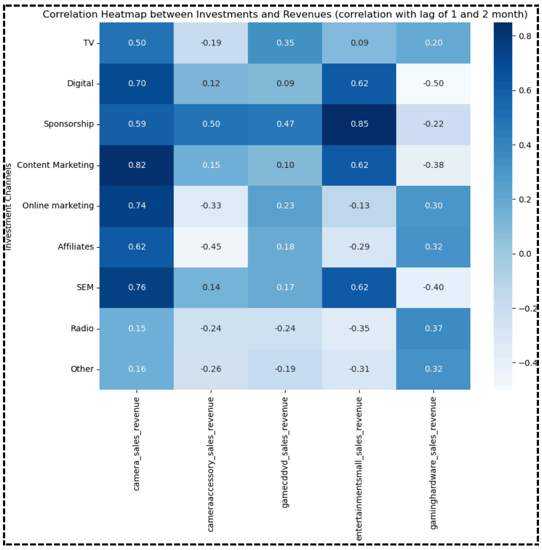

**Payday vs. Non-Payday Weeks:**
**T-statistic:** 18.43; **P-value:** ~8.36e-76 (far below 0.05); **Inference:** Revenue during Payday Weeks is significantly lower than during Non-Payday Weeks. The low p-value means this difference is very unlikely to be due to chance, so the effect is real.

# REFERENCES

## Budget optimization




# LIMITATIONS OF BUDGET OPTIMIZATION

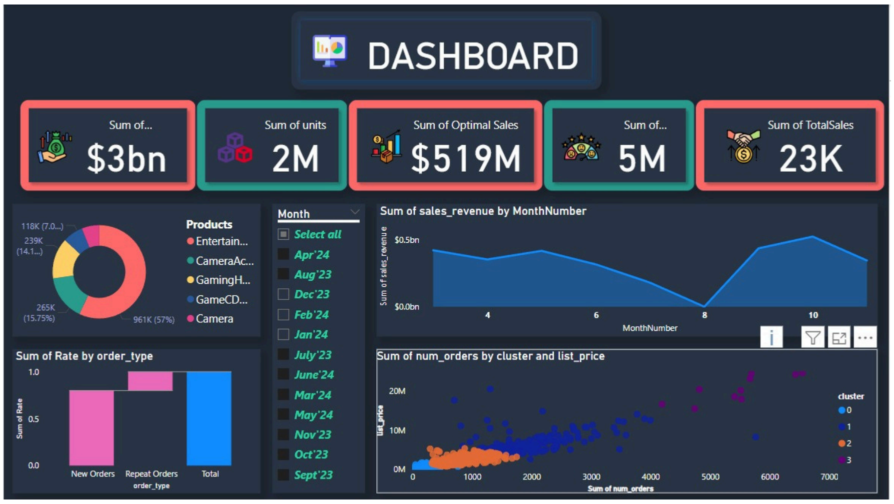
**Potential Limitations**:

1. **Correlation vs. Causation:** While correlation matrices provide useful insights, they don’t necessarily establish causality. Some correlations might be spurious.
2. **Limited Data (12 Months):** With only 12 months of data, noise in correlations might affect the reliability of recommendations.
3. **Fixed Lag Factor (Gamma):** The assumption of a fixed decay parameter (γ) for the lag effect may not hold uniformly across all channels
4. **Ignoring Interactions:** The model currently treats each investment channel independently. There could be synergy effects (e.g., digital ads boosting TV ad effectiveness).

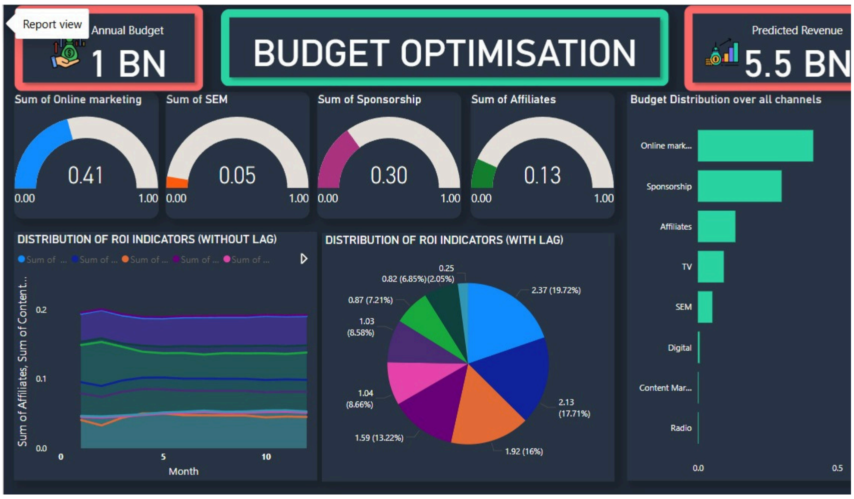
# DASHBOARD

## POWER BI DASHBOARD

The dashboard summarizes $3B sales, 2M units, 23K orders, and $519M optimal sales, with trend analysis



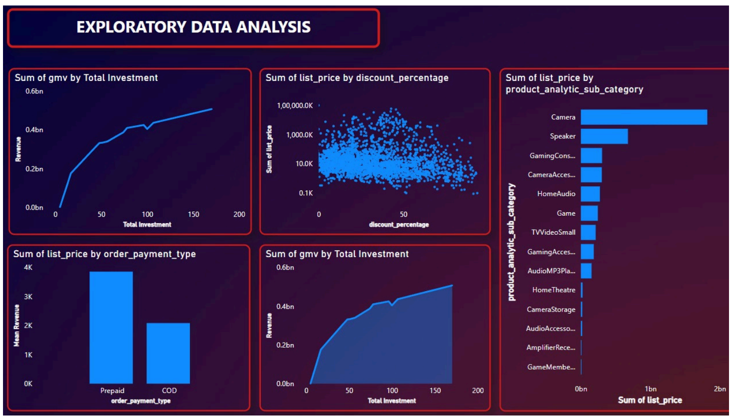

## BUDGET OPTIMISATION DASHBOARD




Annual budget: $1B, predicted revenue: $5.5B, major allocations: online marketing, sponsorship, affiliates, ROI analysis.

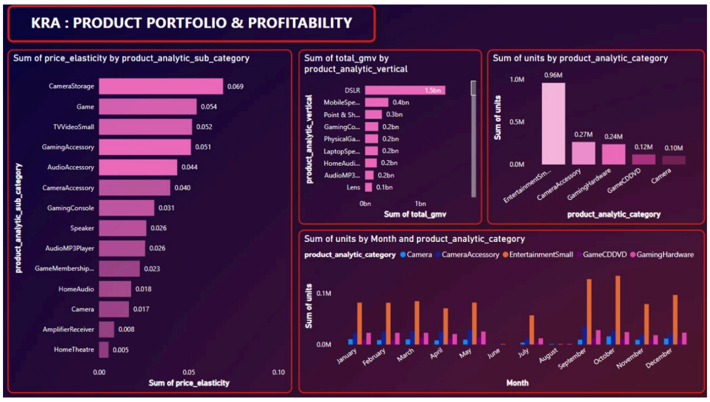
## EDA DASHBOARD

Investment drives revenue growth; cameras lead sales; prepaid orders outperform COD; discounts vary widely.



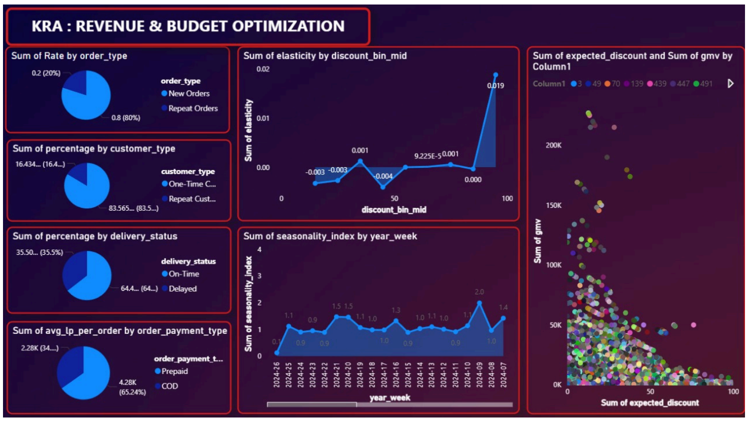
# DASHBOARD

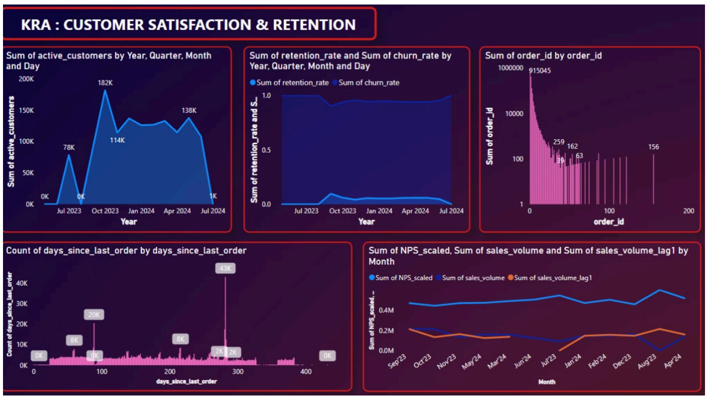
## KRA DASHBOARD



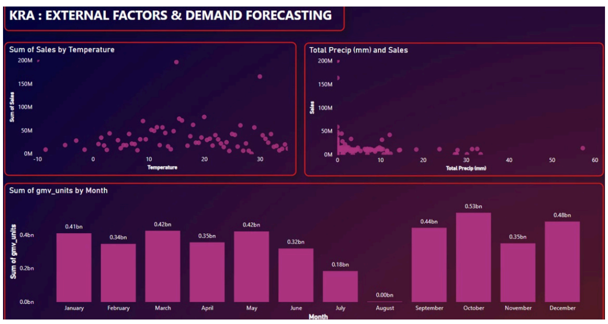

DSLR leads GMV; CameraStorage most elastic; EntertainmentSmall dominates units; seasonal variations in category sales observed.

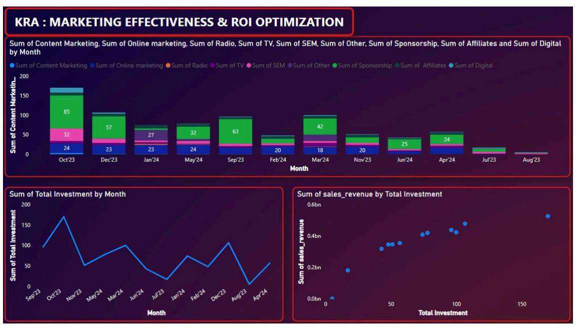
Repeat orders dominate; prepaid orders have higher avg. value; delayed deliveries are common; seasonality and discount elasticity impact GMV.




Marketing effectiveness, customer satisfaction, and external factors drive performance. Investment, retention, and demand forecasting influence revenue, sales, and growth through data-driven insights and strategic optimization.



## Budget optimization example

| | TV | Digital | Sponsorship | Content Marketing | Online Marketing | Affiliates | SEM | Radio | Other | Sum of proportions |
|---|---|---|---|---|---|---|---|---|---|---|
| **SLSQP** | 0.092977 | 0.006146 | 0.300435 | 0.000001 | 0.413976 | 0.134542 | 0.051923 | 1.508900e-17 | 0.000000 | 1.0 |
| **L-BFGS-B** | 0.090681 | 0.106350 | 0.158647 | 0.084319 | 0.125004 | 0.094530 | 0.135733 | 8.569213e-02 | 0.119043 | 1.0 |
| **TNC** | 0.090681 | 0.106350 | 0.158647 | 0.084319 | 0.125004 | 0.094530 | 0.135733 | 8.569213e-02 | 0.119043 | 1.0 |
| **linprog** | 0.001271 | 0.252369 | 0.056607 | 0.000001 | 0.025521 | 0.014550 | 0.241177 | 3.639901e-02 | 0.365114 | 1.0 |