TEAM 2025111

PROBLEM OVERVIEW

Problem Statement

Performance Driver Analysis:
Identify key KPIs that influence revenue growth and analyze historical data to determine primary performance drivers.

Impact Analysis on Marketing ROI:
Quantify the effect of different marketing levers on revenue and assess the return on investment on each marketing channel

Optimising Marketing Budget Allocation:
Recommend an optimal budget split based on data-driven insights and maximize revenue impact by reallocating resources effectively.

Key Considerations

Data Complexity
Distributed datasets containing order details, marketing data, and special event information

23 features

Data Quality Issues
Missing values in numerical and categorical features affect prediction accuracy.

4,096 missing values in **numerical features**

Data Imbalanced
Large dataset (1600,000+ rows) covering order specifics and monetary data

Heavily Skewed

Prominent Features

GMV
The number of days between lead creation and an installation request.

Sub_Category
Product Sub-Category Information

Media Spend
Higher media spend is generally associated with better lead nurturing.

SLA
Number of days it typically takes to deliver the product

Bias in Prediction Models
If not handled, imbalance can cause biased recommendations

Feature Redundancy
The model might overfit to past data patterns, failing to generalize to new customer behaviors

Perceived Importance
If missing values are not properly handled, they can affect the perceived importance of features.

Multiple File Handling
Managing different files may involve addressing inconsistencies in format.

PRE PREPROCESSING

DATA PREPROCESSING

Handling Missing Values

GMV column:
Dropped the missing values

Weather Column:
Imputed using a rolling average

Outlier Detection

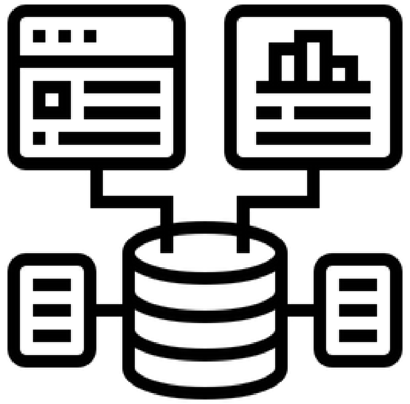
Box plots and Scatter plots
Detect numerous outliers in the dataset

Mathematical Transformation

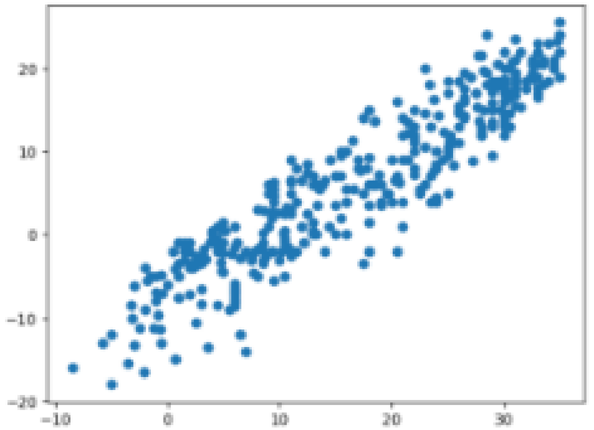
Transformation
Applied log transformation to make the data more Gaussian

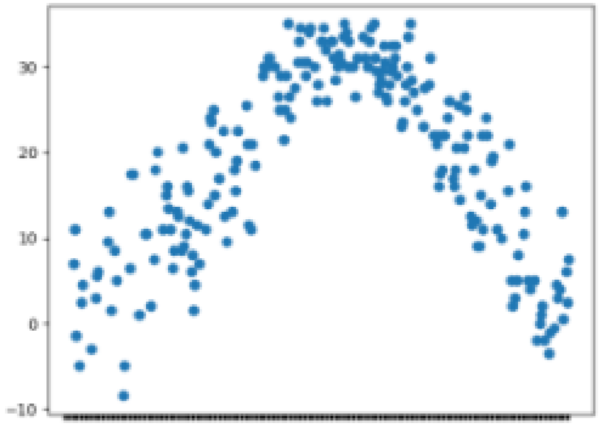
Merging Datasets

Combining SKU, FSN_ID, and order data links products to sales, enabling trend analysis and decisions.



Missing	Strategy
Numerical Column (eg : GMV)	**Dropped(very less missing values)**
Weather Data	**Rolling average imputation**

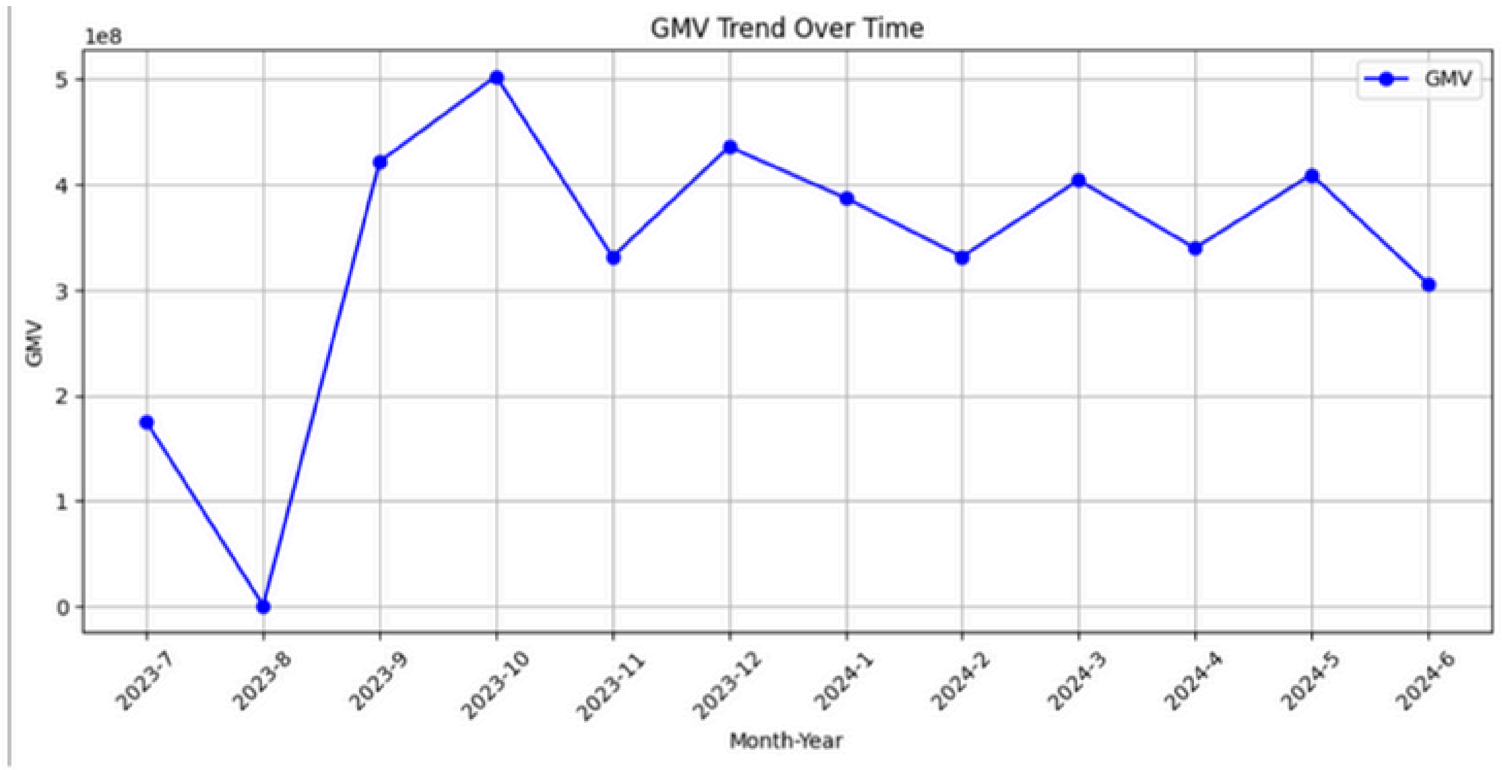




Joining Datasets by FSN_ID:
Two datasets, SKU_details.csv (product info) and Customers_Orders_Data.csv (transaction details), were combined using the common key fsn_id. This created a customized dataset (Customers_Orders_Data.csv) that links product performance with customer sales, helping track buying patterns, evaluate delivery efficiency, and make informed decisions in areas like inventory management and marketing strategies.

Weather Data Handling:
The weather datasets for Ontario (2023 & 2024) contained significant missing values. A rolling moving average technique was applied to fill these gaps. This method replaced missing values with the average of surrounding values, ensuring data consistency and minimizing abrupt fluctuations.

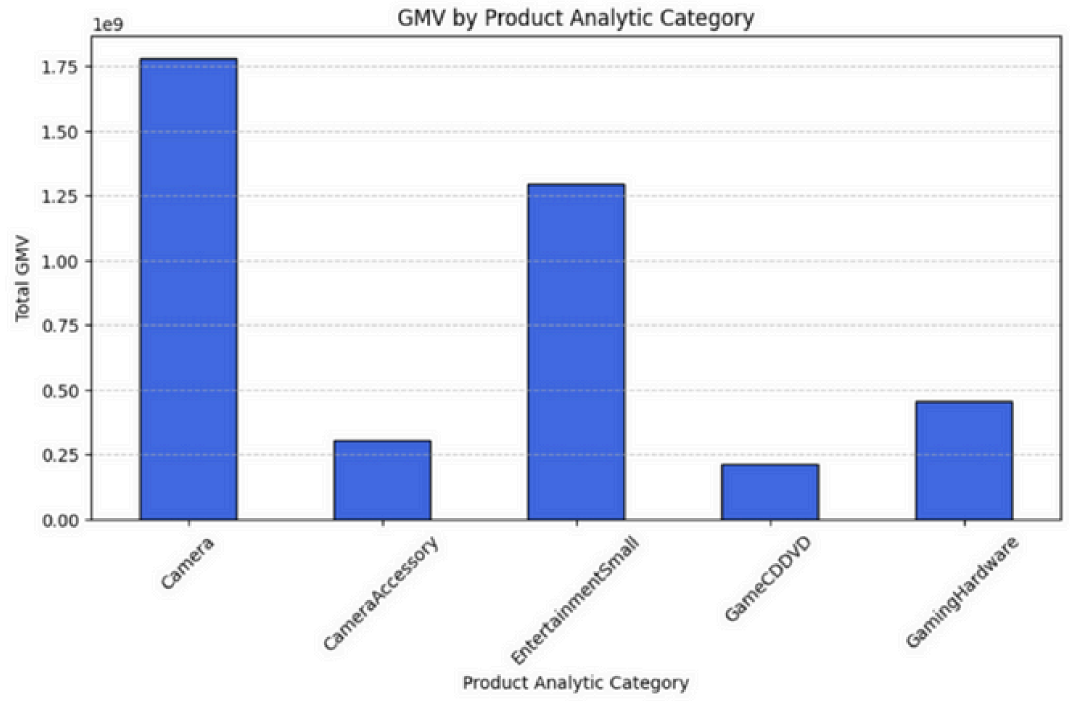
STRATEGIC INSIGHTS

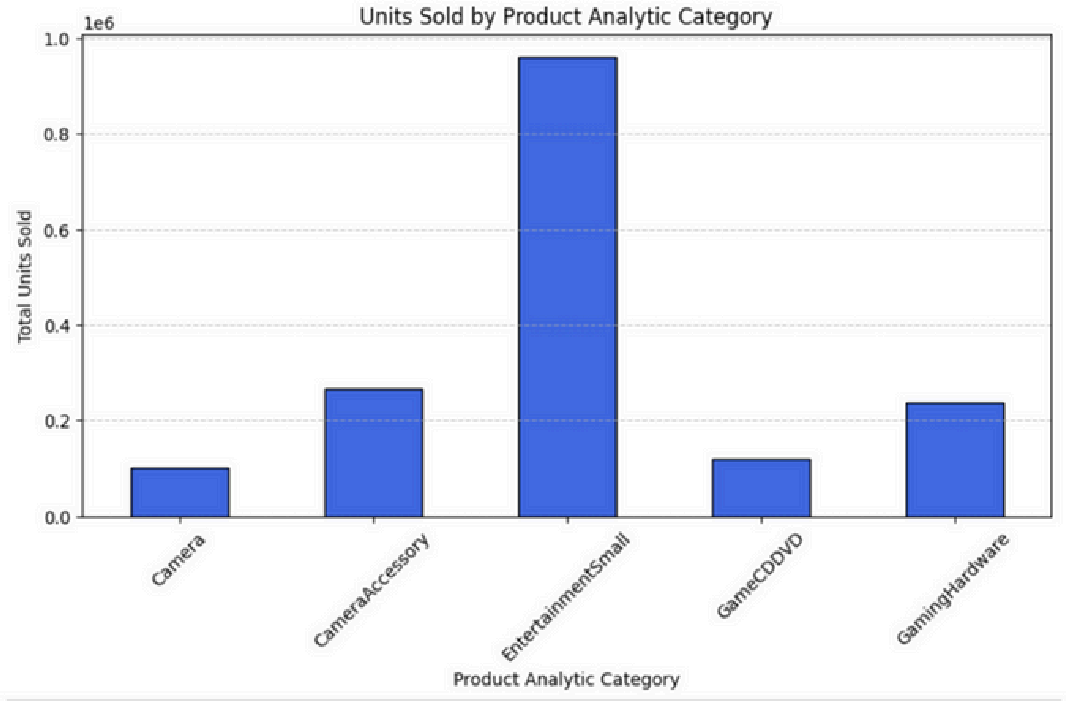


ANALYSIS OF GMV OVER TIME

- GMV fluctuated from July 2023 to June 2024.
- August 2023: Sharp drop, followed by a peak in October 2023, likely seasonal demand.
- December 2023 & April 2024: Notable peaks, likely sales events.
- January & June 2024: Dips, possibly due to seasonal slowdowns.
- Despite early setbacks, GMV recovered and stabilized with periodic fluctuations. Further analysis to follow.

ANALYZING THE PRODUCT_ANALYTICAL_CATEGORY



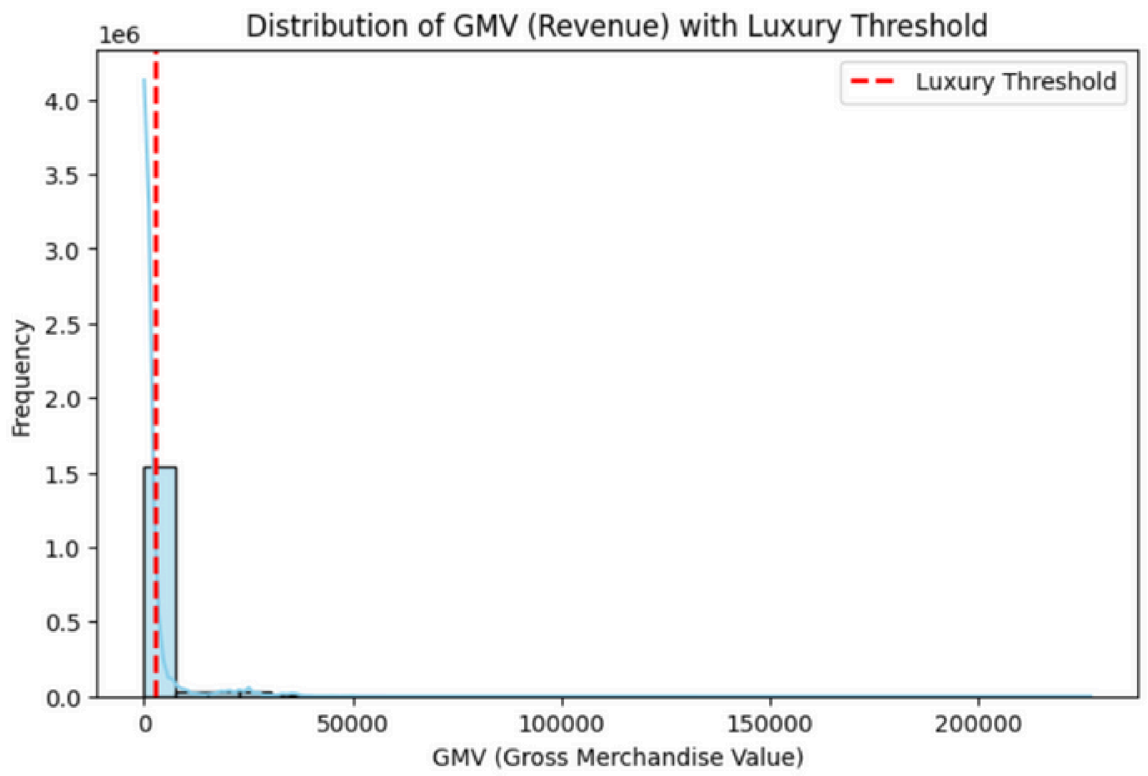


- Camera: High GMV, low unit sales → High-value products.
- Entertainment Small: Highest unit sales, lower GMV share → Lower per-unit value.
- Opportunity: Increase Camera sales for higher GMV.
- Strategy: Sustain Entertainment Small’s volume for steady revenue.

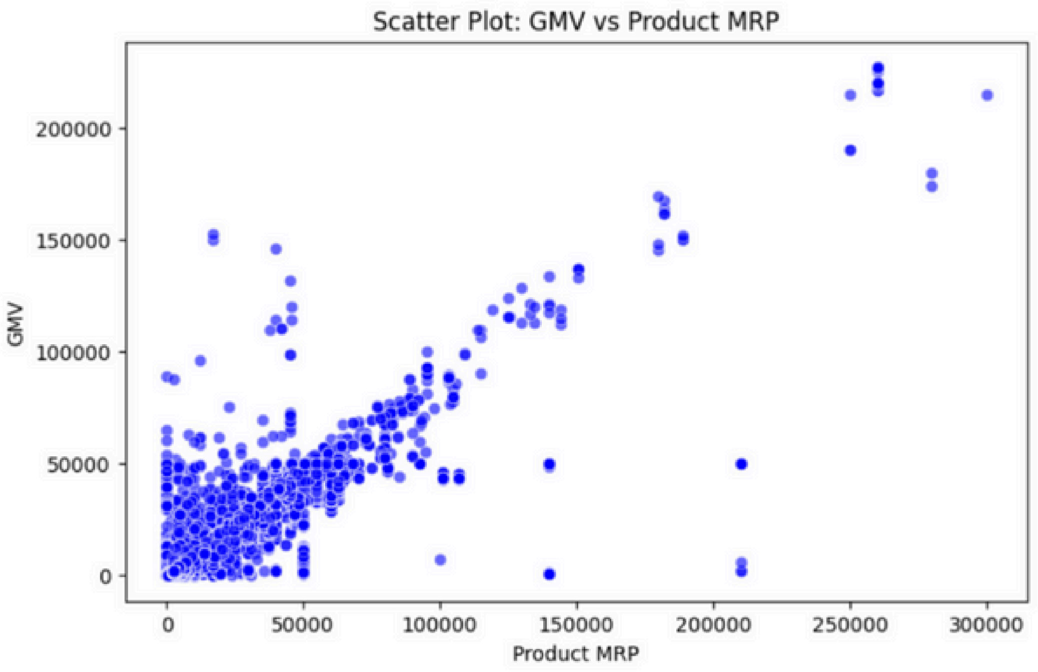
STRATEGIC INSIGHTS

ANALYZING THE DISTRIBUTION OF GMV

- GMV distribution is highly skewed, with most transactions at low values.
- Luxury threshold (80th percentile) shows few high-value products.
- A long tail indicates some extremely high-value transactions.
- The majority of products are mass-market; strategy: increase luxury sales or boost mass-market volume to maximize GMV.

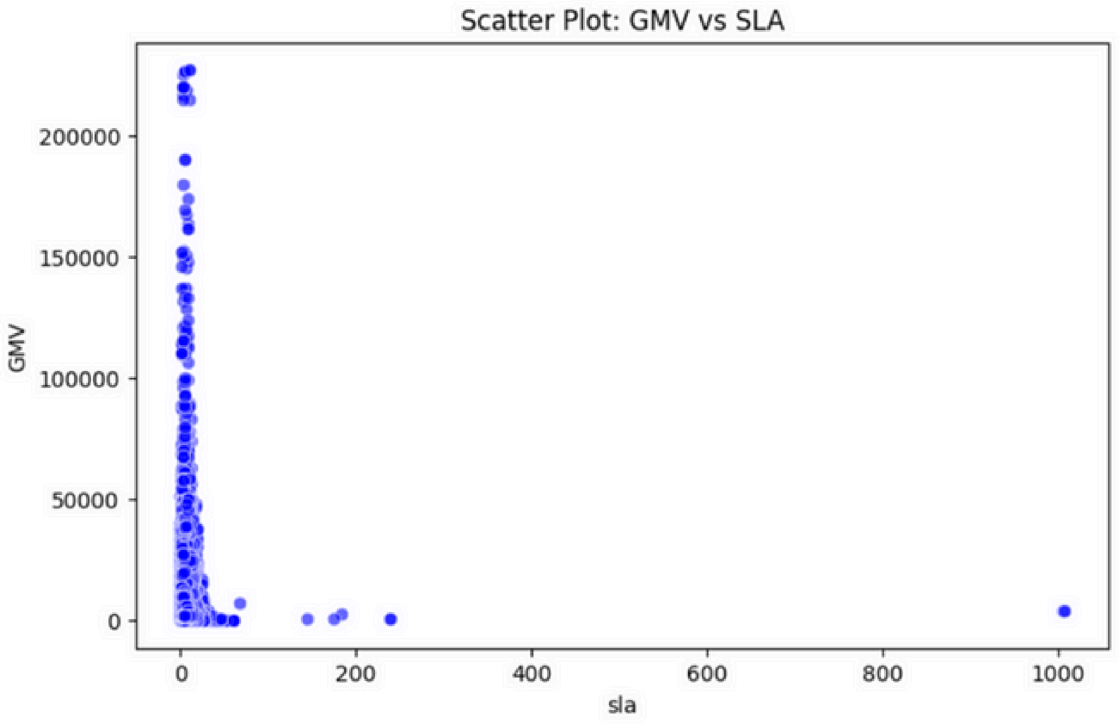


ANALYZING THE PRODUCT_ANALYTICAL_CATEGORY



- The plot suggests a positive correlation between GMV (Gross Merchandise Value) and Product MRP.
- Higher MRP products tend to have higher GMV

Most high GMV transactions occur at very low SLA values, indicating that faster deliveries are associated with higher revenue generation.



STRATEGIC INSIGHTS

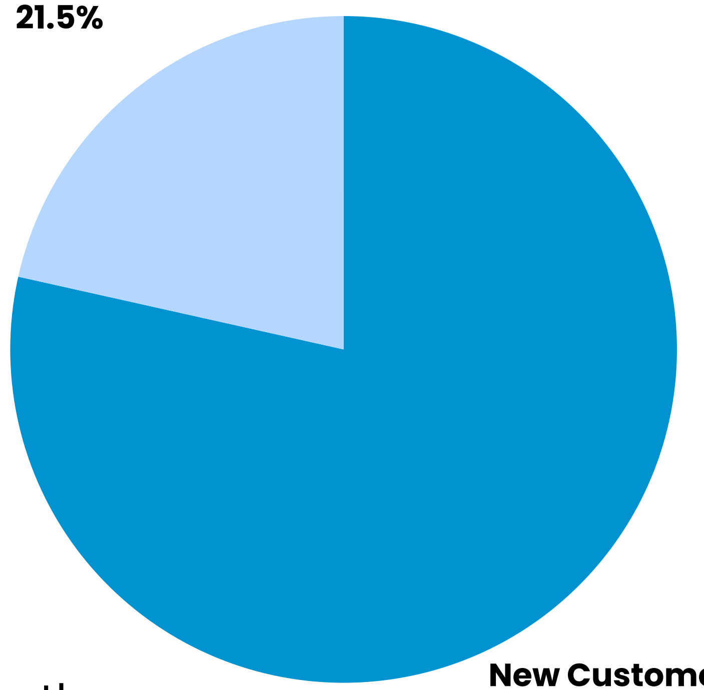
ANALYZING CUSTOMER DEMOGRAPHICS

Customer Overview
- Total Customers: 1.25M
- New Customers (1 order): 984K (78.5%)
- Repeat Customers (>1 order): 269K (21.5%)

Revenue Insights
- Avg. GMV per New Customer: ₹2,438.82
- Avg. GMV per Repeat Customer: ₹2,494.41

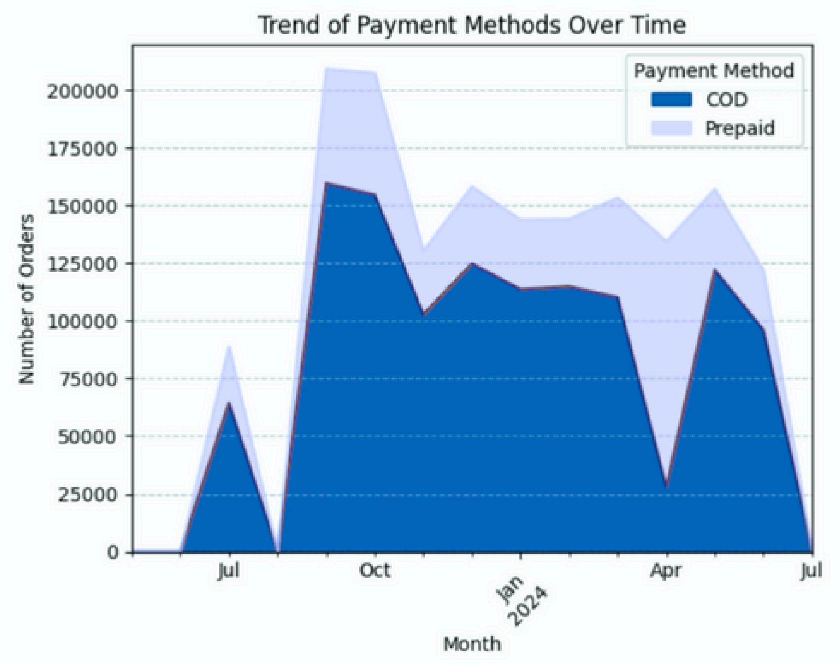
Sales Distribution
- High-Sales Regions (Top 20%) – ₹1.28M
- Medium-Sales Regions – ₹354K
- Low-Sales Regions (Bottom 20%) – ₹9.65K

Key Takeaways & Action Plan
- Prioritize High-Sales Regions to maximize revenue growth
- Increase repeat purchases by targeting existing customers
- Expand presence in medium-sales regions to unlock potential growth



Comparison of payment methods

In general, people prefer COD over digital payments. But particularly in April. Due to tax filing, people are cautious about spending cash so they prefer digital payments like credit cards which they can pay off later.



ANALYSIS OF GMV on Special Days

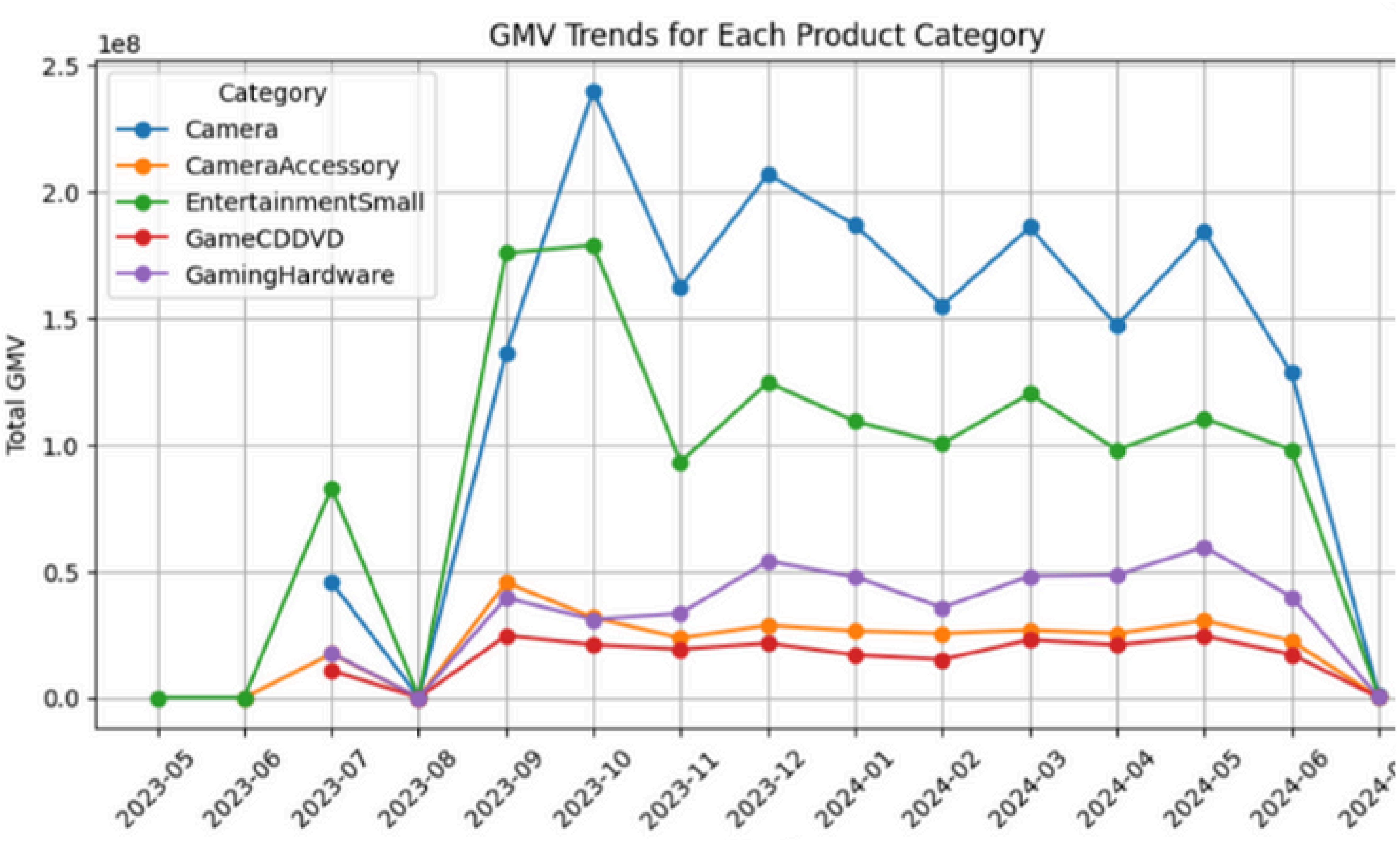
DISCOUNT DAYS

- Discounts increase GMV (₹3,184.06 vs ₹2,348.70) but don't affect order volume (1.02 units/order).
- SLA remains at ~5 days.
- COD to Prepaid ratio stays at 73:27.
- Key Insight: Discounts drive higher-value purchases, but strategies are needed to boost volume.

HOLIDAYS

- Order volume remains stable (~1 unit/order) on holidays and non-holidays.
- GMV is slightly higher on non-holidays (₹2,466 vs. ₹2,357), likely due to offline shopping shifts.
- SLA stays consistent at 5-6 days, ensuring smooth operations.
- Key Insight: Targeted promotions could boost holiday sales and offset offline competition.

Analysis of GMV trends product wise

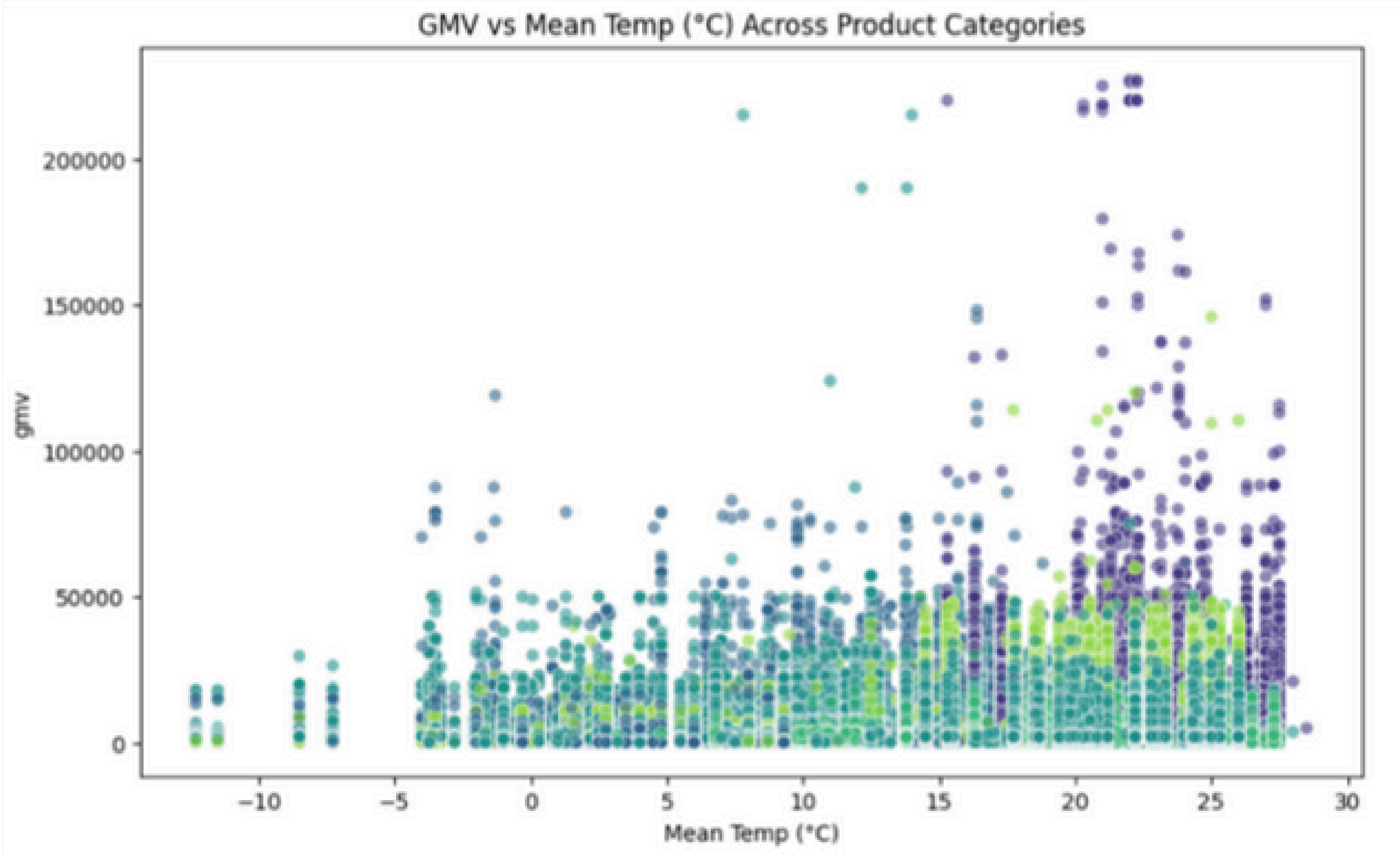


- There is a dip in sales in all categories in the period of June to August.
- This is explained by seasonal variation as June to August is summer in Canada. During summers people go on vacations and prefer outdoor activities, so they spend less on items like gaming setups and luxury camera accessories.
- The marketing budget is also very low compared to the latter months. This also explains why sales are so low.
- Another reason for low sales of Gaming Hardware is that most gaming companies launch their new products in Q4. So people wait for new arrivals during the prior two months.

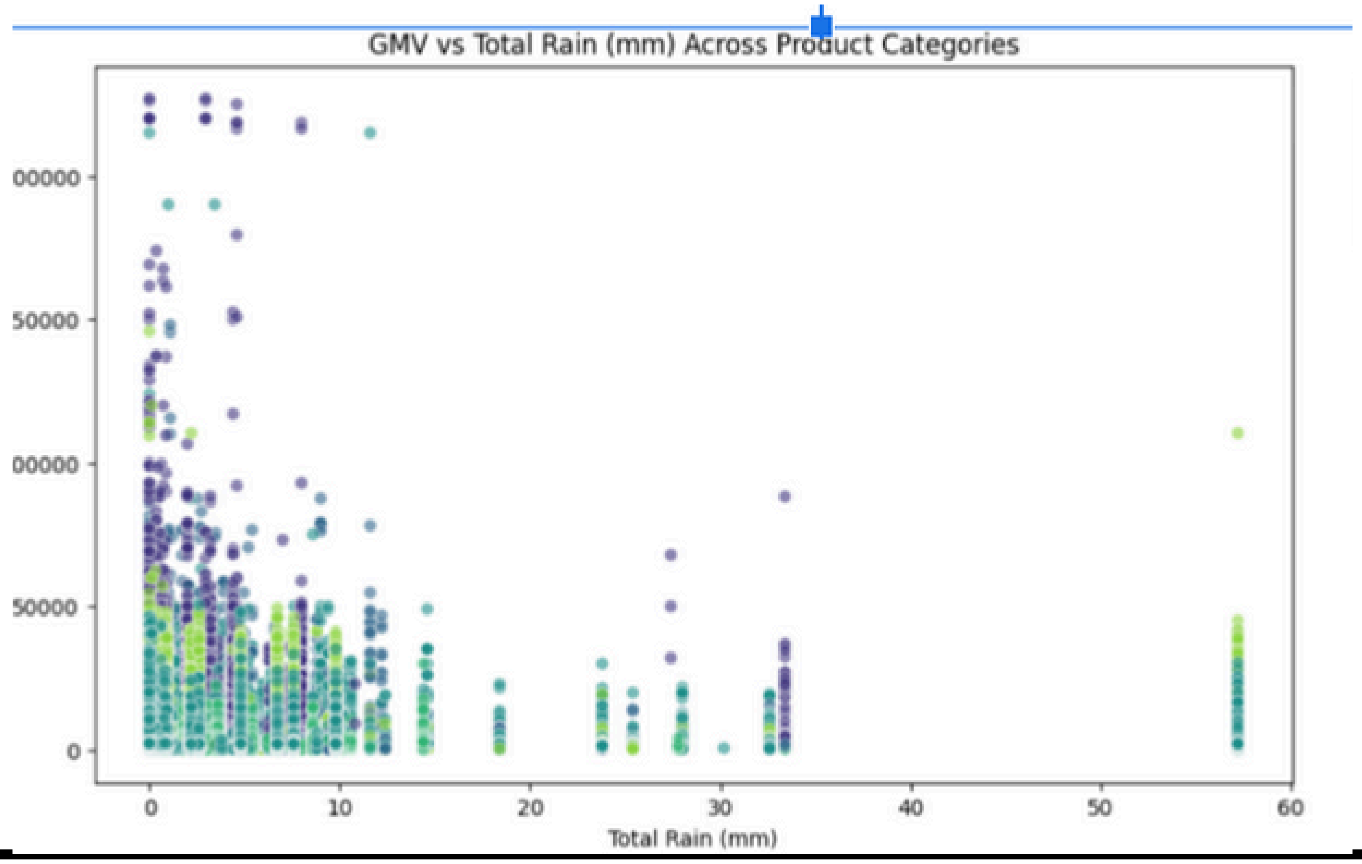
ANALYSIS OF GMV VS WEATHER

Temperature & Product Sales

- Cameras stand out from other categories, as confirmed by statistical tests (ANOVA and TSHD). They are mostly bought when the temperature is higher and rainfall is lower.
- Luxury and non-luxury products show similar temperature trends, meaning weather affects them in roughly the same way.
- Game Membership Cards were purchased significantly more on rainy days in the non-luxury segment, likely because people tend to stay indoors and look for entertainment.



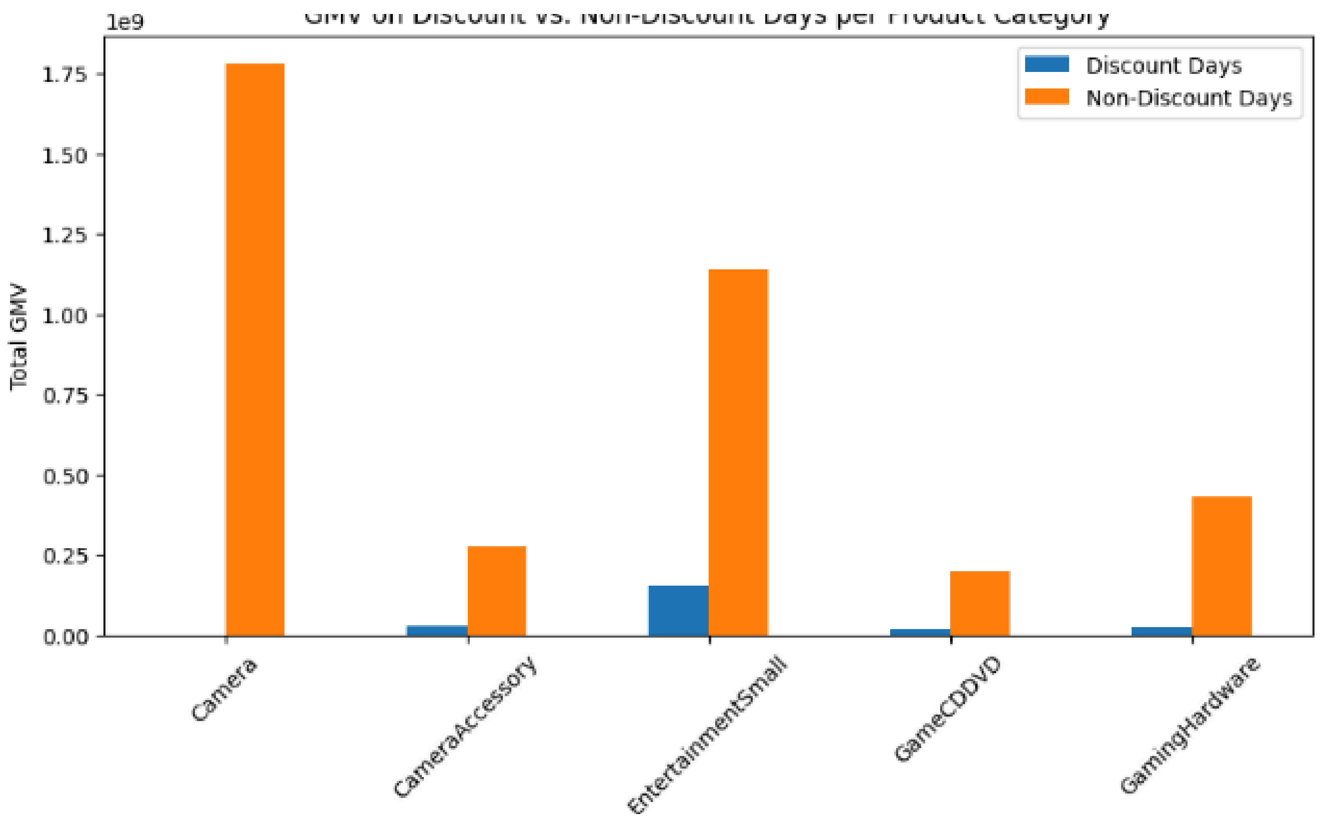
Rainfall & Sales Patterns



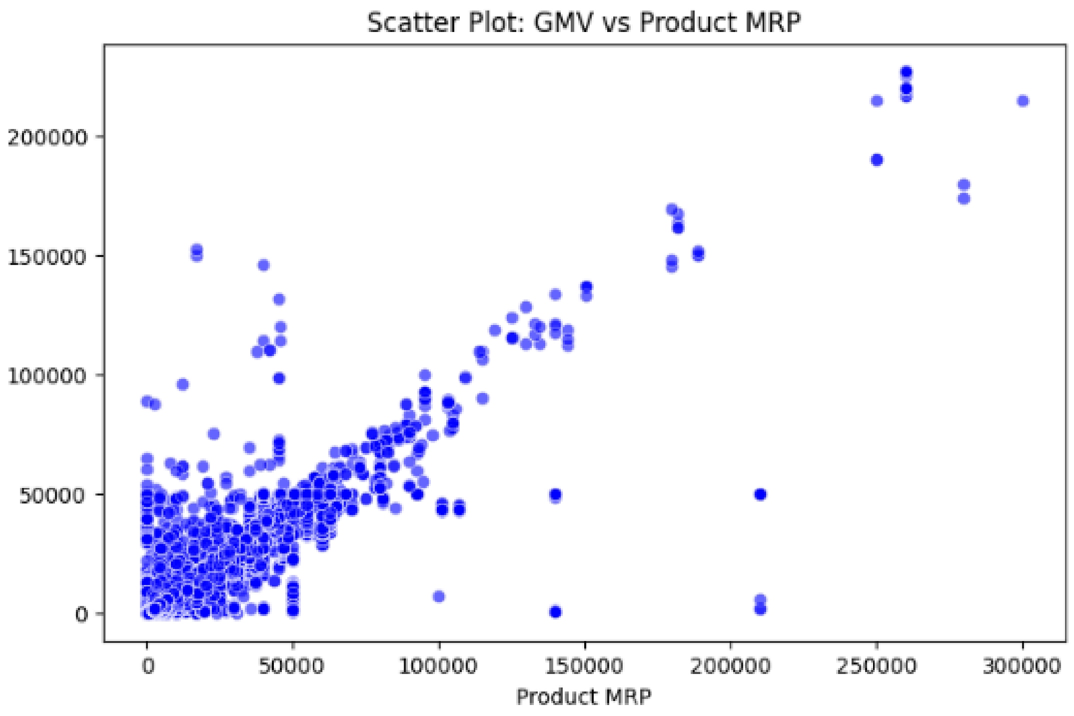
- Cameras sell best when it's dry, with demand dropping when rainfall increases.
- Most other categories follow a similar pattern, but gaming-related products, especially Game Membership Cards, sell more when it rains, showing they are a popular choice for indoor activities.

STRATEGIC INSIGHTS

ANALYSIS OF GMV VS DISCOUNT

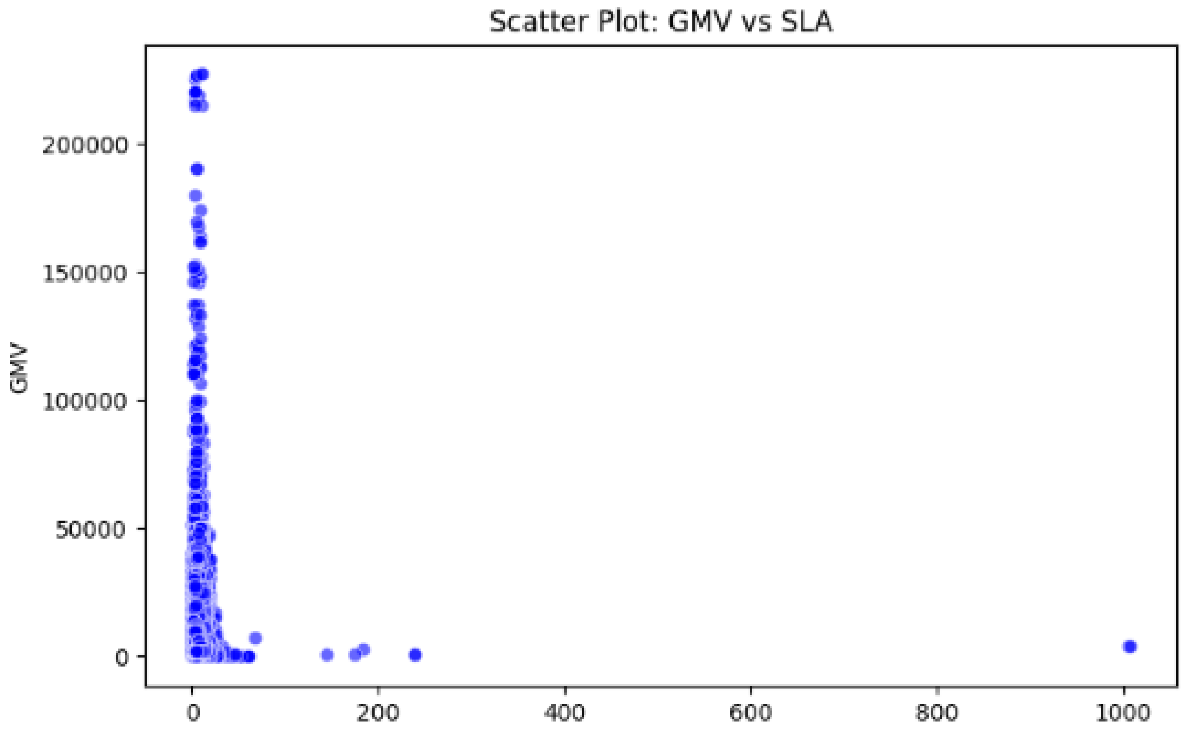


- The plot shows that **non-discount days generate significantly higher GMV** across all product categories, especially for **Cameras and Entertainment Small**, indicating strong demand even without discounts.
- While discounts slightly boost sales in some categories, most GMV still comes from non-discounted purchases, suggesting that discounts are not the primary sales driver for these high-value items.



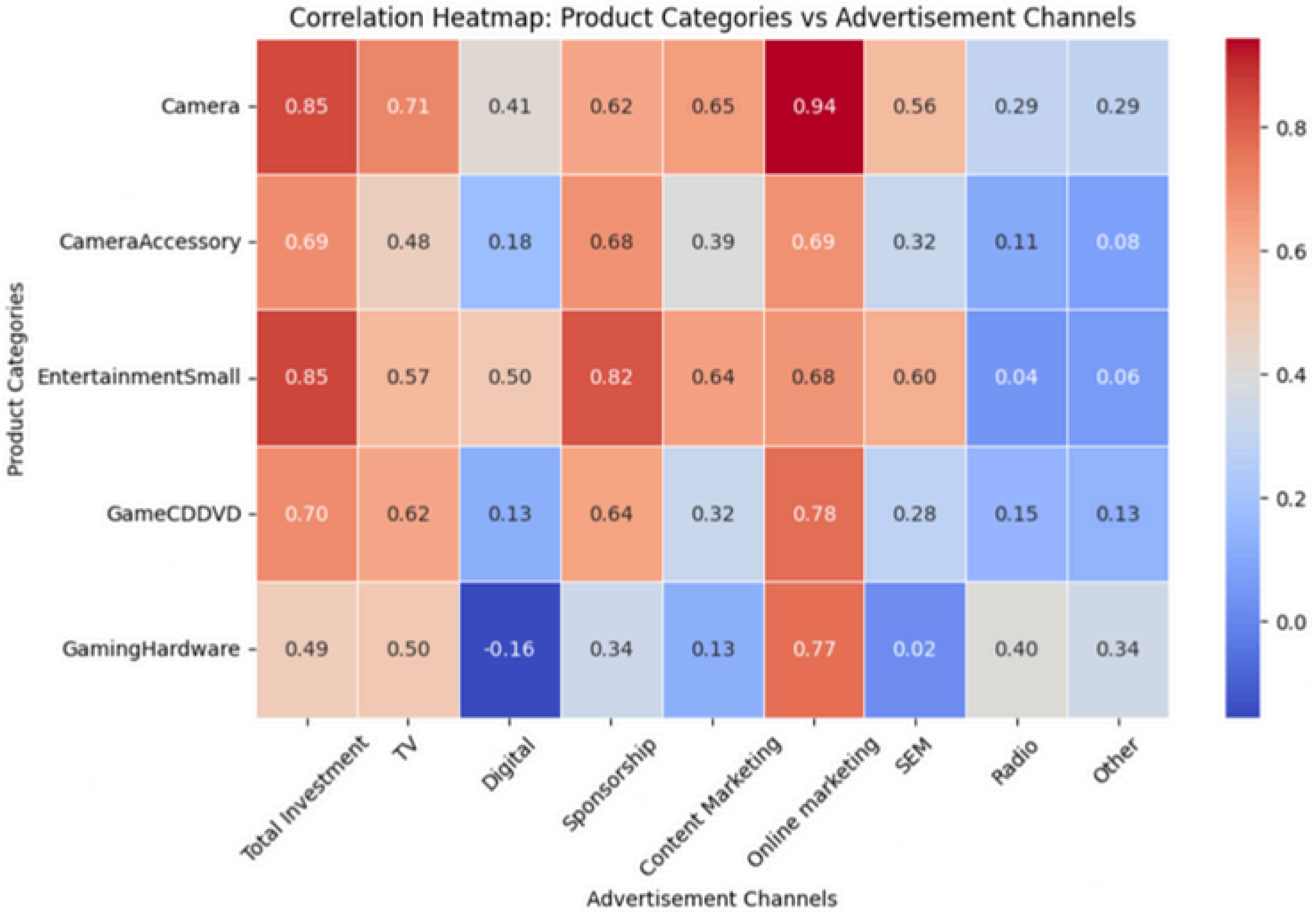
- The plot suggests a positive correlation between GMV and Product MRP.
- Higher MRP products tend to have higher GMV.

- Most high GMV transactions occur at very low SLA values, indicating that faster deliveries are associated with higher revenue generation.
- As SLA increases, GMV tends to be very low, suggesting that long delivery times may negatively impact high-value sales.



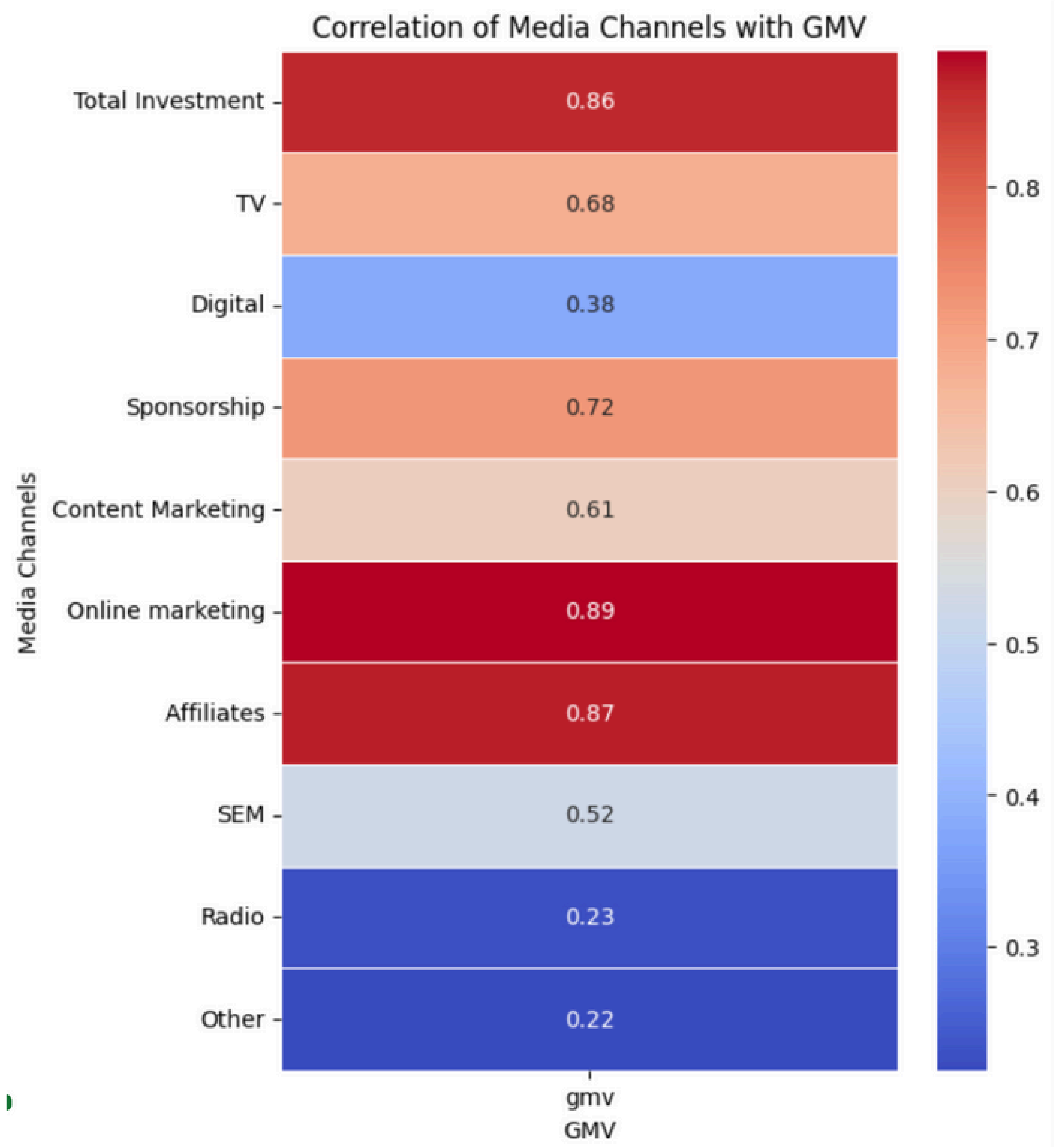
STRATEGIC INSIGHTS

ANALYZING MEDIA CHANNEL vs PRODUCT_ANALYTICAL_CATEGORY



- **Best Performing Channel:** Online Marketing excels across all categories, especially Camera (0.94), EntertainmentSmall (0.68), and GamingHardware (0.77).
- **Category Highlights**
  a. Camera: Strong with Online Marketing (0.94) and TV (0.71).
  b. EntertainmentSmall: Driven by Sponsorship (0.82) and Total Investment (0.85).
  c. GamingHardware: Online Marketing (0.77) is key.
  d. Weak Channels: Digital Advertising, Radio, and SEM show poor or negative correlations across most categories.
- Camera seems to have the best correlation statistics with all the marketing channels, indicating it is the most compatible product category and the easiest one to sell.

STRATEGIC INSIGHTS



GMV vs Total Investment

Overall, investment in marketing does lead to higher GMV

GMV vs TV

Fair bit of correlation denoting TV marketing is still useful for some audiences.

GMV vs Digital

Need to enforce novel marketing strategies to capture the digitla market

GMV vs Sponsorship

Fairly important, especially for catching eyes at events of scale

GMV vs Content Marketing

Relevant in specific scenarios

GMV vs Online Marketing

Strongest affect, can look to boost investment here

GMV vs Affiliates

Must try to form dense influencer partnership network

GMV vs SEM

Reassessment required to improve visibility

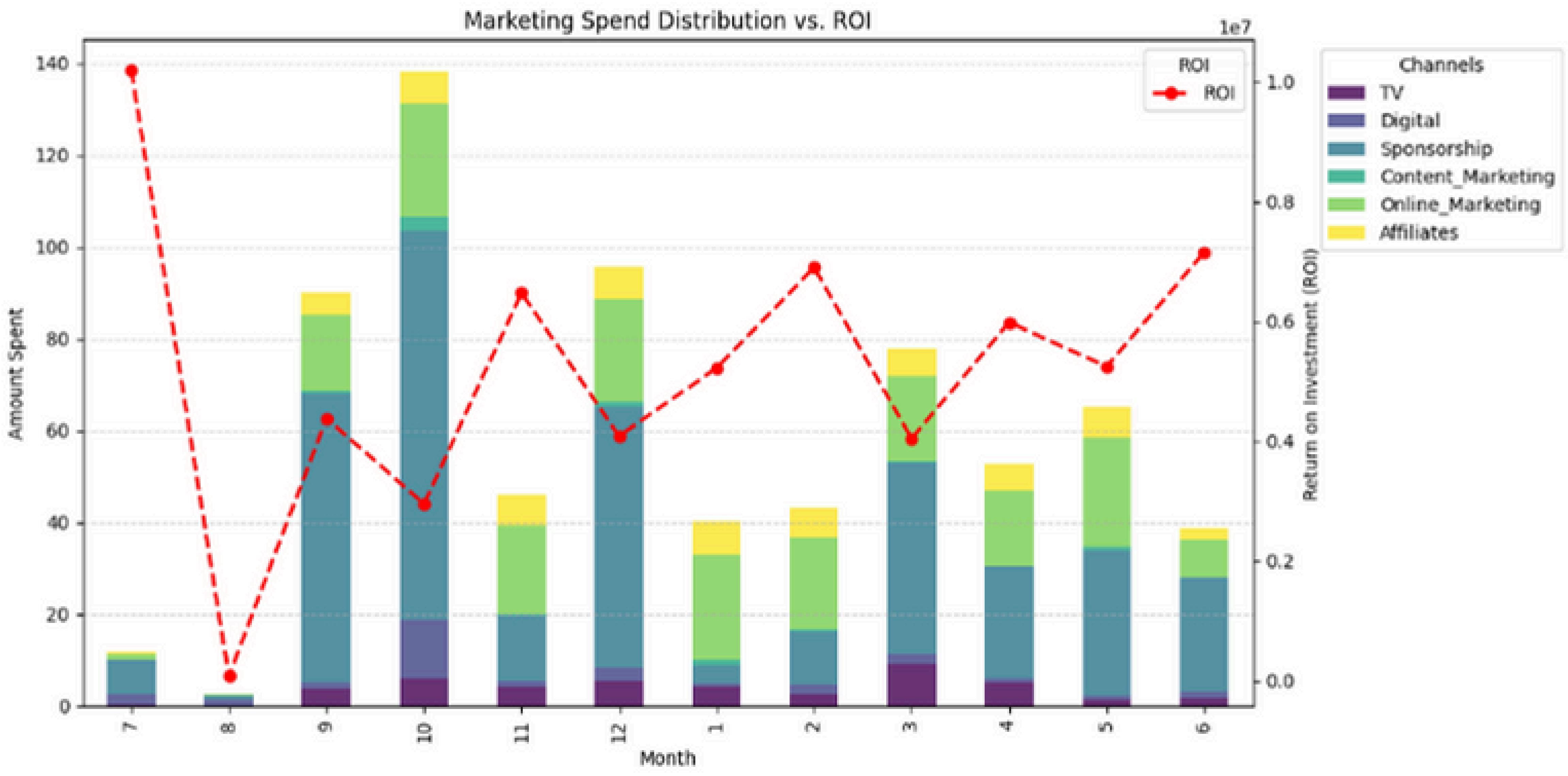
GMV vs Radio

Do away with the outdated channel

GMV vs Other

Decrease focus in irrelevant channels

Impact analysis on marketing ROI



1. **Fluctuating ROI**
   - The red dashed line (ROI) shows significant fluctuations over the months.
   - ROI appears to peak in some months despite lower spending, indicating non-linear returns on investment.

2. **High Investment, Low ROI Cases**
   - Some months with high marketing spend (e.g., Month 10) did not yield the highest ROI, suggesting diminishing returns or inefficient allocation.

3. **Low Investment, High ROI Cases**
   - Months with relatively lower spend (e.g., Month 8) still show spikes in ROI, implying that external factors or organic growth may have played a role.

4. **Dominance of Sponsorship & Digital**
   - Sponsorship and Digital channels are the largest contributors to overall spending, indicating a heavy reliance on these channels for marketing impact.

5. **Diversification Effect**
   - Months with a more balanced marketing mix (varied investments across multiple channels) tend to show better ROI stability, suggesting the importance of a diversified strategy.

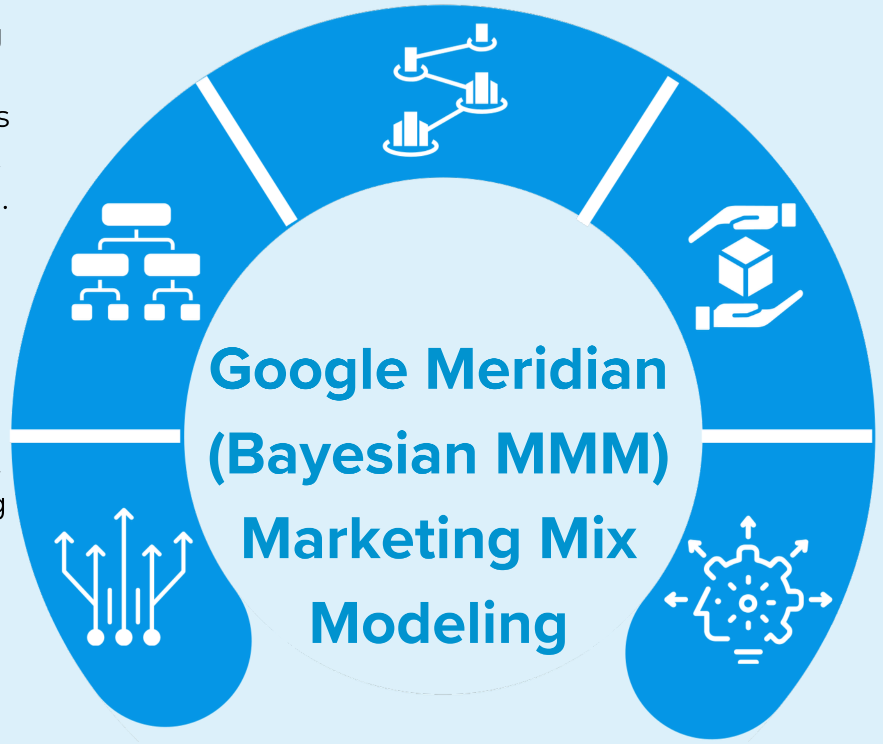
6. **Need for Optimization**
   - ROI variations indicate that simply increasing spending does not always guarantee higher returns. A data-driven approach is needed to identify the most effective channels for future campaigns.

THE MODEL

Bayesian Regression & Probabilistic Inference – Uses Bayesian priors and MCMC sampling for accurate and uncertainty-aware predictions.

Saturation & Diminishing Returns – Models how increasing ad spend leads to reduced effectiveness, preventing overspending.

Budget Optimization & Forecasting – Predicts future performance and suggests optimal budget allocation for maximum ROI.



Adstock Effect & Carryover – Captures the delayed impact of marketing efforts, ensuring past campaigns influence future conversions.

External Factor Integration – Incorporates seasonality, economic trends, and organic growth for realistic sales forecasting.

LightWeight_MMM:

Marketing Mix Modeling (MMM) is used by advertisers to measure advertising effectiveness and inform budget allocation decisions across media channels.

1. Data Collection & Preprocessing: Collect historical data, create features, handle missing values, and standardize inputs for better performance.

2. Model Selection & Bayesian Regression: Define priors, use Bayesian regression (NumPyro), and apply hierarchical modeling for dependencies.

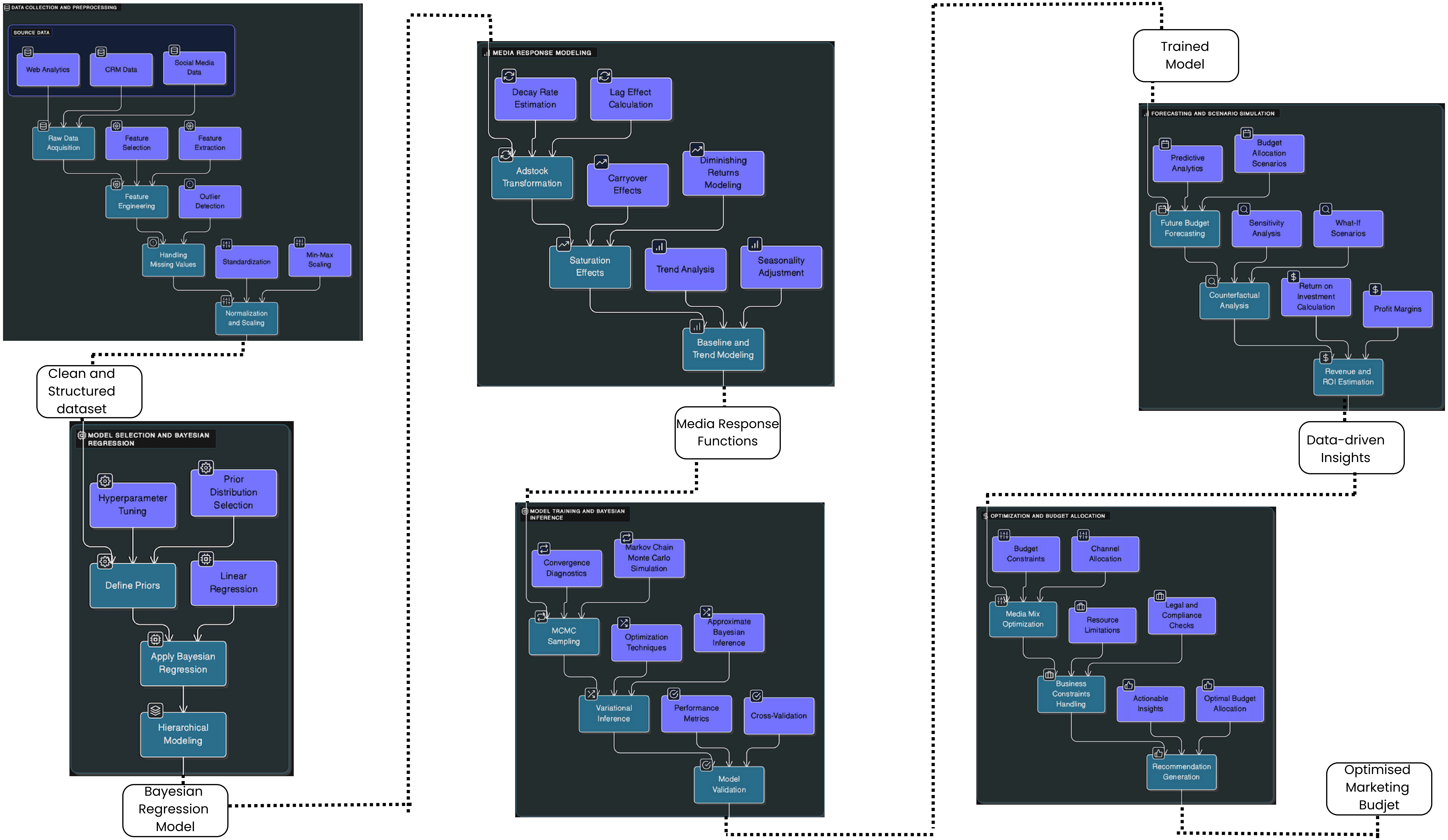
3. Media Response Modeling: Model adstock effects, simulate diminishing returns, and account for baseline and trend influences.

4. Model Training & Bayesian Inference: Use MCMC sampling, optional variational inference, and validate with cross-validation.

5. Forecasting & Scenario Simulation: Predict performance under various budget scenarios and optimize for ROI.

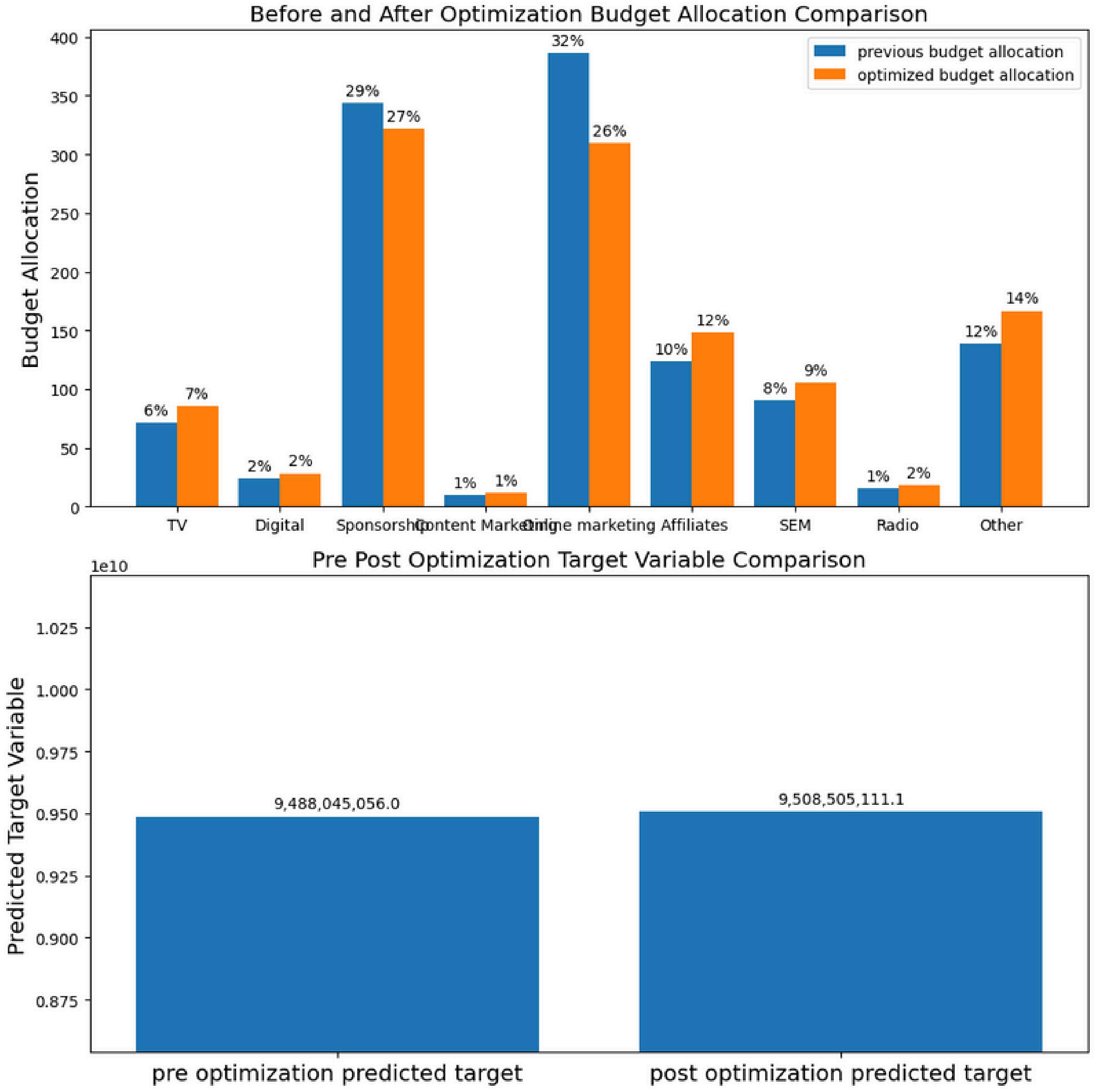
6. Optimization & Budget Allocation: Optimize media mix with constraints and provide actionable insights for media efficiency.

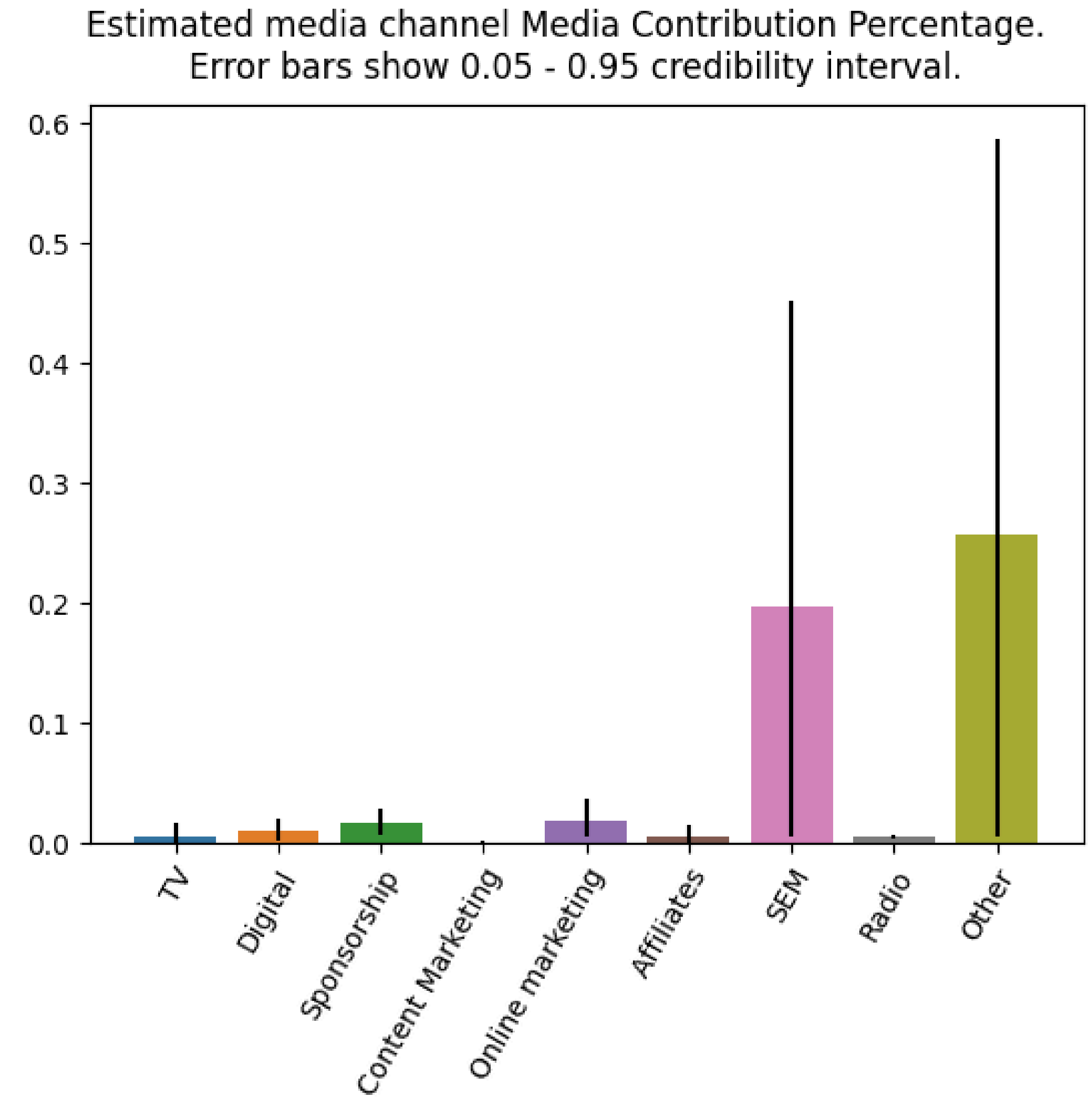
7. Deployment & Monitoring: Deploy models, track performance, and automate reporting on media effectiveness and ROI.



DATA COLLECTION AND PREPROCESSING
SOURCE DATA
Web Analytics
CRM Data
Social Media Data
Raw Data Acquisition
Feature Selection
Feature Extraction
Feature Engineering
Outlier Detection
Handling Missing Values
Standardization
Min-Max Scaling
Normalization and Scaling
Clean and Structured dataset
MODEL SELECTION AND BAYESIAN REGRESSION
Hyperparameter Tuning
Prior Distribution Selection
Define Priors
Linear Regression
Apply Bayesian Regression
Hierarchical Modeling
Bayesian Regression Model
MEDIA RESPONSE MODELING
Decay Rate Estimation
Lag Effect Calculation
Adstock Transformation
Carryover Effects
Diminishing Returns Modeling
Saturation Effects
Trend Analysis
Seasonality Adjustment
Baseline and Trend Modeling
Media Response Functions
MODEL TRAINING AND BAYESIAN INFERENCE
Convergence Diagnostics
Markov Chain Monte Carlo Simulation
MCMC Sampling
Optimization Techniques
Approximate Bayesian Inference
Variational Inference
Performance Metrics
Cross-Validation
Model Validation
Trained Model
FORECASTING AND SCENARIO SIMULATION
Predictive Analytics
Budget Allocation Scenarios
Future Budget Forecasting
Sensitivity Analysis
What-If Scenarios
Counterfactual Analysis
Return on Investment Calculation
Profit Margins
Revenue and ROI Estimation
Data-driven Insights
OPTIMIZATION AND BUDGET ALLOCATION
Budget Constraints
Channel Allocation
Media Mix Optimization
Resource Limitations
Legal and Compliance Checks
Business Constraints Handling
Actionable Insights
Optimal Budget Allocation
Recommendation Generation
Optimised Marketing Budjet

Result





Previous Budget Allocation

Array([71.377396, 23.512438, 343.9902 , 9.596348, 386.7427 , 123.65756 , 90.39968 , 15.039522, 138.9256], dtype=float32)

Optimal Budget Allocation

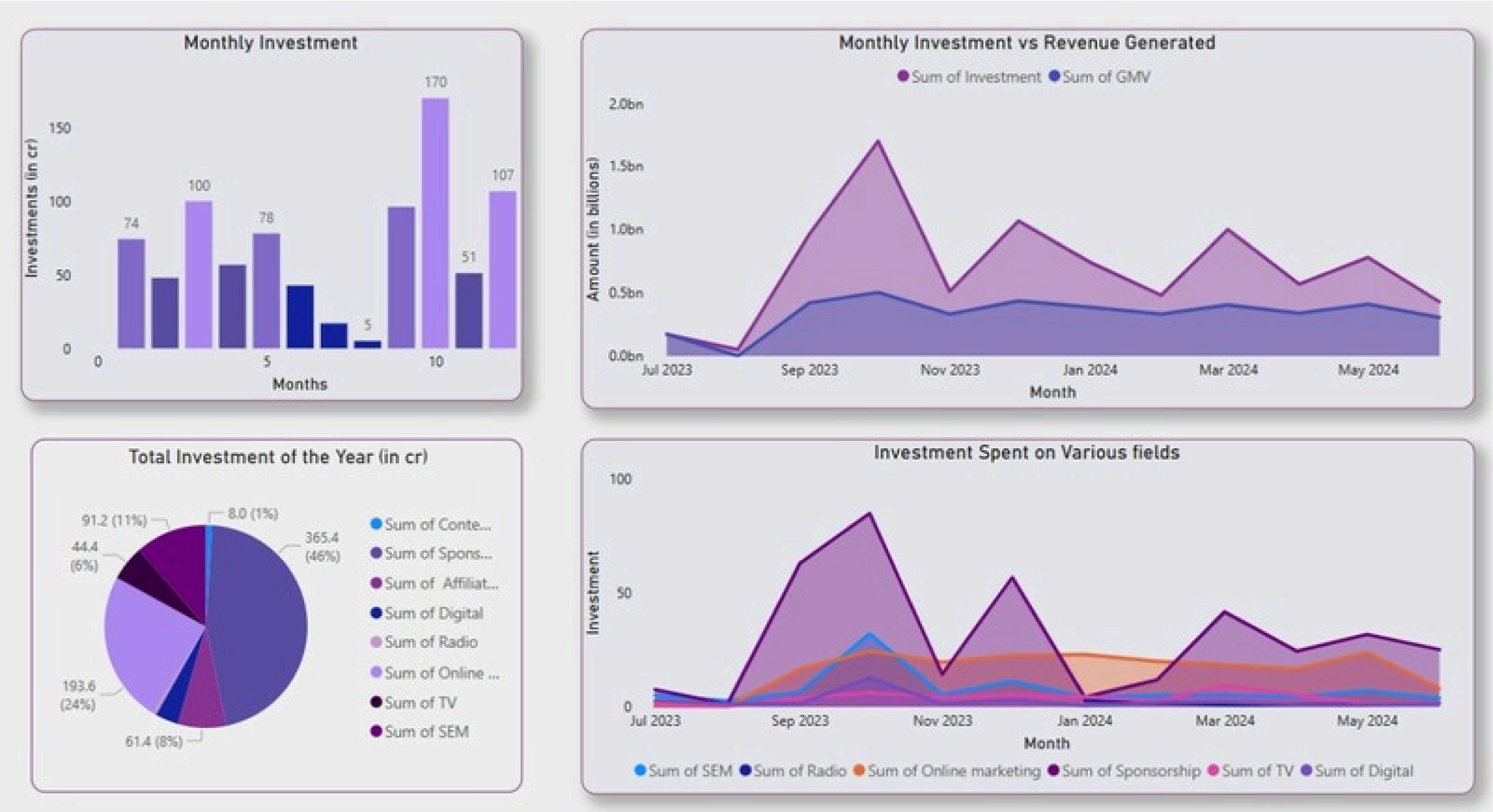
Array([85.683365, 28.224972, 322.10532 , 11.24245 , 309.50433 , 148.44191 , 105.58991 , 18.053852, 166.77005], dtype=float32)

THANK YOU

APPENDIX

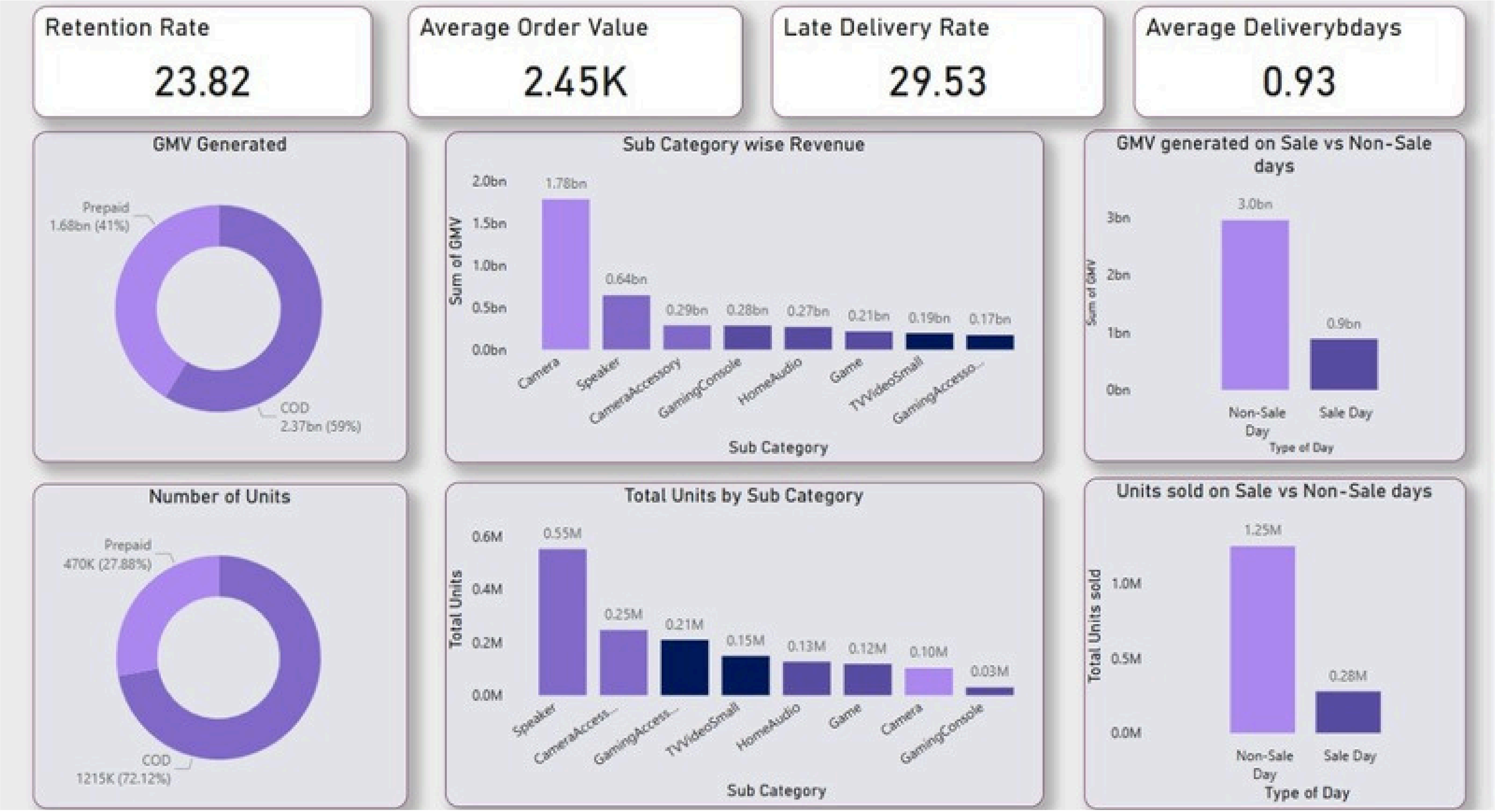

Total GMV of the Year
4.05bn
Total Order Placed
1.65M
Total Items Bought
1.69M
Number of Customers
1.20M
January
February
March
April
May
June
July
August
September
October
November
December
Total GMV per Category
2bn
1bn
0bn
Sum of GMV
1.8bn
1.3bn
0.5bn
0.3bn
0.2bn
Camera
Entertainm...
GamingHar...
CameraAcc...
GameCDD...
Category
Revenue generated per Month
0.6bn
0.4bn
0.2bn
0.0bn
Sum of GMV
Jul 2023
Oct 2023
Jan 2024
Apr 2024
Jul 2024
Months
Number of Units per Category
1.0M
0.5M
0.0M
Sum of GMV
0.96M
0.27M
0.24M
0.12M
0.10M
Entertain...
CameraAc...
GamingHa...
GameCDD...
Camera
Category
Total Units sold per Month
0.2M
0.1M
0.0M
Sum of Units
Jul 2023
Oct 2023
Jan 2024
Apr 2024
Jul 2024
Months

APPENDIX



Monthly Investment
Investments (in cr)
Months
74
100
78
5
170
51
107
Monthly Investment vs Revenue Generated
Sum of Investment
Sum of GMV
Amount (in billions)
Month
Total Investment of the Year (in cr)
91.2 (11%)
8.0 (1%)
365.4 (46%)
44.4 (6%)
193.6 (24%)
61.4 (8%)
Sum of Conte...
Sum of Spons...
Sum of Affiliat...
Sum of Digital
Sum of Radio
Sum of Online ...
Sum of TV
Sum of SEM
Investment Spent on Various fields
Investment
Month
Sum of SEM
Sum of Radio
Sum of Online marketing
Sum of Sponsorship
Sum of TV
Sum of Digital

APPENDIX



Retention Rate
23.82
Average Order Value
2.45K
Late Delivery Rate
29.53
Average Deliverybdays
0.93
GMV Generated
Prepaid
1.68bn (41%)
COD
2.37bn (59%)
Sub Category wise Revenue
Sum of GMV
2.0bn
1.5bn
1.0bn
0.5bn
0.0bn
1.78bn
0.64bn
0.29bn
0.28bn
0.27bn
0.21bn
0.19bn
0.17bn
Camera
Speaker
CameraAccessory
GamingConsole
HomeAudio
Game
TVVideoSmall
GamingAccesso...
Sub Category
GMV generated on Sale vs Non-Sale days
3bn
2bn
1bn
0bn
3.0bn
0.9bn
Non-Sale Day
Sale Day
Type of Day
Number of Units
Prepaid
470K (27.88%)
COD
1215K (72.12%)
Total Units by Sub Category
Total Units
0.6M
0.4M
0.2M
0.0M
0.55M
0.25M
0.21M
0.15M
0.13M
0.12M
0.10M
0.03M
Speaker
CameraAccess...
GamingAccess...
TVVideoSmall
HomeAudio
Game
Camera
GamingConsole
Sub Category
Units sold on Sale vs Non-Sale days
Total Units sold
1.0M
0.5M
0.0M
1.25M
0.28M
Non-Sale Day
Sale Day
Type of Day