TEAM 2025108

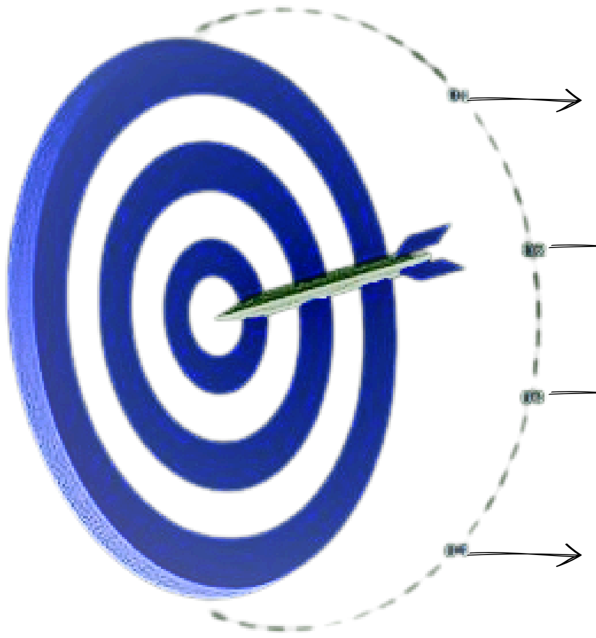
Table of contents

PROBLEM STATEMENT OVERVIEW

PROBLEM STATEMENT

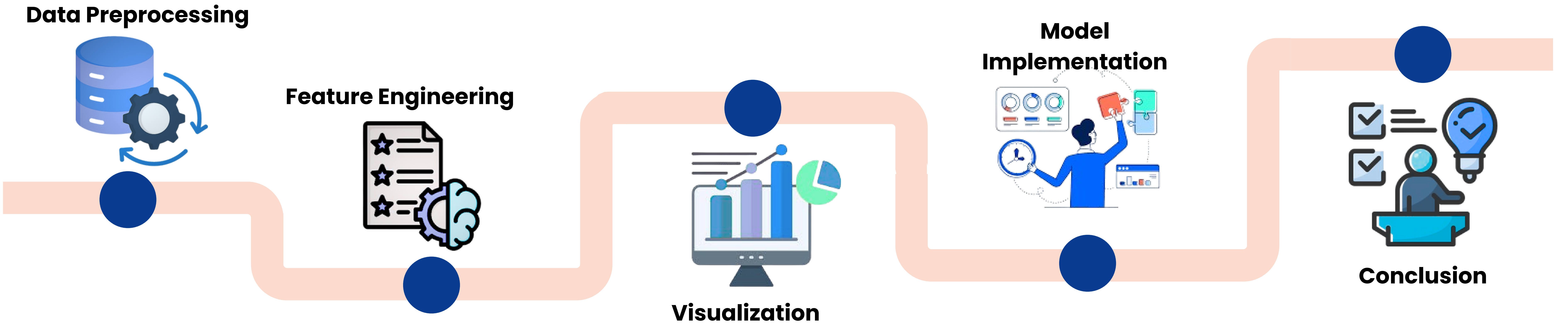
- ElectroMart, a fast-growing e-commerce retailer, aims to maximize marketing ROI after heavy ad spending with limited revenue growth.
- Key challenges include inefficient budget allocation, ineffective promotions, and a lack of integrated insights on seasonality and economic factors.

OBJECTIVES



- Key Performance Indicators (KPIs) Evaluation
- Marketing ROI Analysis
- Optimized Budget Allocation
- Data-Driven Decision Making

WORKFLOW



DATA OVERVIEW & PREPROCESSING

DATA DESCRIPTION

Datasets used to assess marketing effectiveness

1. **Customer Orders**
2. **Media Spend**
3. **Weather**
4. **Promotions**
5. **NPS Ratings**
6. **Holidays**

Key variables like ad spend, order value, and discounts were analyzed using statistical techniques and predictive models for budget optimization and revenue forecasting.

DATA PREPROCESSING

Customer Orders

- Filtered date range
- Handled missing values
- Removed duplicates

Media & Sales Data

- Cleaned media investment records
- Structured sales calendar
- Mapped revenue to key events.

Payday & Holiday Impact

- Validated dates
- Analyzed revenue trends around key periods.

NPS & Stock Index

- Aligned customer satisfaction scores with marketing spend for correlation analysis

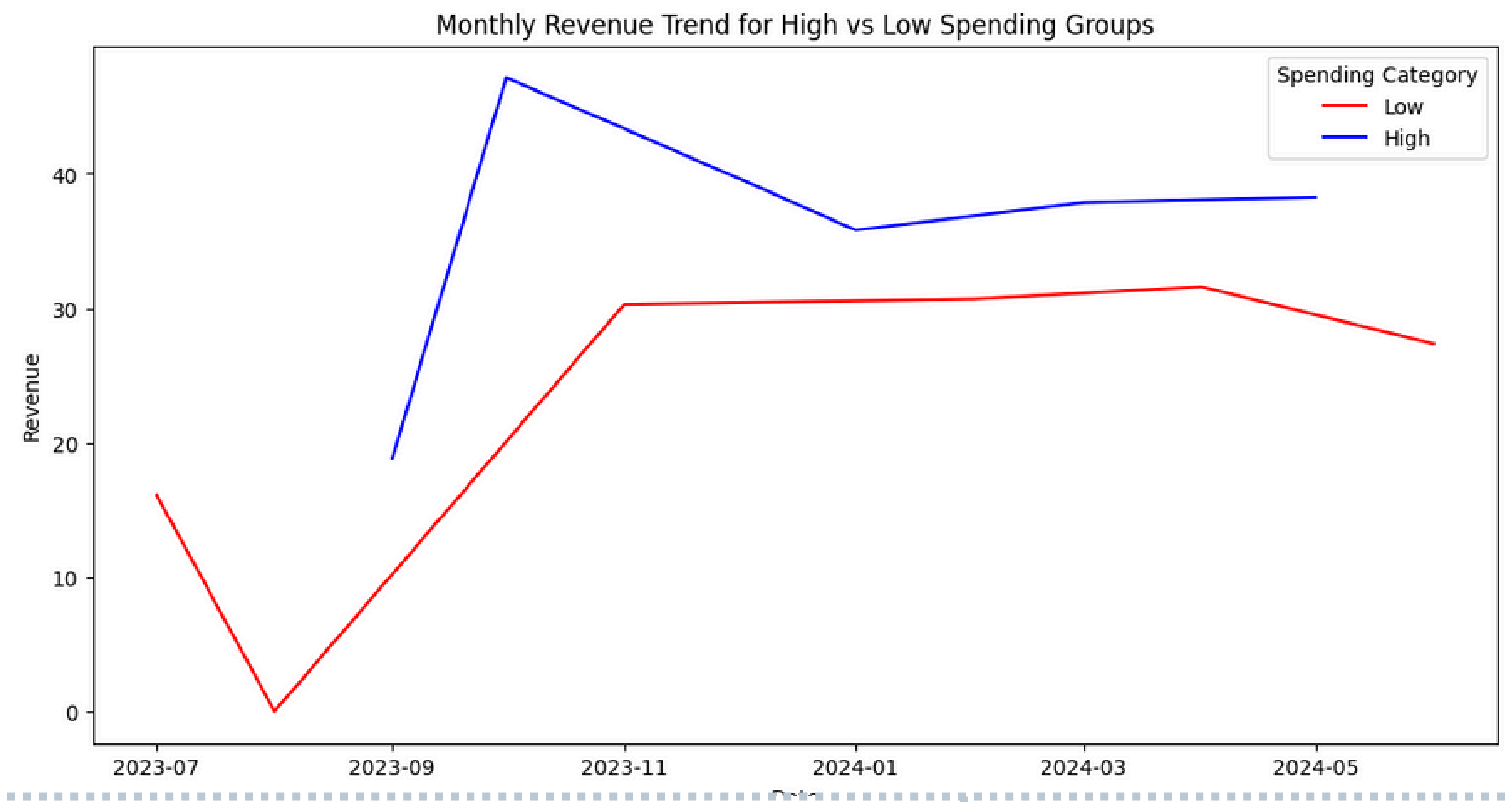
Product Hierarchy

- Organized category-wise data for deeper revenue insights.

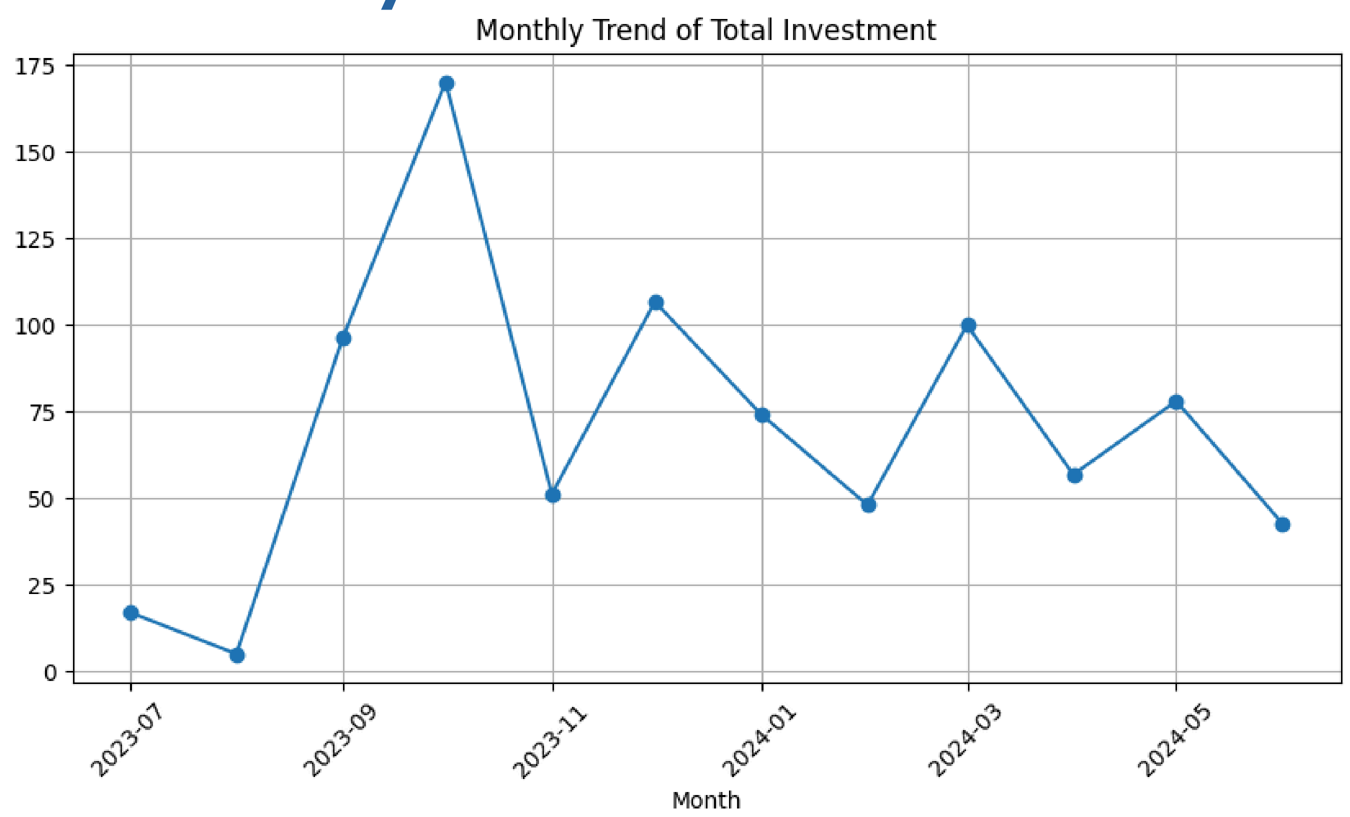
EXPLORATORY DATA ANALYSIS

MARKETING INVESTMENT & REVENUE ANALYSIS

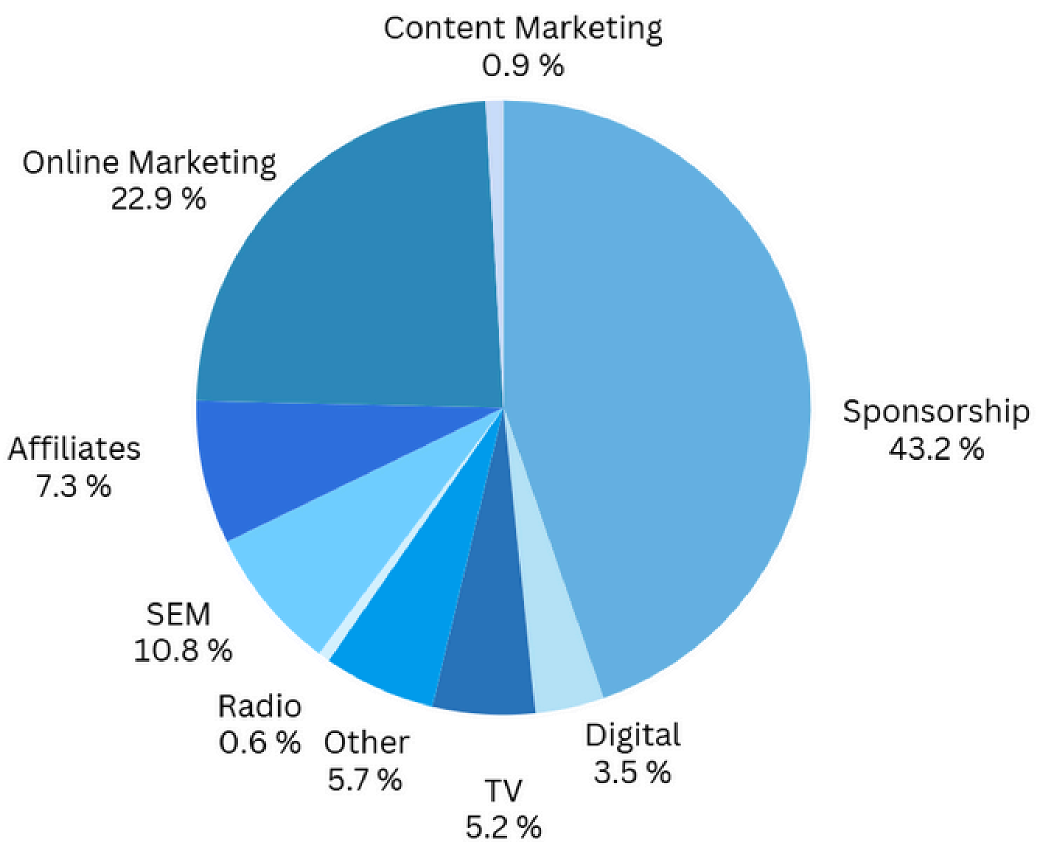
Monthly Revenue Trend for High vs Low Spending Groups



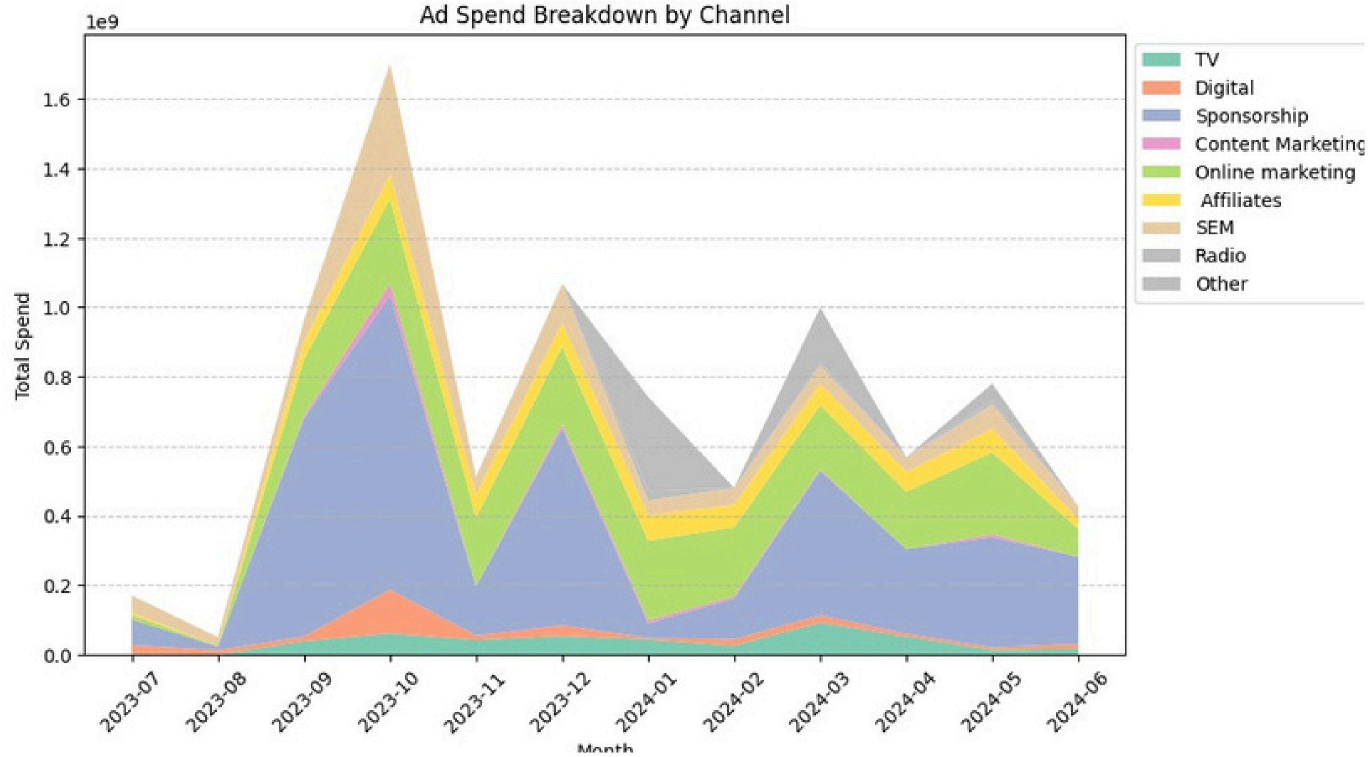
Monthly Trend of Total Investment



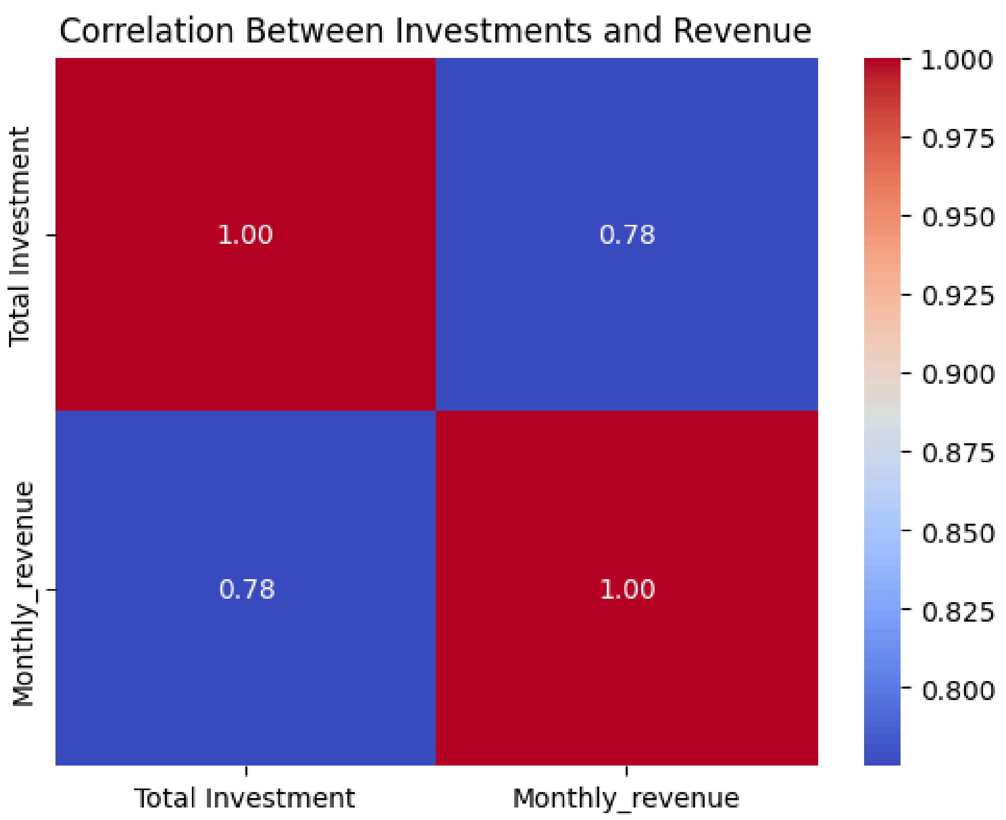
Breakdown of Investment by Channel



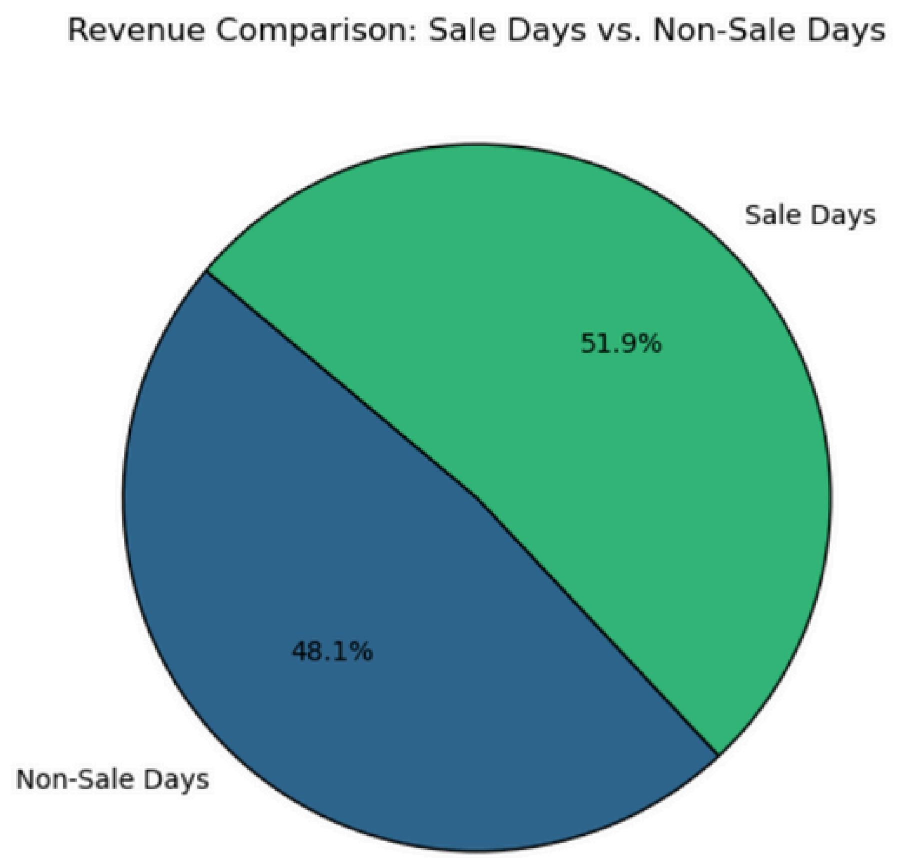
Trends in Ad Investments Across Channels

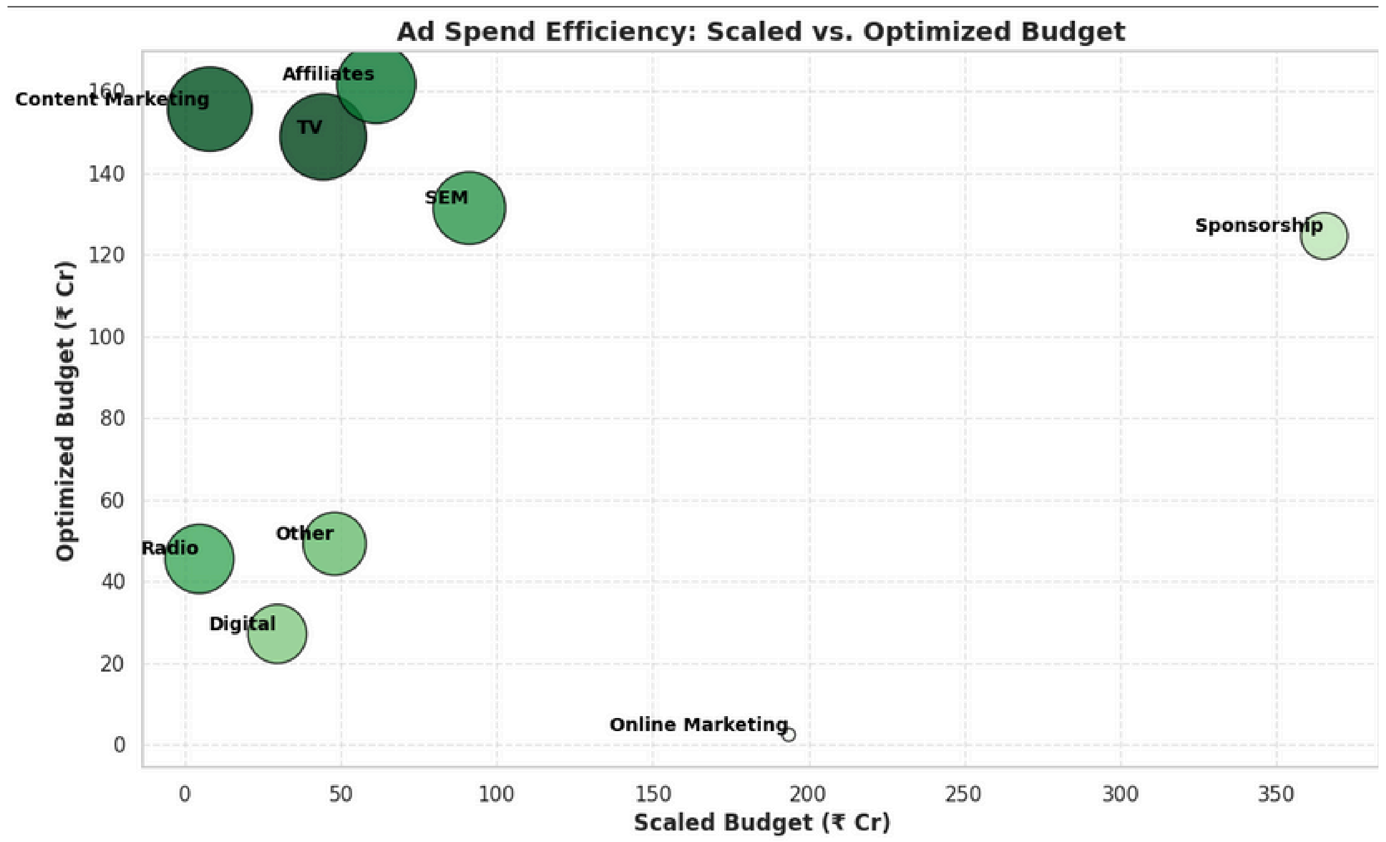


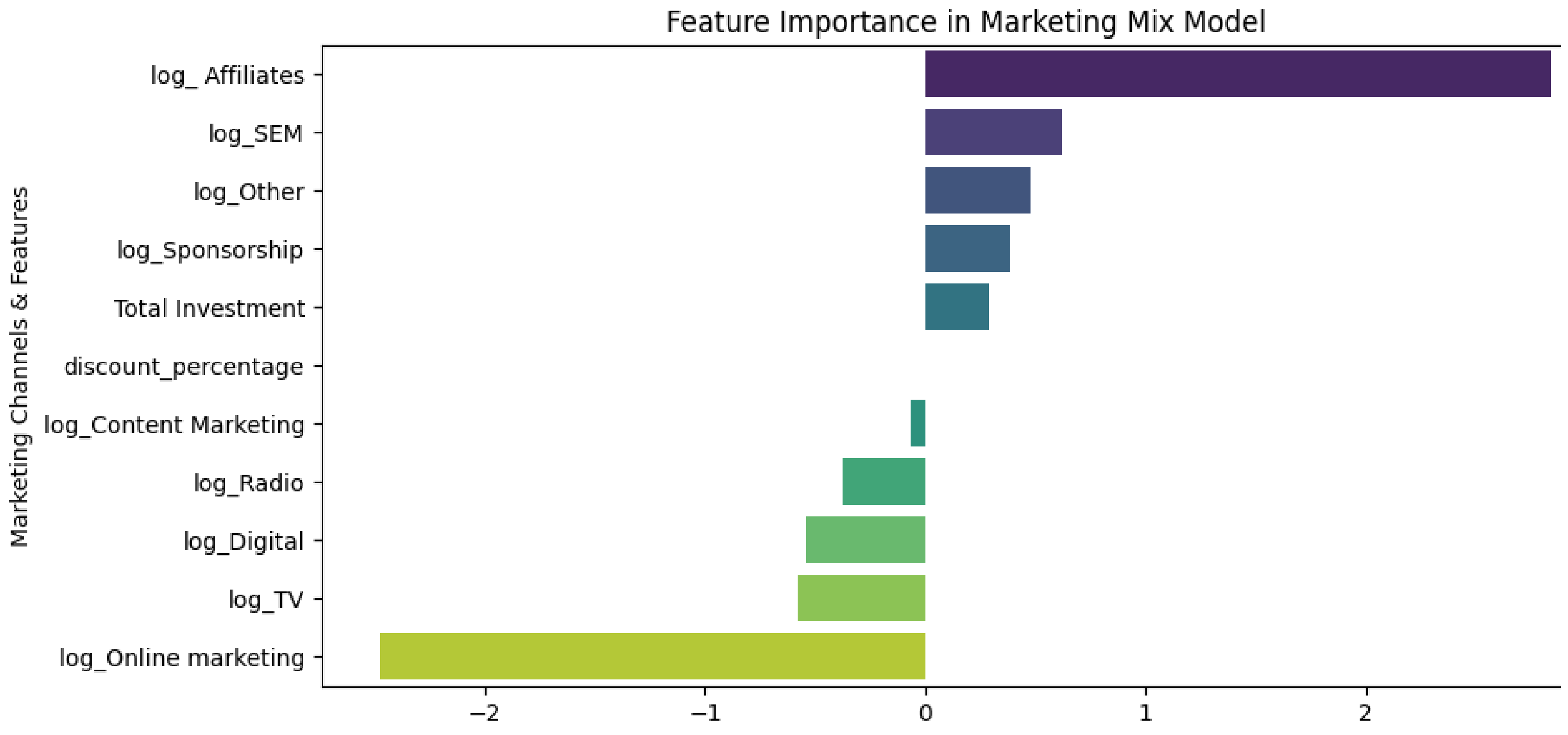
Investment vs. Revenue

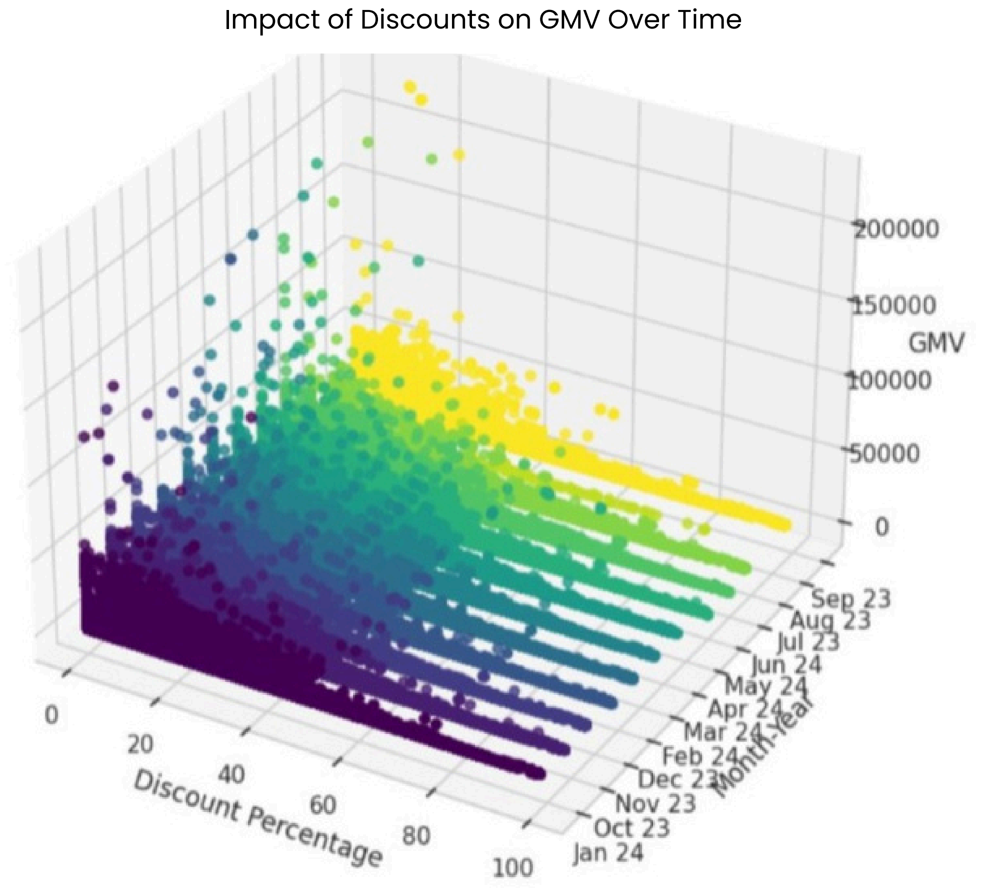


IMPACT OF SALES & DISCOUNT

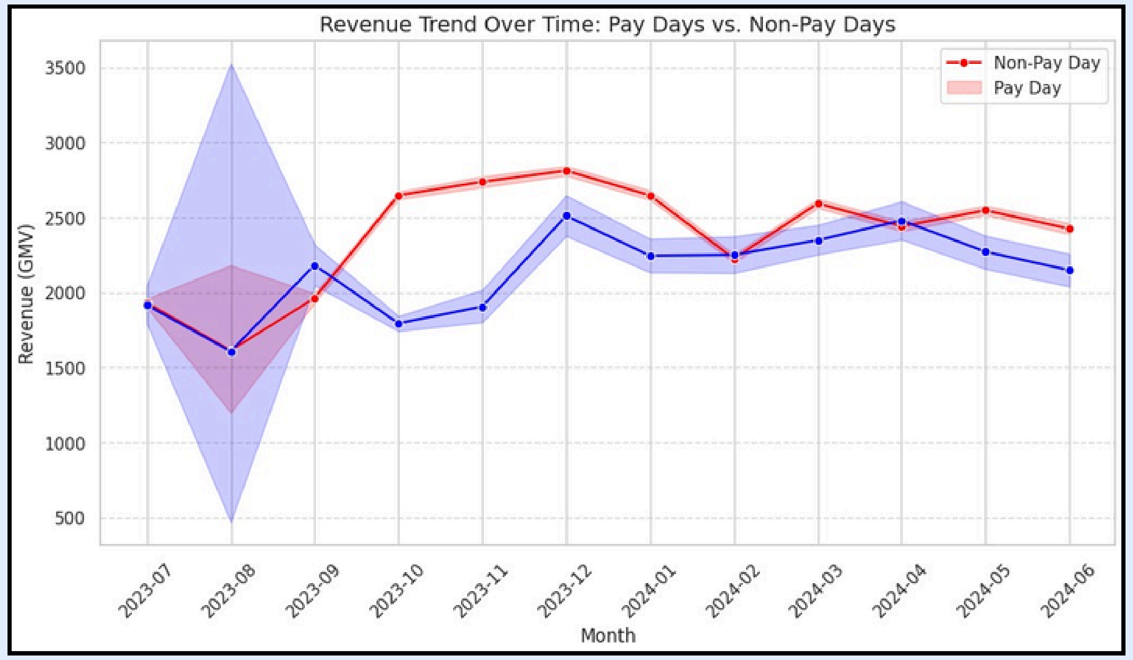








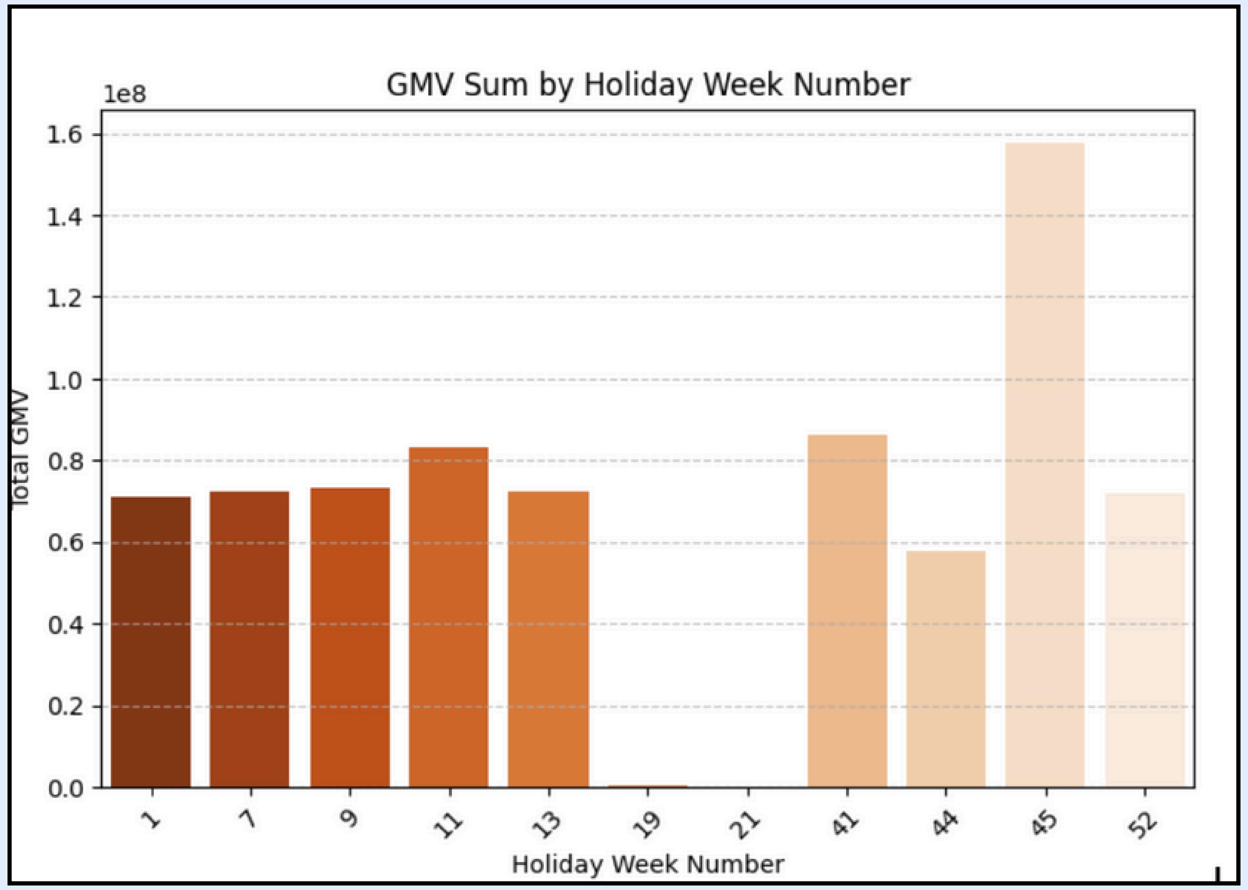
REVENUE TRENDS: PAY DAYS AND HOLIDAYS

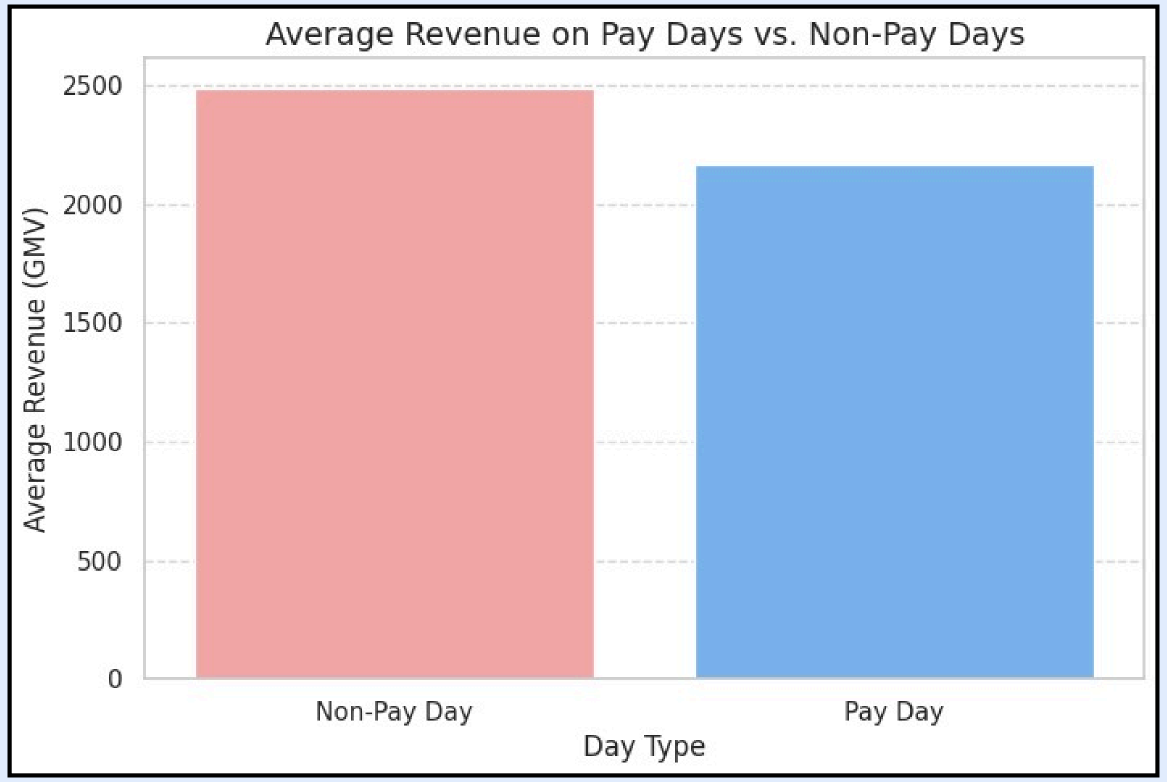


Revenue is consistently higher on non-pay days than on pay days, with smaller fluctuations over time.

GMV peaks during major holidays (Weeks 45 & 52), with moderate spikes around Halloween, steady early-year sales, and lows in weeks with minor holidays.

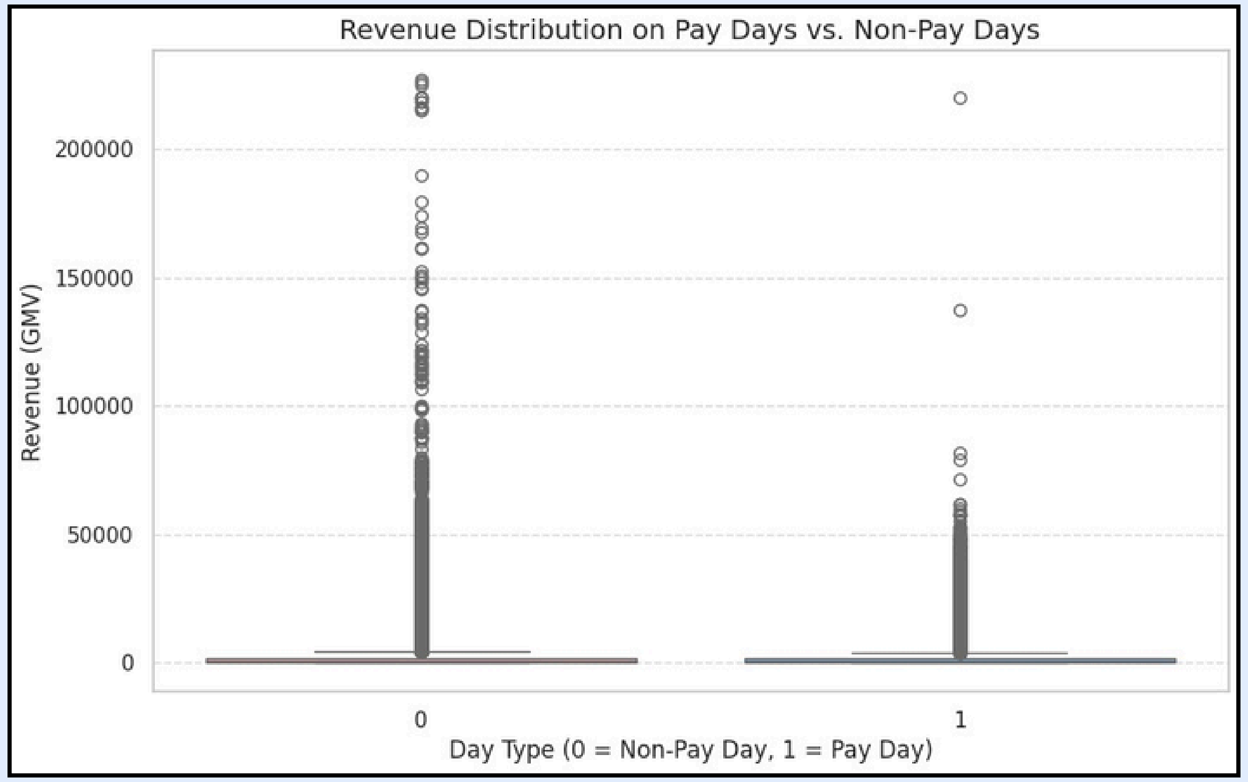
Revenue on both pay and non-pay days shows a similar median, but pay days have fewer extreme outliers, indicating slightly less variability.





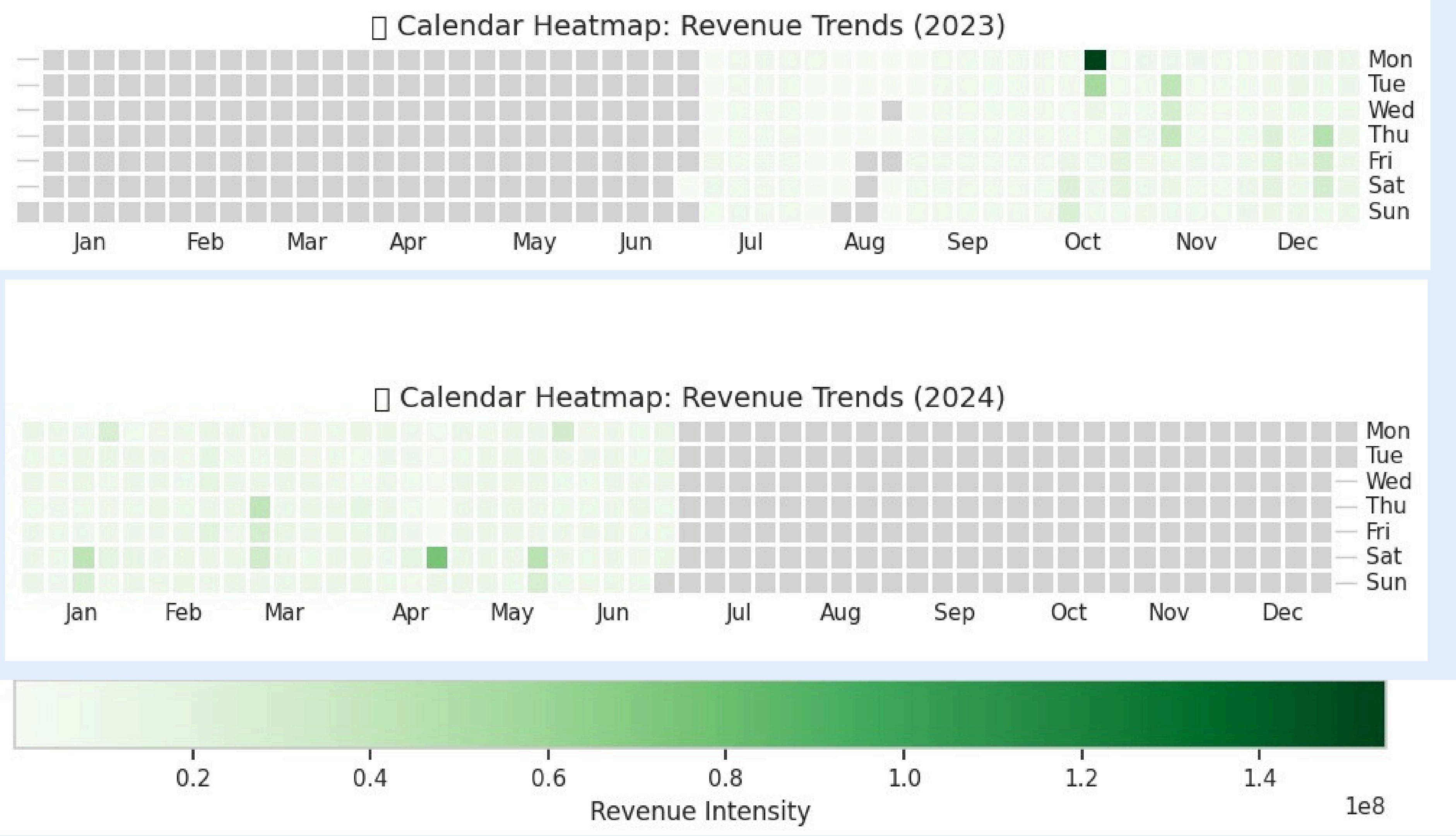
Pay Day and Non Pay Day Revenue

Revenue distribution is similar on pay and non-pay days, but both have high-value outliers, with occasional revenue spikes.



KEY DATA PATTERNS & TRENDS

Calendar Heatmap



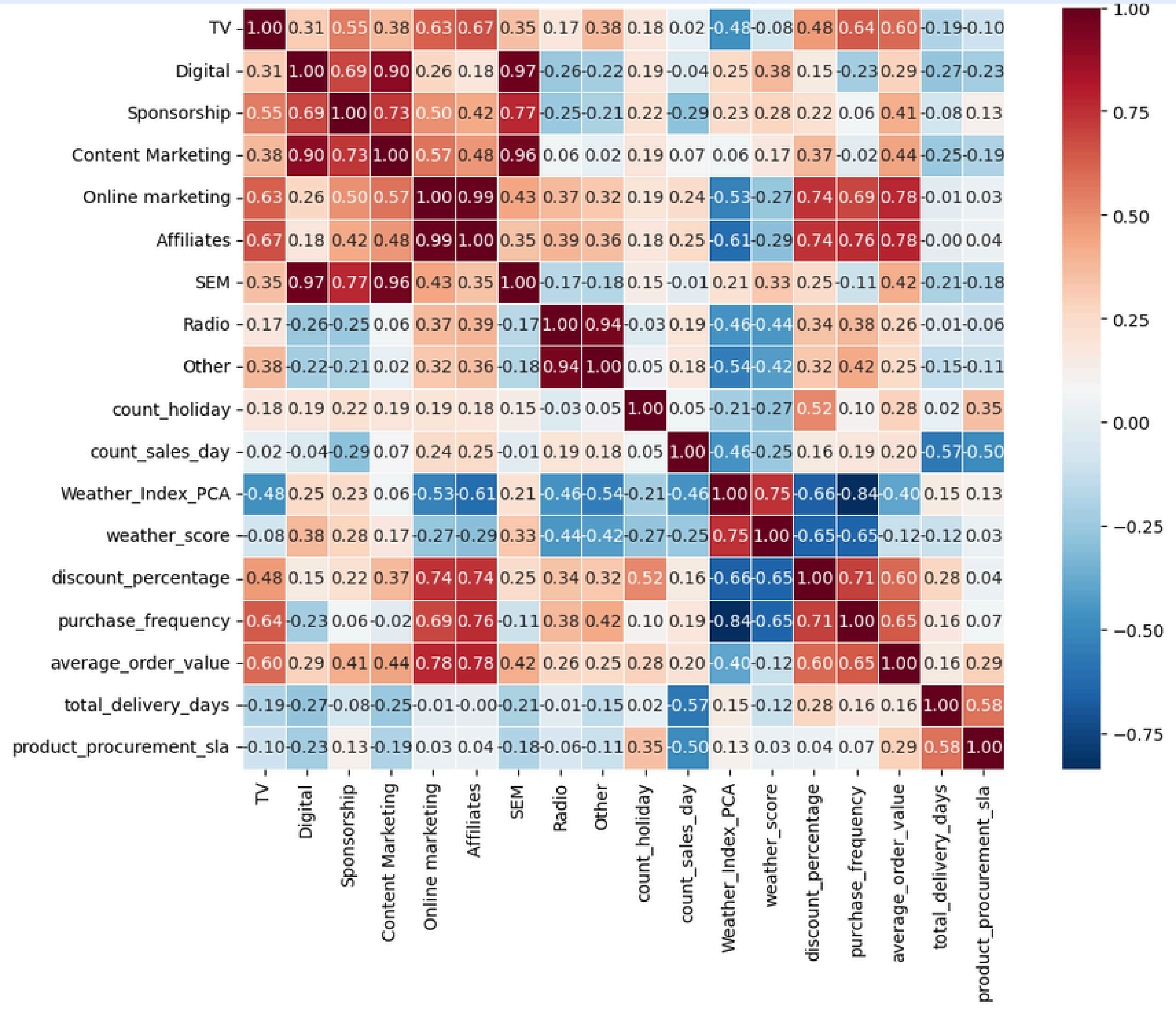
Each block represents a day, with darker shades indicating higher revenue generated on that day.

Radar Plot

Radar plots show how key factors like units sold, discount percentage, and weather score vary across daily data clusters, highlighting unique patterns in each cluster.

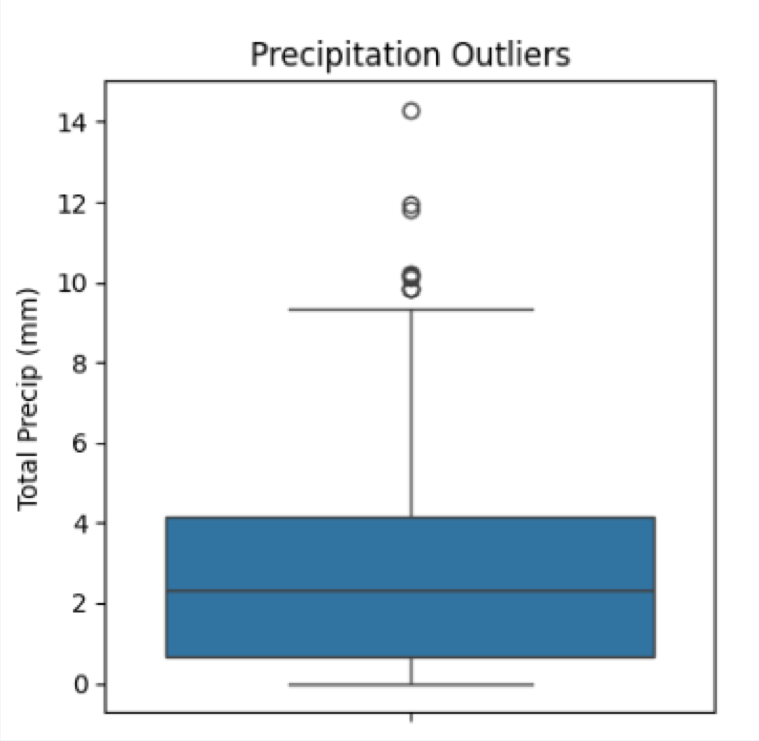
VARIABLE CORRELATIONS & OUTLIERS

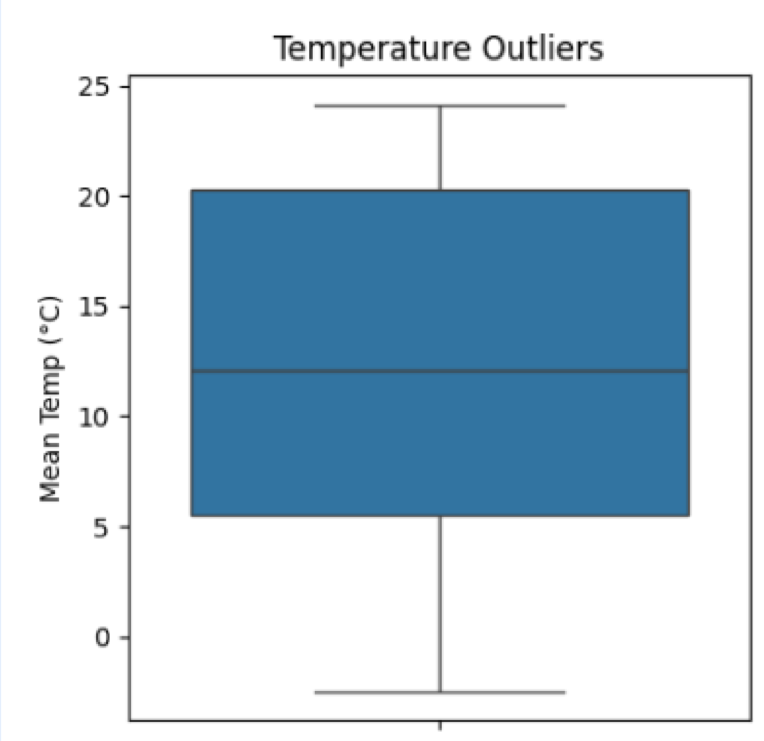
Correlation Matrix



The heatmap shows pairwise correlations between variables, with warm colors (red) indicating strong positive relationships and cool colors (blue) indicating strong negative ones. This helps identify patterns, dependencies, and potential multicollinearity in the data.

Outlier Detection



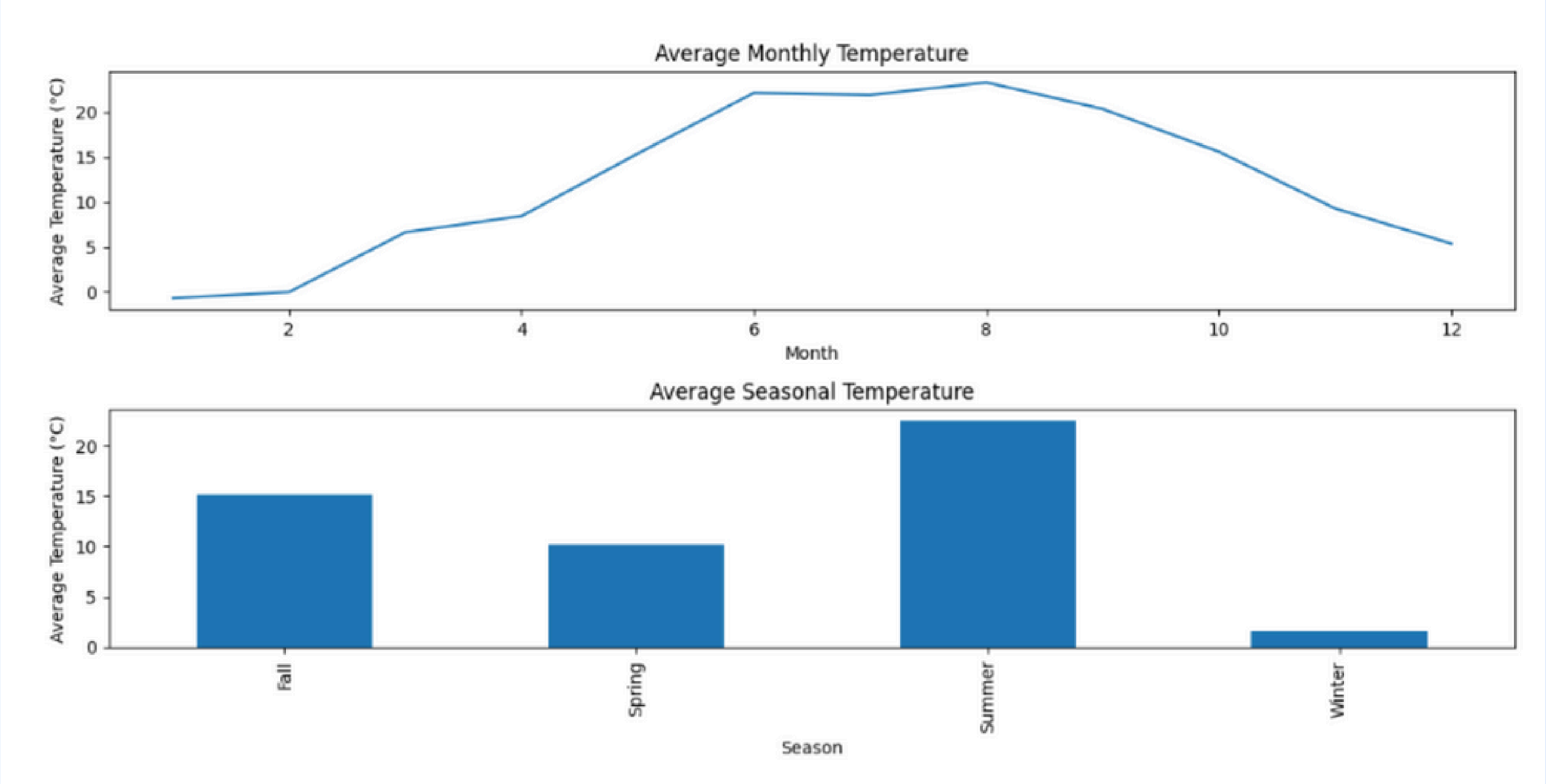


- Boxplots were used to identify outliers in temperature and precipitation.
- Unusual temperature spikes or extreme precipitation events were flagged for further investigation.

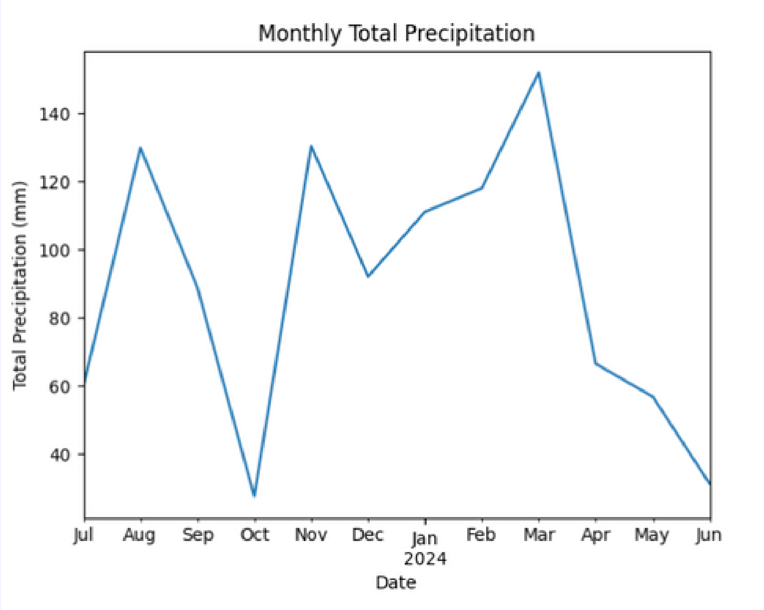
Several extreme precipitation values indicate unusually high rainfall events, while temperature data shows no significant outliers, suggesting stable variations.

SEASONAL WEATHER PATTERNS

Seasonal Trends

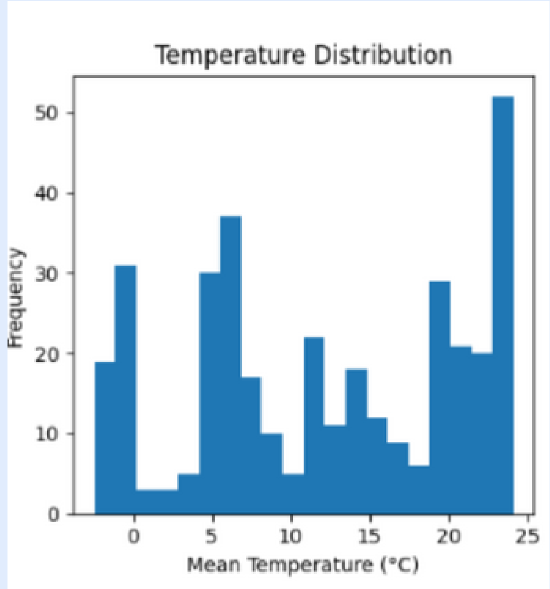


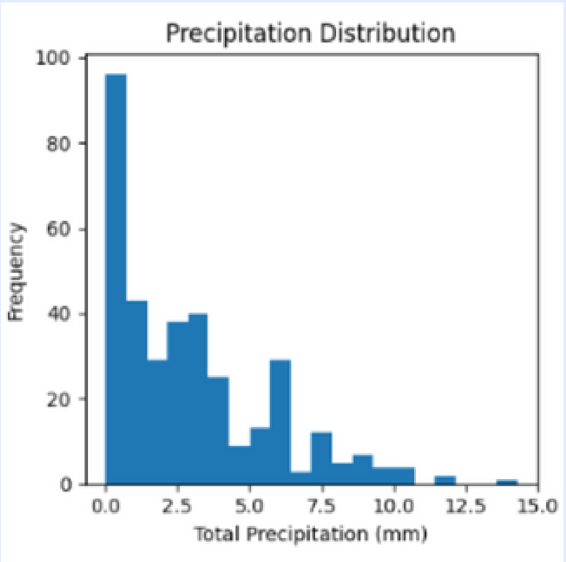
Monthly Total Precipitation

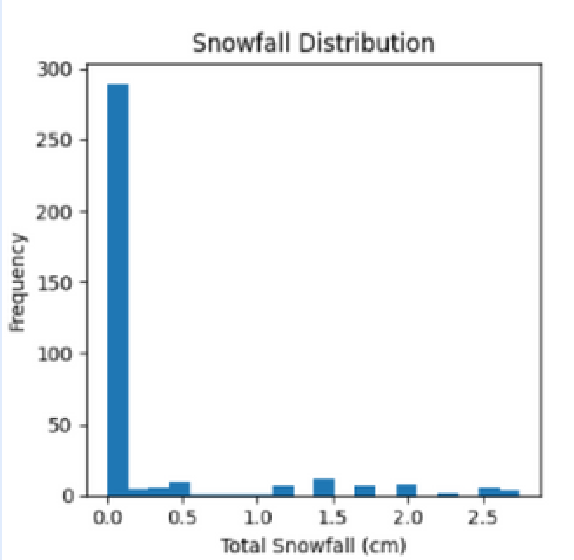


Precipitation peaks in August, November, and March, with March 2024 highest, showing seasonal variation and wetter winters and late summers.

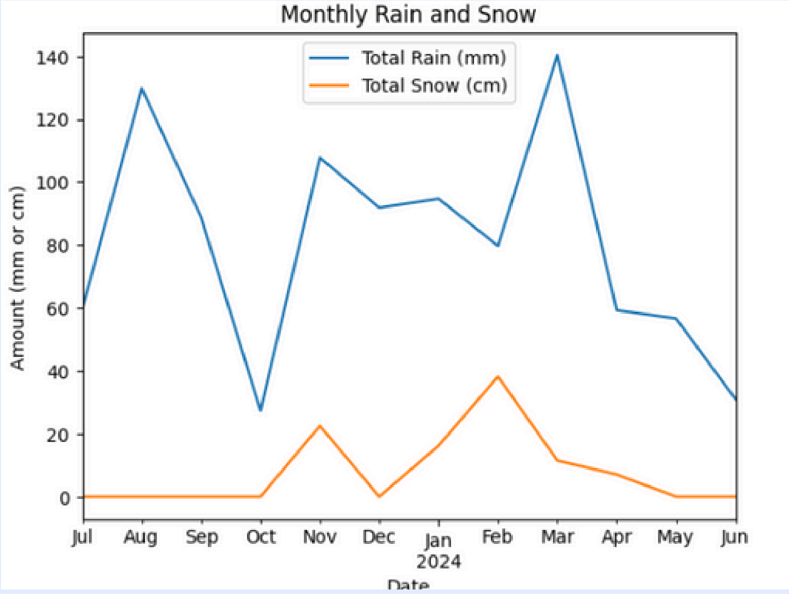
Temperature, Precipitation, and Snowfall Trends







The histogram analysis revealed the distribution of average temperature, precipitation, and snowfall, assessing normality and flagging extreme values.

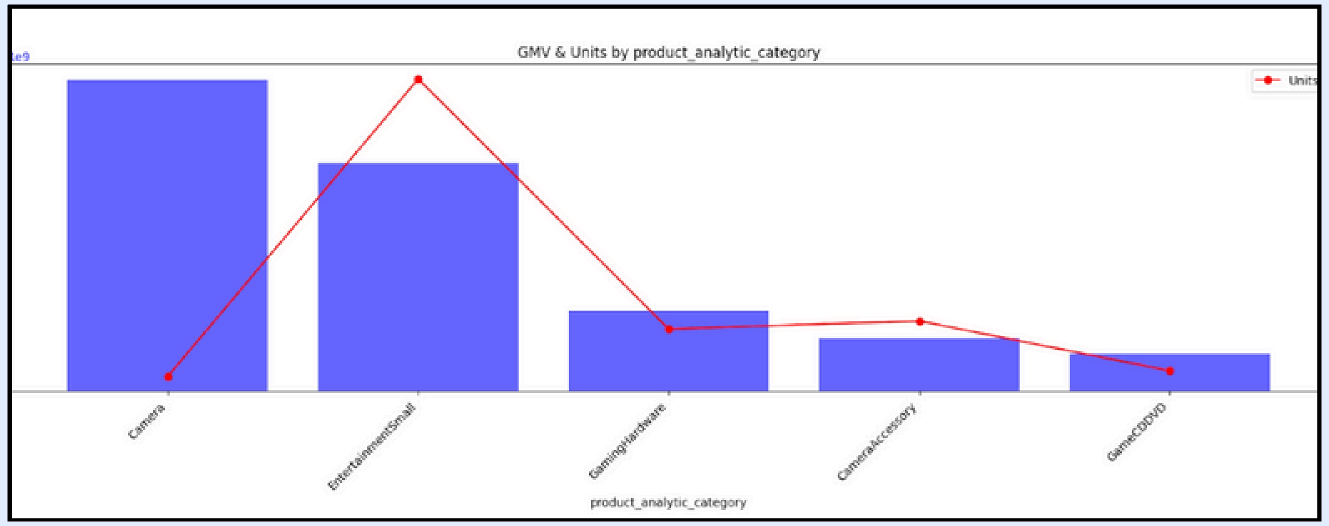


Monthly Rain and Snow Trend

Rainfall peaks in August and March, while snowfall is highest in November and February, showing an inverse, seasonal pattern.

Revenue Insights: Categories, Luxury vs. Mass-Market, and Discounts

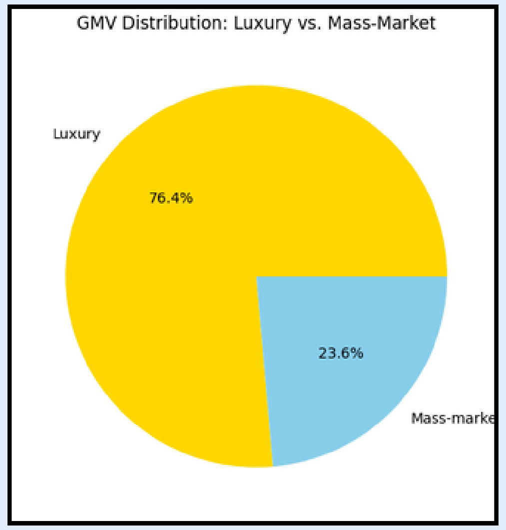
Category-Wise Revenue & Units



- Cameras dominate GMV, followed by EntertainmentSmall with strong market interest.
- GamingHardware is a niche but valuable segment.
- Camera Accessories & GameCD/DVD have lower GMV due to sales volume or pricing.
- EntertainmentSmall sees high unit sales despite lower GMV.

High-value electronics drive revenue, making them essential for strategic focus.

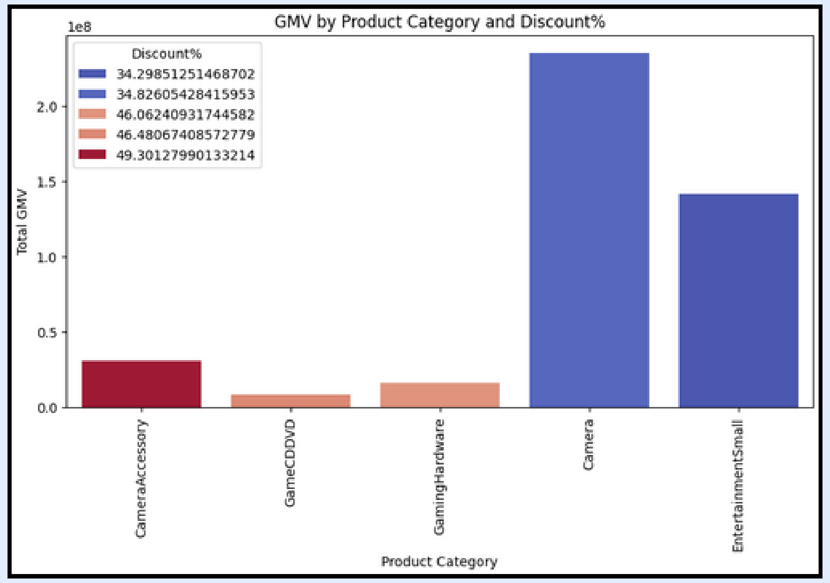
Luxury vs. Mass-Market Product Sales

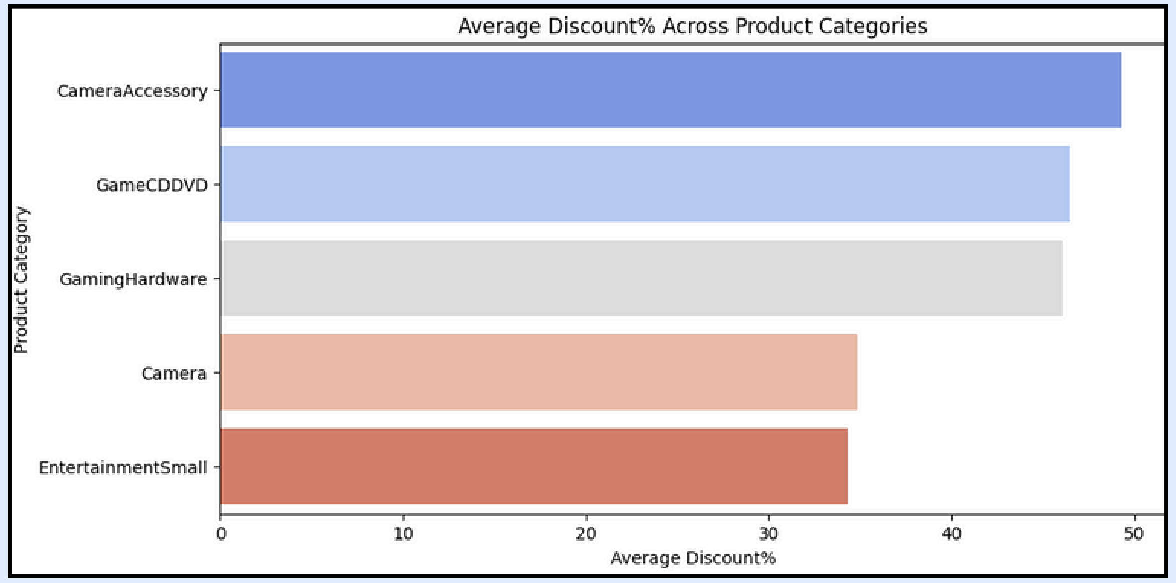


- Luxury products dominate GMV, contributing 76.4%.
- Mass-market products account for only 23.6%, suggesting lower revenue contribution.

High-value products drive overall revenue, making luxury a key segment for profitability.

Discount Effect on Different Categories





- Camera Accessories, GameCDDVD, and Gaming Hardware have the highest discounts but generate lower GMV.
- Camera and EntertainmentSmall have relatively lower discounts but dominate GMV

Moderate discounts indicate strong demand, while high discounts don't always boost revenue.

Order Fulfillment, Payment Preferences & Revenue Trends

Revenue vs. SLA & Delivery Days

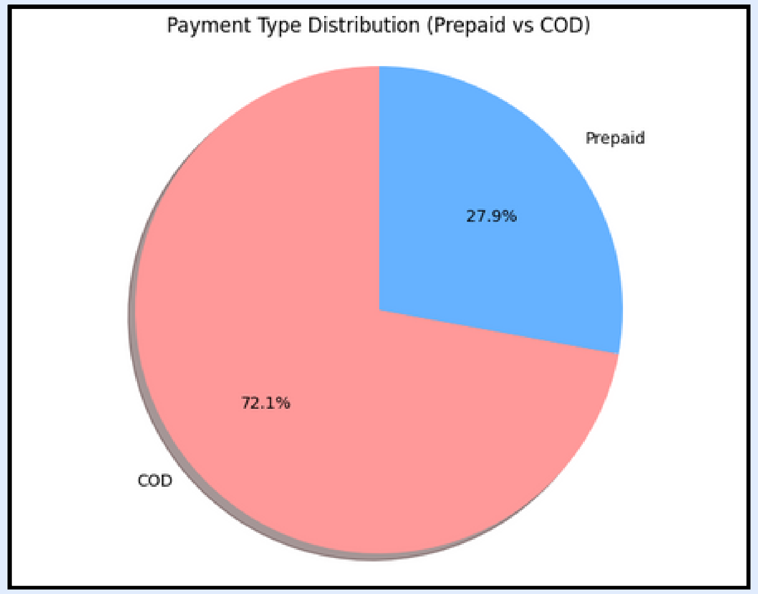
Most expected delivery times (SLA) are concentrated at lower values.

Actual delivery days tend to be higher than expected for some transactions.

Higher revenue orders have longer delivery times. Many orders exceed expected delivery times. Lower delivery time has given lesser revenue.

High-value orders are prioritized for faster delivery while lower-value orders frequently experience fulfillment delays.

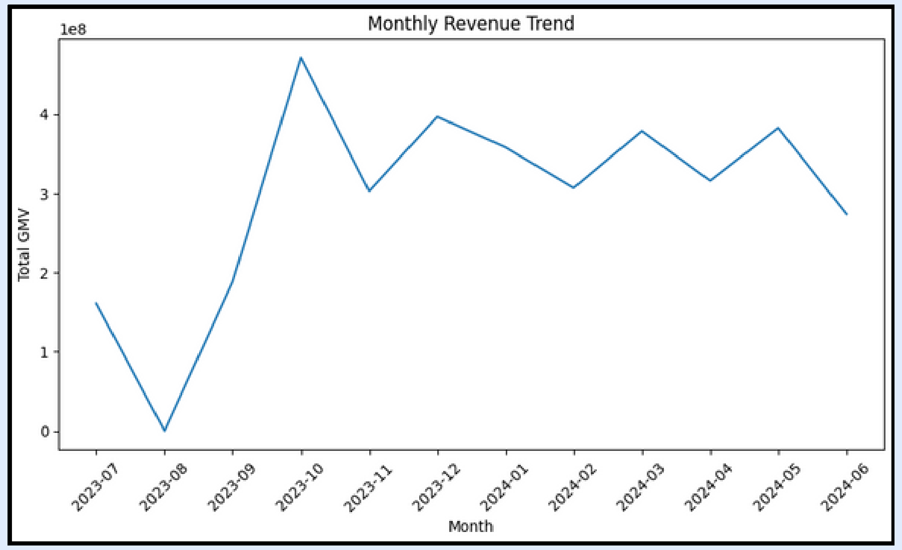
Payment Type (Prepaid vs. COD) Distribution

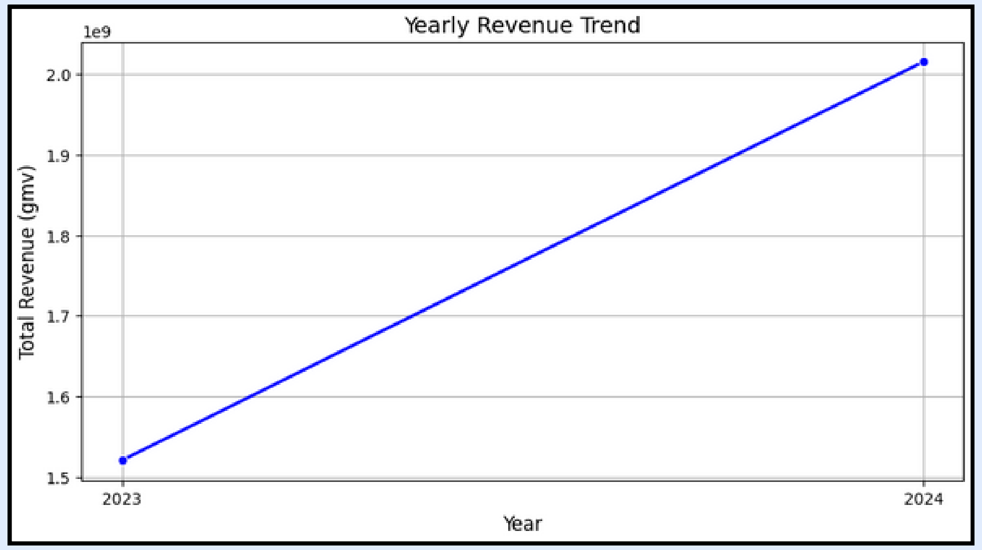


Cash on Delivery (COD) accounts for 72.1% of transactions while only 27.9% of orders are prepaid.

There is a strong preference for post-payment purchases.

Revenue growth





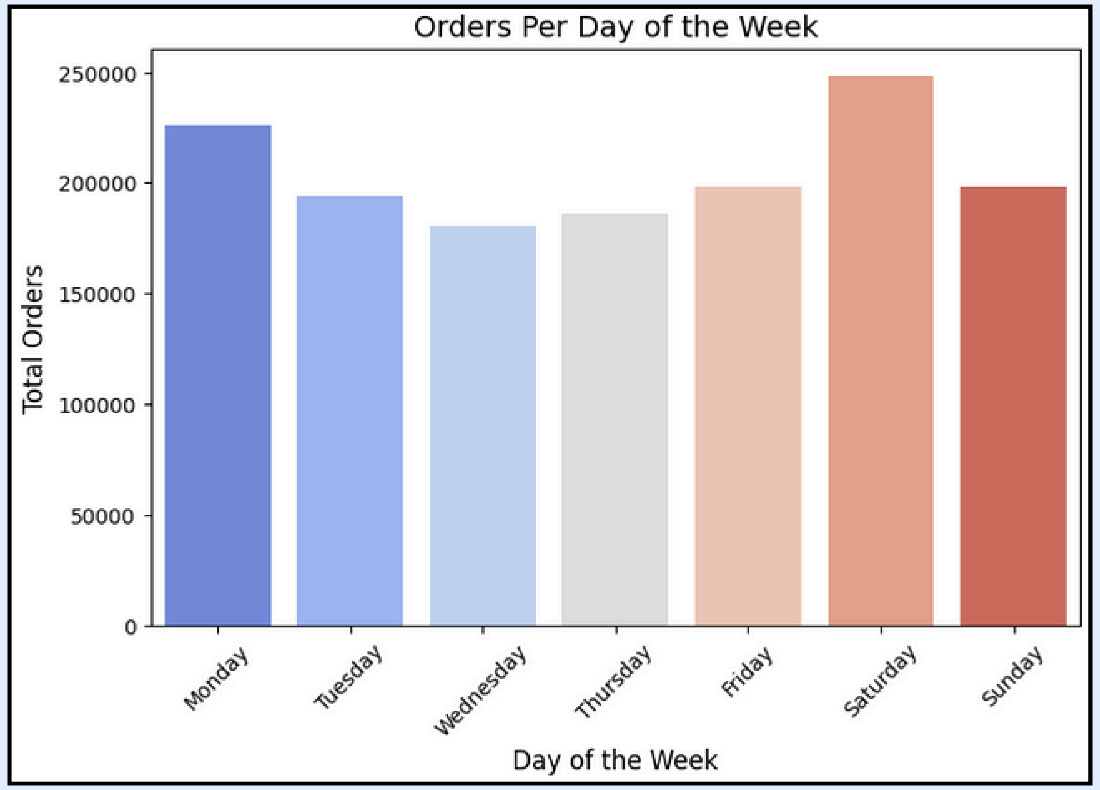
Revenue spiked in Oct 2023, fluctuated in early 2024, peaked in Mar & May, then declined in June.

Revenue (GMV) increased significantly from 2023 to 2024 in a linear trend.

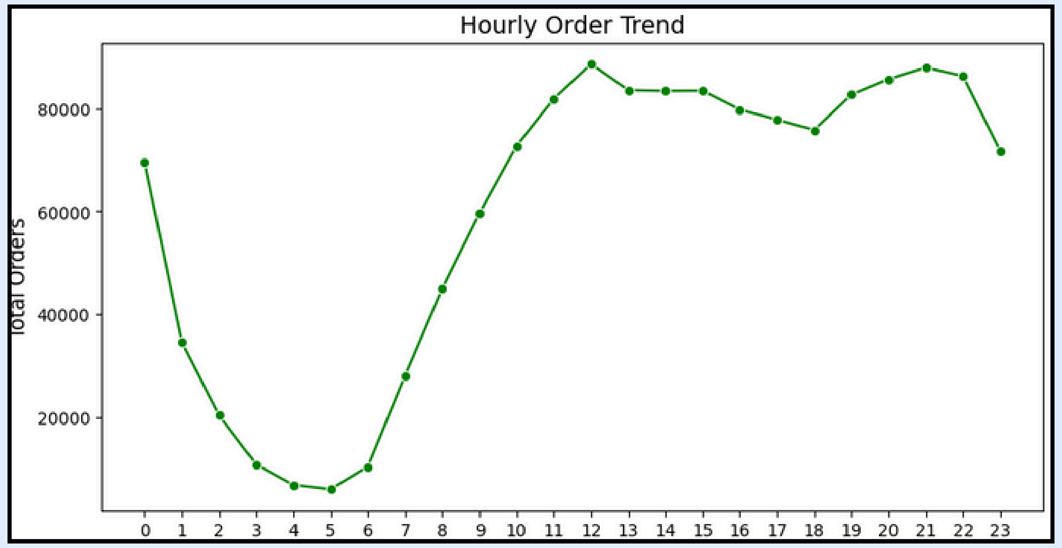
The business has experiencing consistent overall revenue growth but with noticeable monthly volatility.

Order and Revenue Trends: Daily, Hourly, and Seasonal Insights

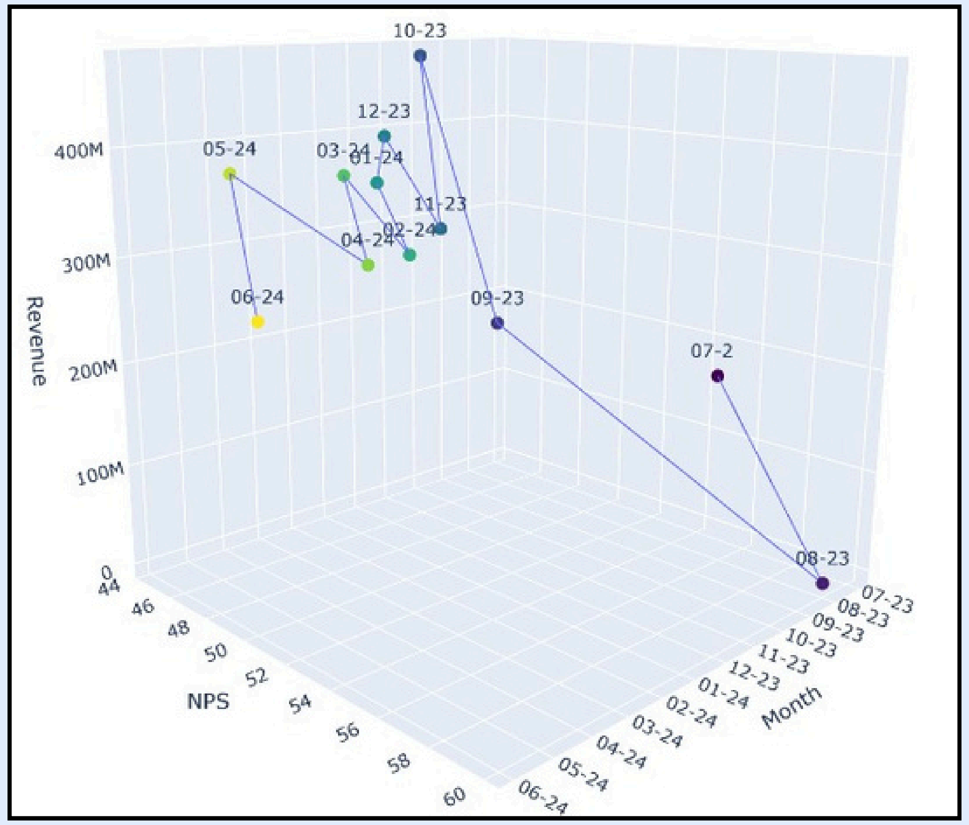
Orders Per Day of the Week



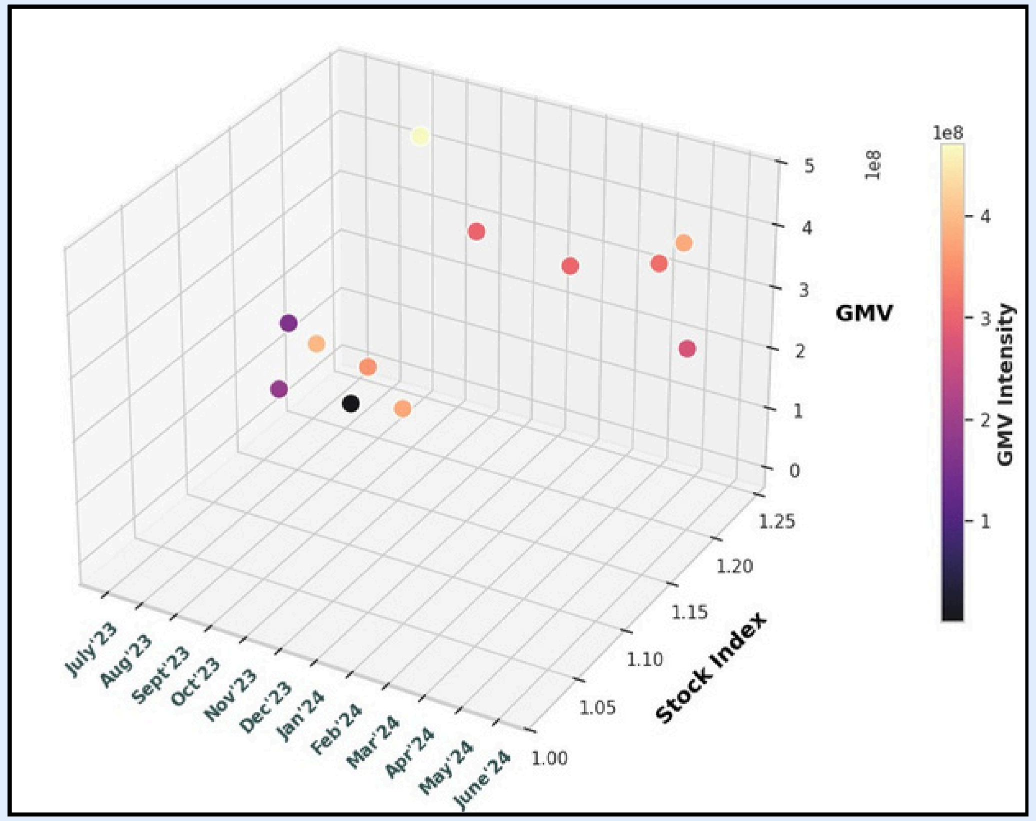
Saturday sees the highest order volume, followed by Monday, while midweek (Tuesday-Thursday) experiences lower orders. Friday and Sunday are moderate, with a slight dip on Sunday.



3D Trend of NPS, Revenue, and Time



Revenue peaks in Oct 2023, May 2024, and June 2024 suggest promotions, seasonal demand, or market influences.



Certain months see higher GMV due to seasonal trends, promotions, or market influences, making them key for strategic sales planning.

Hourly Order Trend

Orders peak around midnight, drop sharply until early morning, then rise steadily after 6 AM, peaking between 12 PM - 1 PM before stabilizing in the afternoon with a slight evening dip.

FEATURE ENGINEERING

FEATURE ENGINEERING

Category & Customer Insights Features

- Product Category Performance Score = Total Revenue from Category / Total Units Sold
- Luxury vs. Mass-Market Products = Defined by GMV > 80th percentile
- Customer Retention Rate = (Repeat Customers / Total Customers) × 100
- Order Frequency Score = Total orders per customer over a set period
- Sale Period Flag = 1 if order falls within a sale period, else 0

Sales & Pricing Features

- List Price = GMV * Units
- Discount % = 100 × (Product MRP – List Price) / Product MRP
- Delivery Time (Total) = Deliverybdays + Deliverycdays
- Customer Purchase Frequency = Total orders per customer
- Average Order Value (AOV) = GMV / Total Orders
- Revenue = ∑GMV (analyzed monthly, during sales, payday weeks, holidays, etc.)

Seasonality & External Impact Features

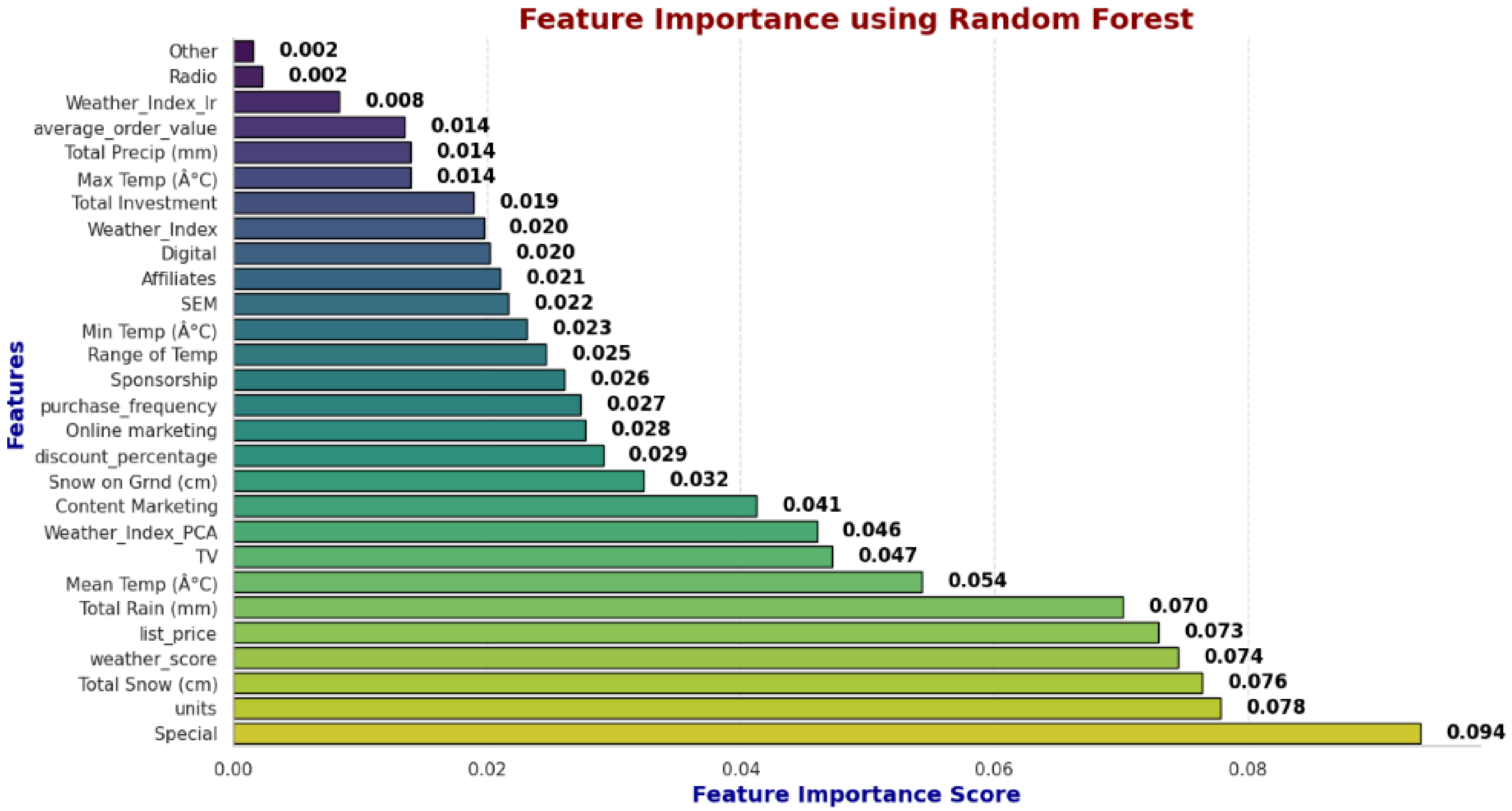
- Payday Week Flag = 1 if order falls within payday week, else 0
- Holiday Week Flag = 1 if order falls within a holiday period, else 0
- NPS Growth Rate = Current Month NPS - Previous Month NPS
- Revenue Change % = (Current Month Revenue - Previous Month Revenue) / Previous Month Revenue

Weather-Based Feature Engineering

- Extreme Temperature Flag → 1 if Max Temp < -3°C or >23°C, else 0.
- Rainy Day Indicator → 1 if Total Rain > 0, else 0.
- Snowy Day Indicator → 1 if Total Snow > 0, else 0.
- 7-Day Rolling Avg Temp → Captures temperature trends over a week.
- Lag Features → Yesterday's temperature & precipitation to analyze weather persistence effects.
- Weather Impact Index → Custom score based on temperature deviation, wind speed, and precipitation to assess extreme weather effects on sales.

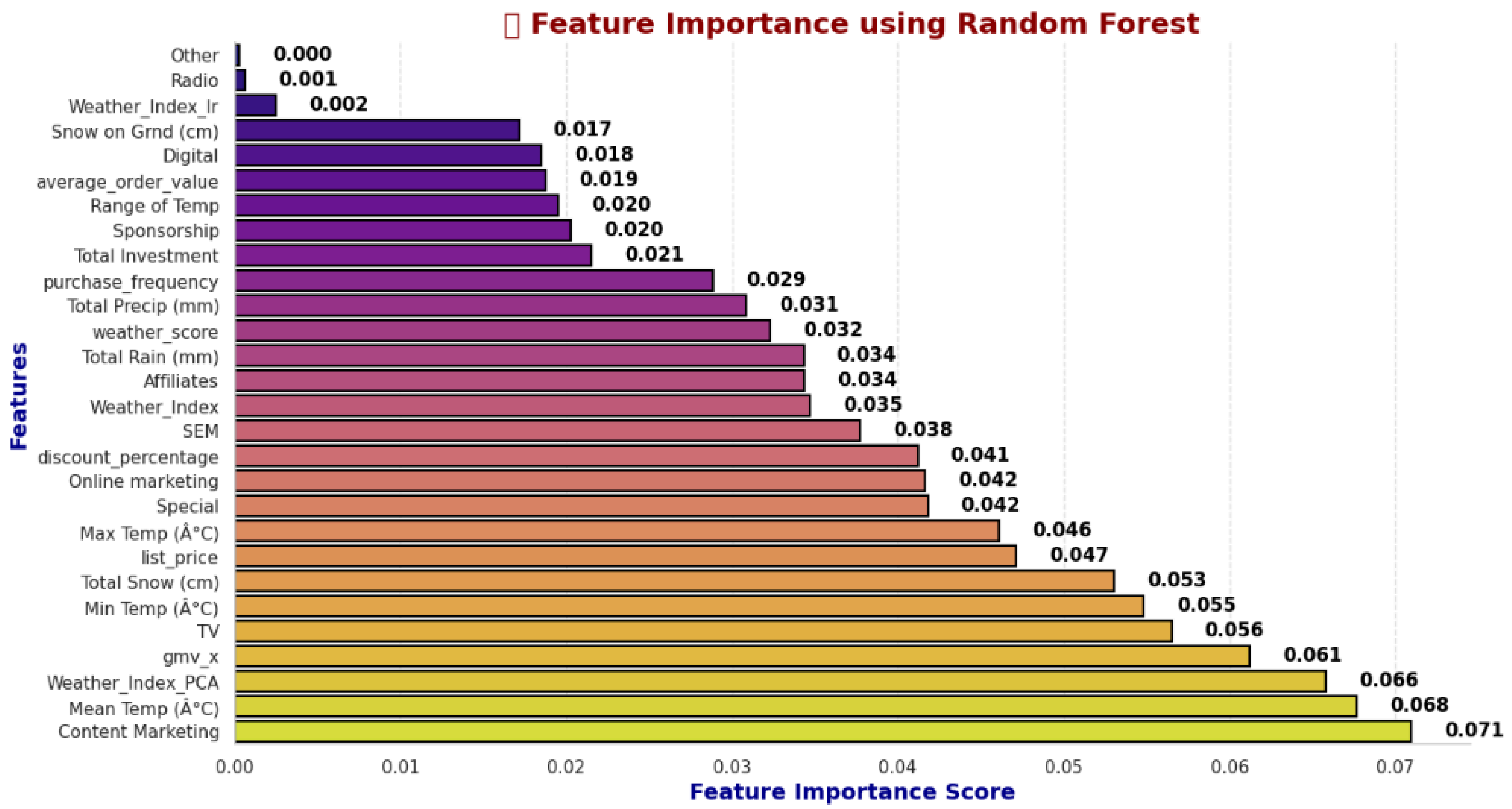
FEATURE ENGINEERING

Based on GMV



Units sold, special events, and total snowfall are the top drivers of GMV, indicating sales and weather significantly impact revenue.

Based on units

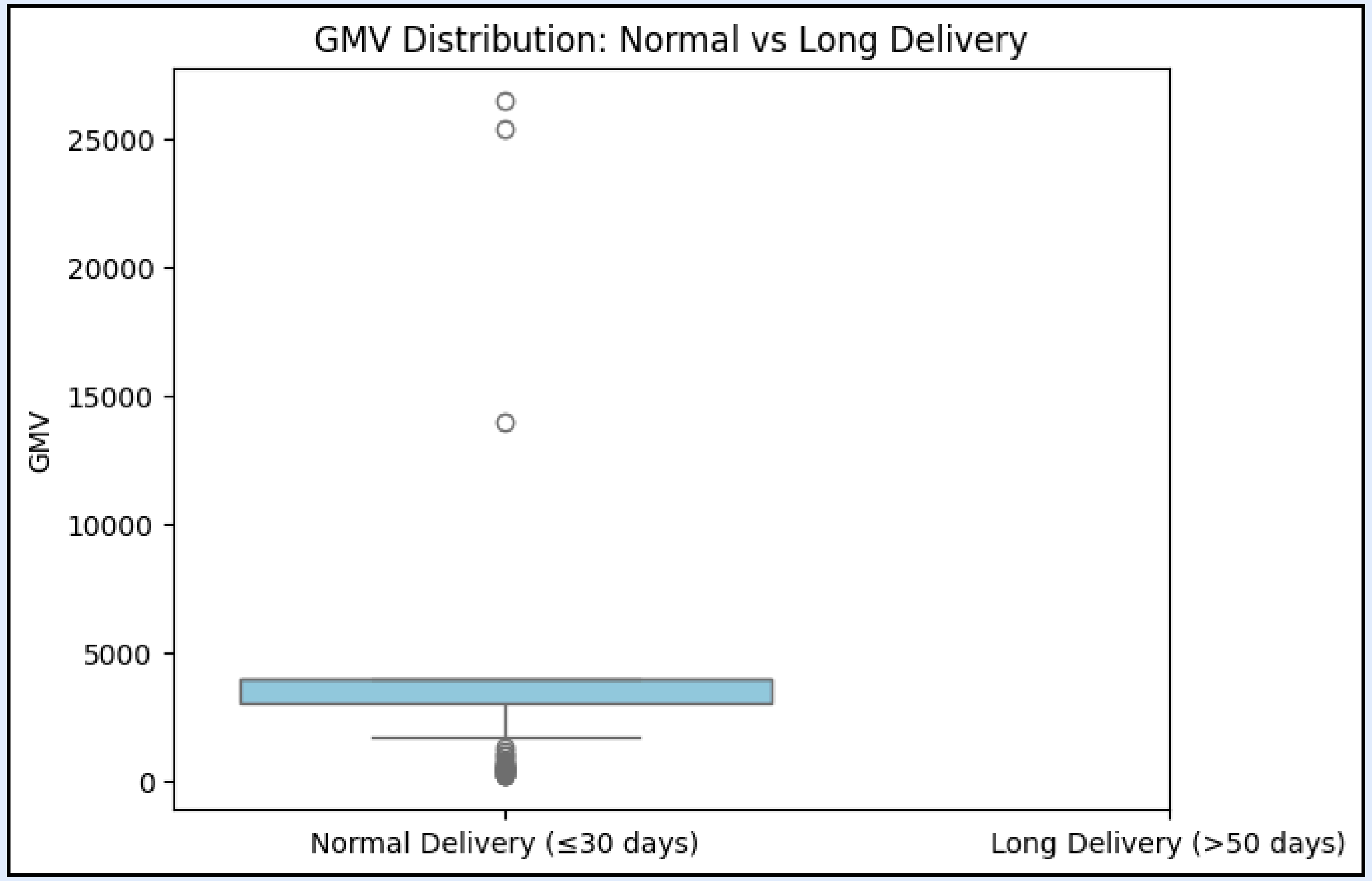


Special events, GMV, and discounts are key influencers of units sold, highlighting the importance of promotions and event-driven strategies.

HYPOTHESIS TESTING

Impact of Delivery Time on GMV

- Long delivery orders (>50 days) have higher GMV (₹3522.89 vs. ₹2469.22) but are rare (518/14L).
- Small effect size (d = 0.254) limits overall GMV impact despite significance.

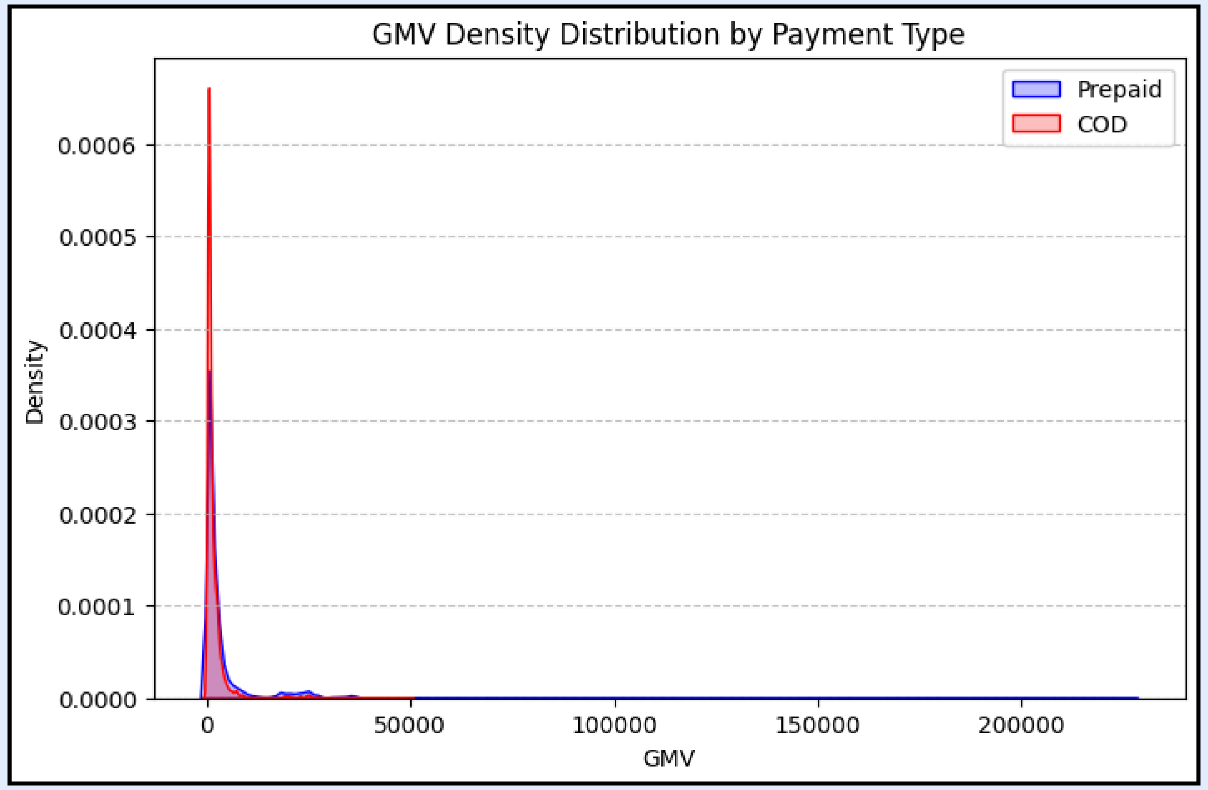


Extreme Weather Conditions Drive Higher GMV

- Weather does not impact order volume (p = 1.000) but significantly affects GMV (p < 0.001)
- Extreme weather (high/low temps) sees higher GMV, while moderate weather has the lowest.

Prepaid Orders Drive Higher GMV

- Prepaid orders generate significantly higher GMV than COD (p ≈ 0)
- Despite fewer prepaid orders, they have higher spending, suggesting incentives can boost revenue..

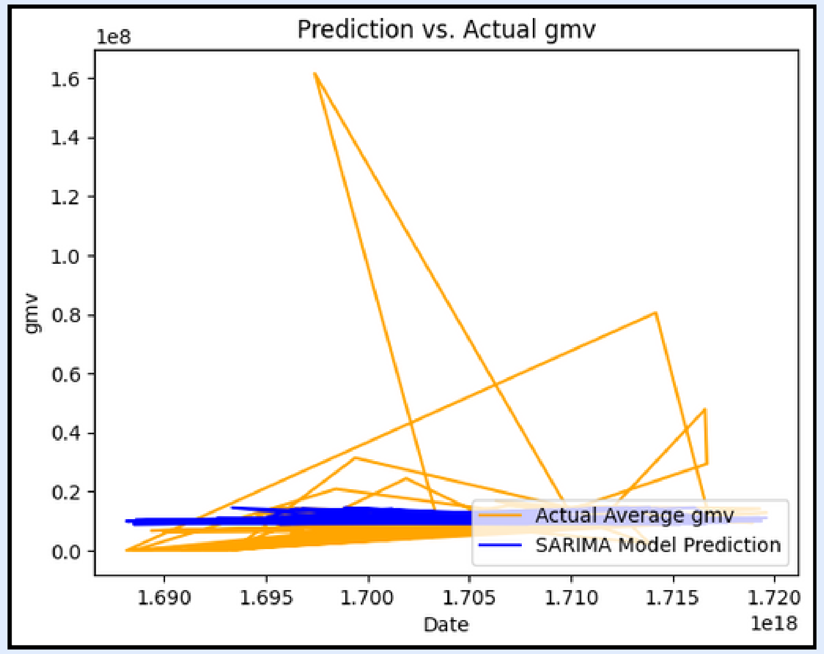


BUDGET OPTIMIZATION

TIME SERIES FORECASTING

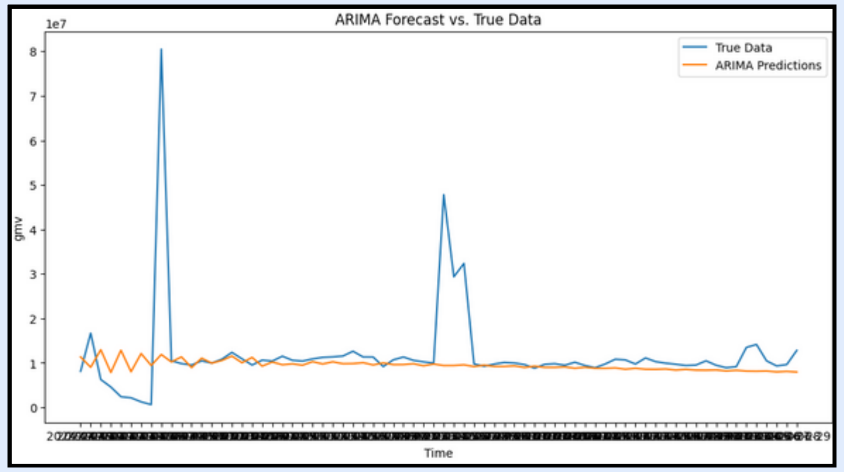
SARIMA

- Chosen to capture seasonal patterns in GMV.
- Optimal order (2,0,1)(0,1,1,12) determined from ACF, PACF & Dickey-Fuller tests
- Rolling forecast approach achieved MAPE of **10%**.



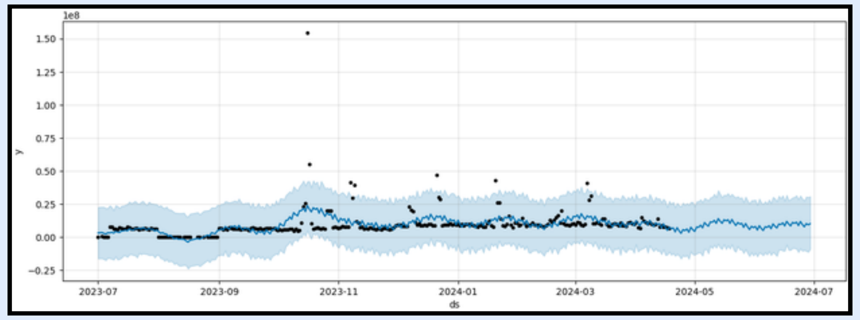
ARIMA

- Chosen for trend-based GMV prediction.
- Manual grid search (p=9, d=1, q=3) & Mango library optimization.
- Achieved MAPE of **16.09%** by hyperparameter tuning



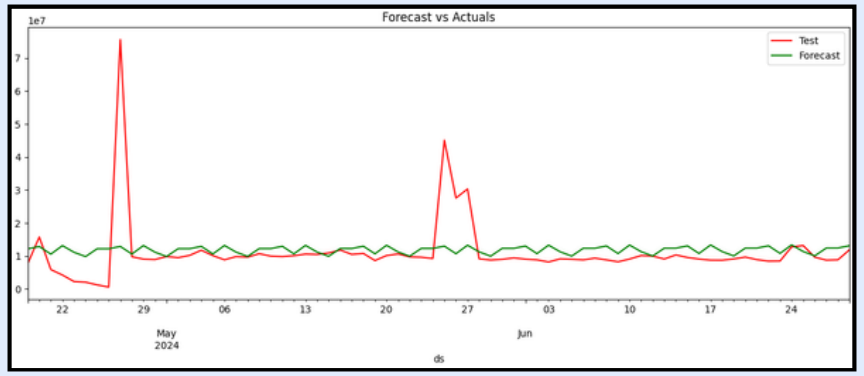
PROPHET

- Decomposes time series into trend & seasonality using date & GMV.
- Integrated Canadian holidays to improve accuracy.
- Achieved **30%** MAPE by hyperparameter tuning.



LSTM

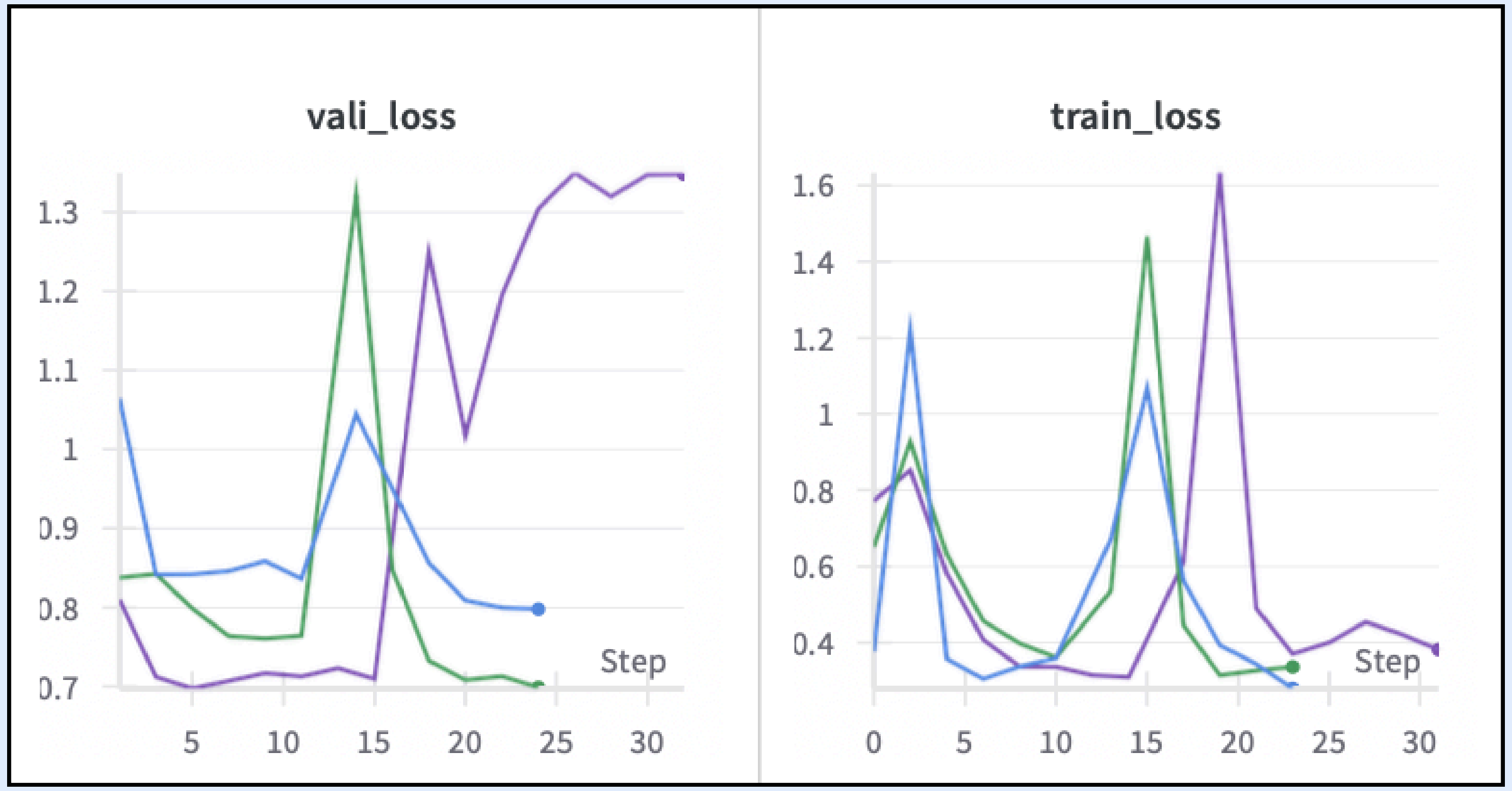
- Captured temporal dependencies for GMV prediction.
- Integrated weather, customer orders, and other factors.
- Hyperparameter tuning achieved **64.8%** MAPE.



TIME SERIES FORECASTING

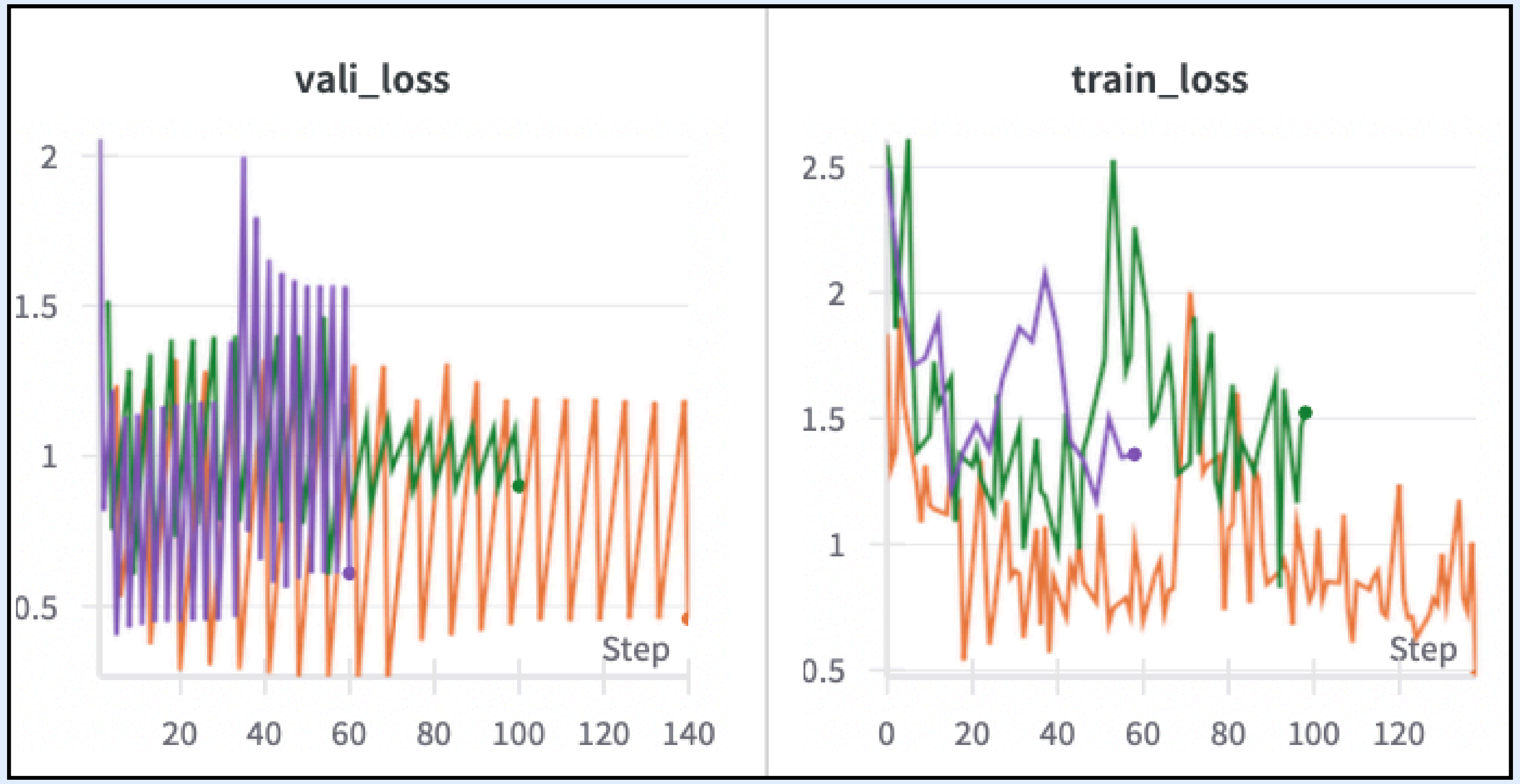
INFORMER

- Addresses the limitations of Transformer-based models in long-sequence forecasting.
- Trained for 6 epochs with early stopping.
- Performance gains over ARIMA & SARIMA by leveraging multivariate data & reducing computations.

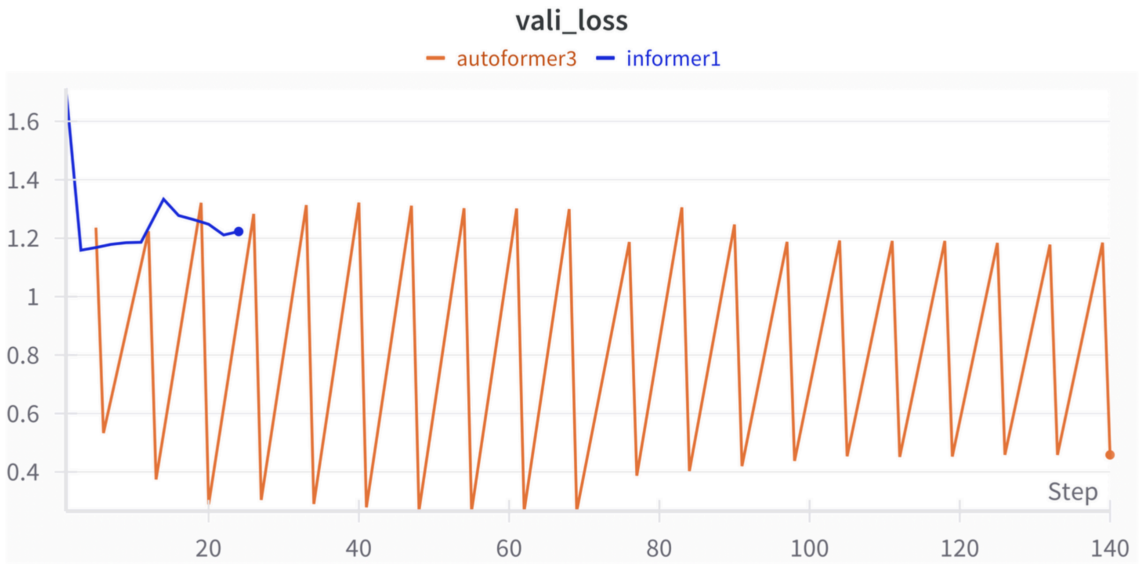


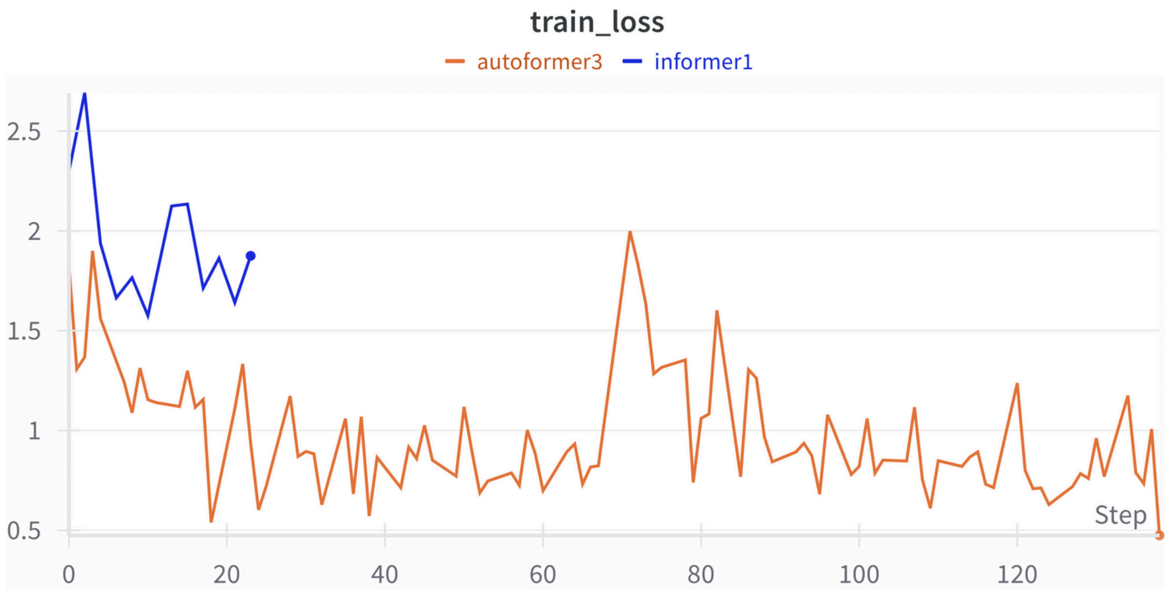
AUTOFORMER

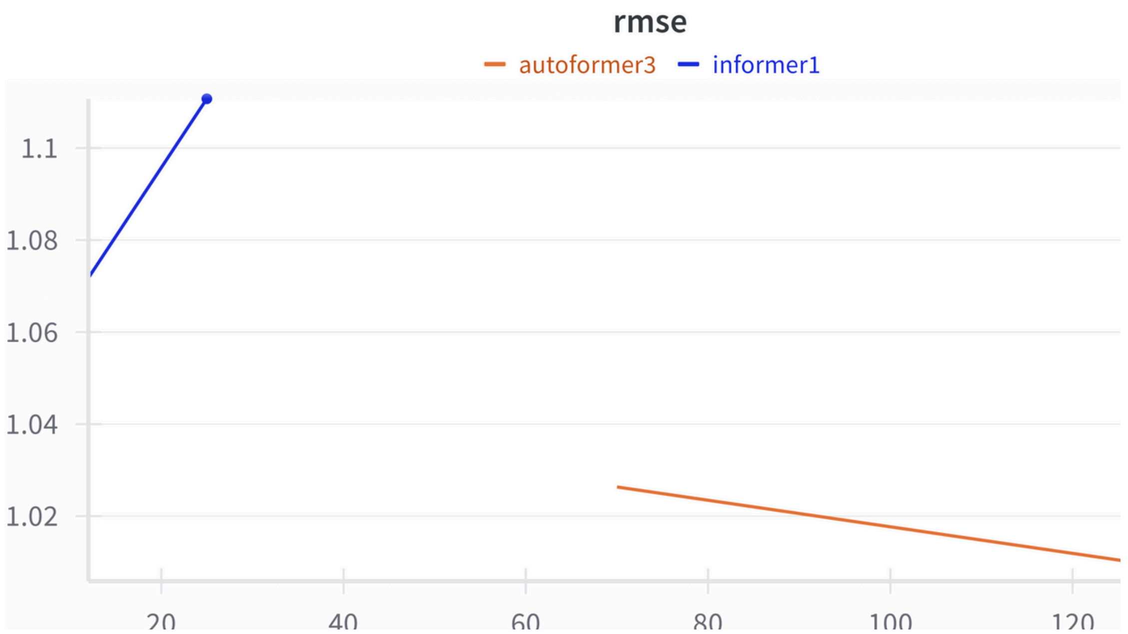
- Designed to decompose the time series into trend and seasonal components.
- Trained for 10 epochs with early stopping.
- Performance gains over Informer Model as it uses Auto-Correlation attention instead of dot-product attention.

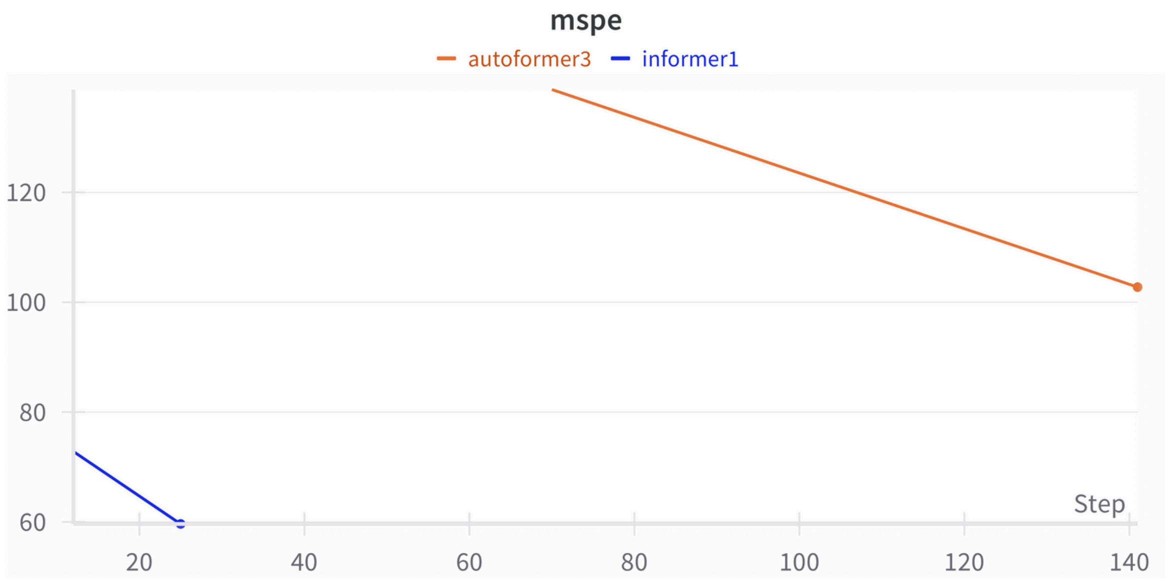


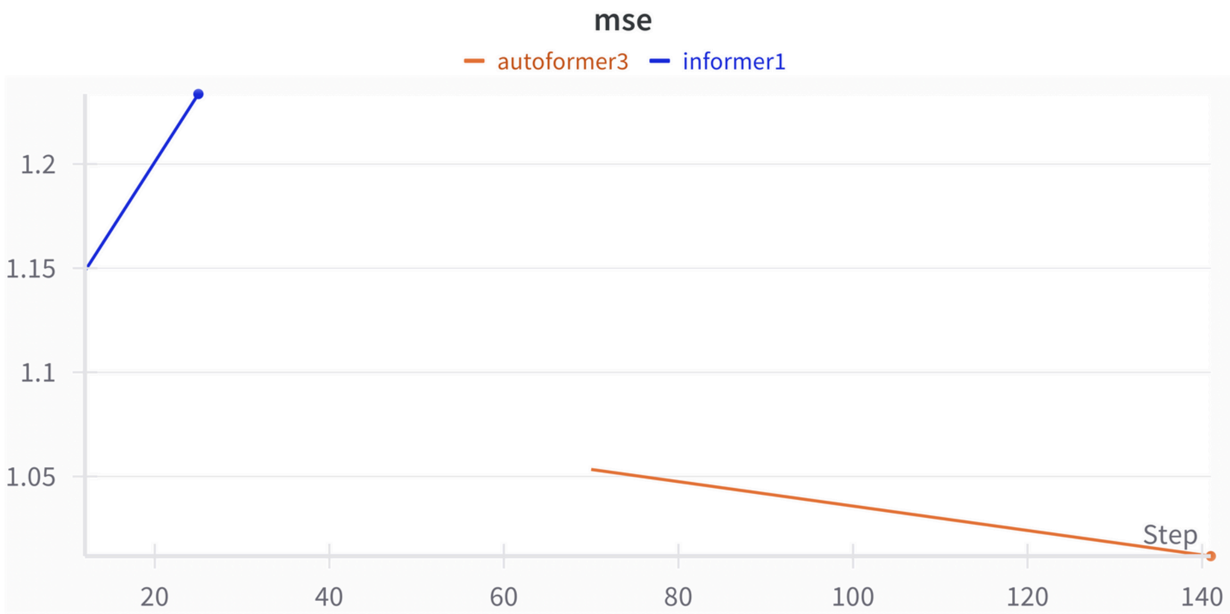
AUTOFORMER v/s INFORMER

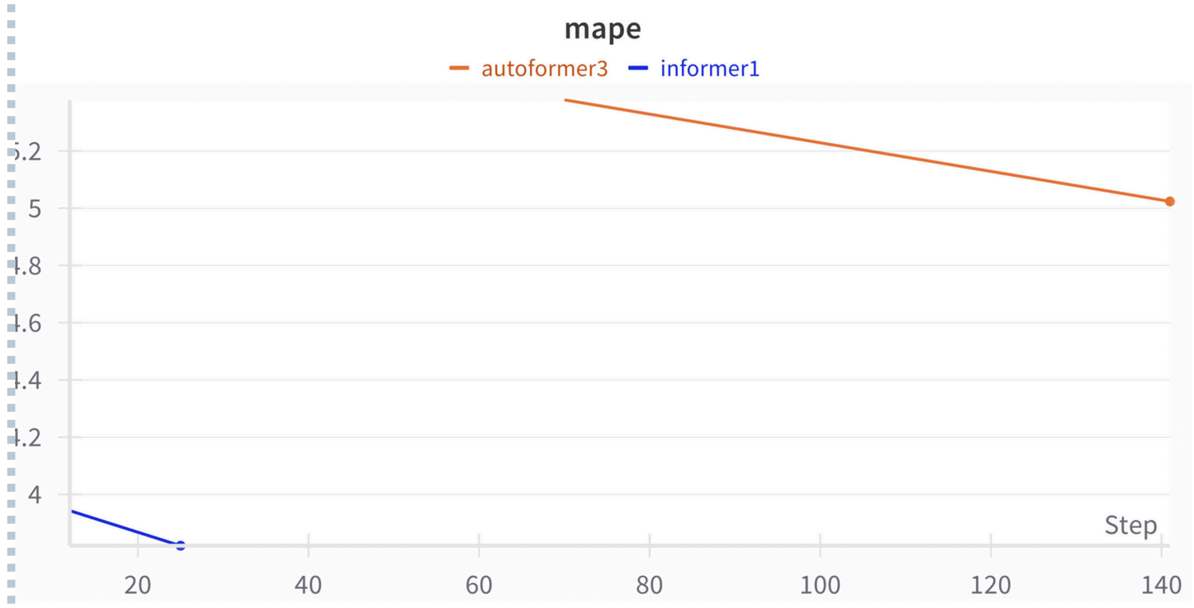




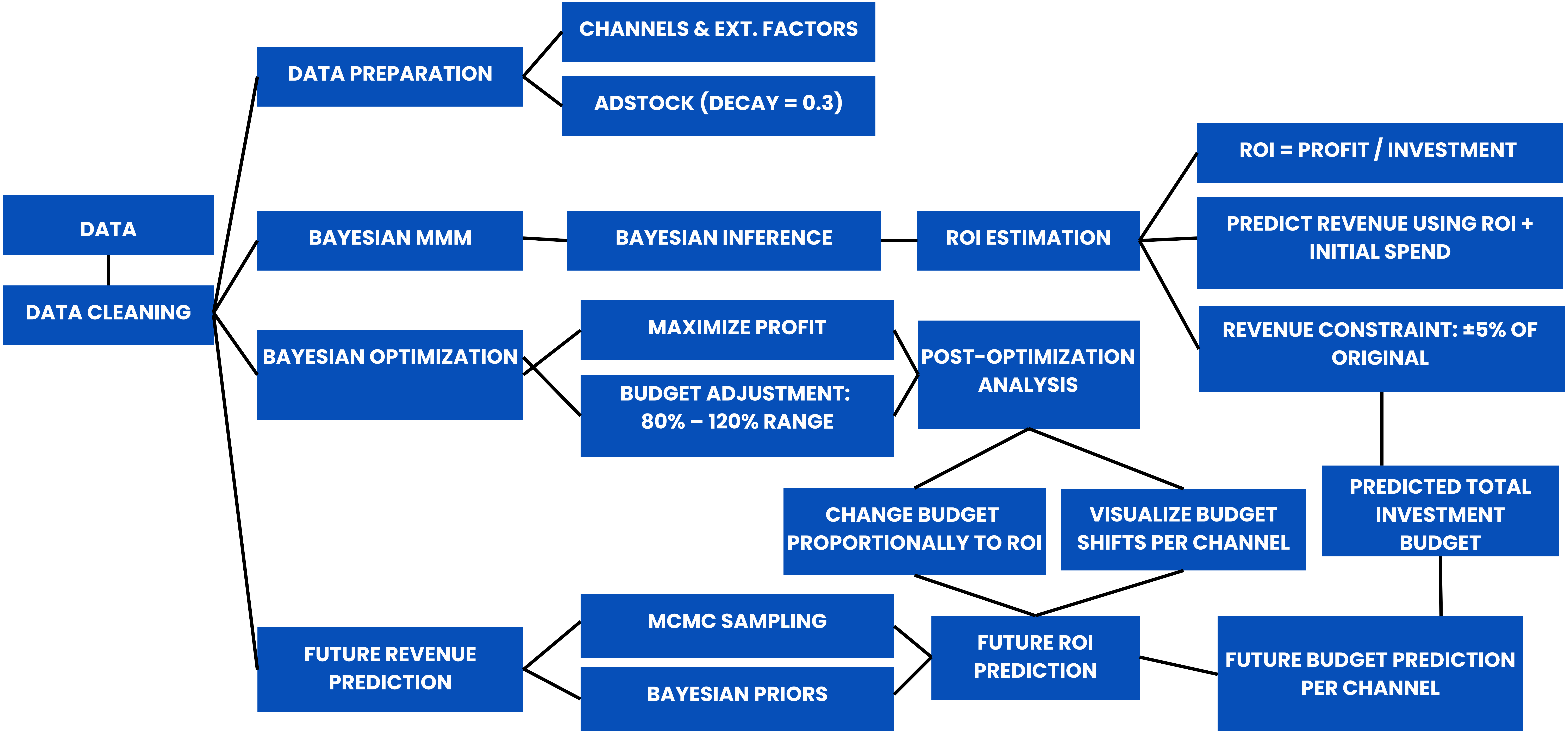








SOLUTION 1: BUDGET REDUCTION



BUDGET REDUCTION STRATEGY

1 Bayesian ROI Estimation

- Defined marketing channels & external factors.
- Applied adstock (decay = 0.3) for carryover effects.
- Set Bayesian priors for base revenue, channels, and external effects.
- Modeled revenue using PyMC; sampled with NUTS (3,000 samples, 4 chains).

2 Bayesian ROI Estimation

- Estimated channel-wise ROI via Bayesian inference.
- ROI = Profit / Investment per channel
- Predicted original revenue from estimated ROIs.

3 Optimization Approach

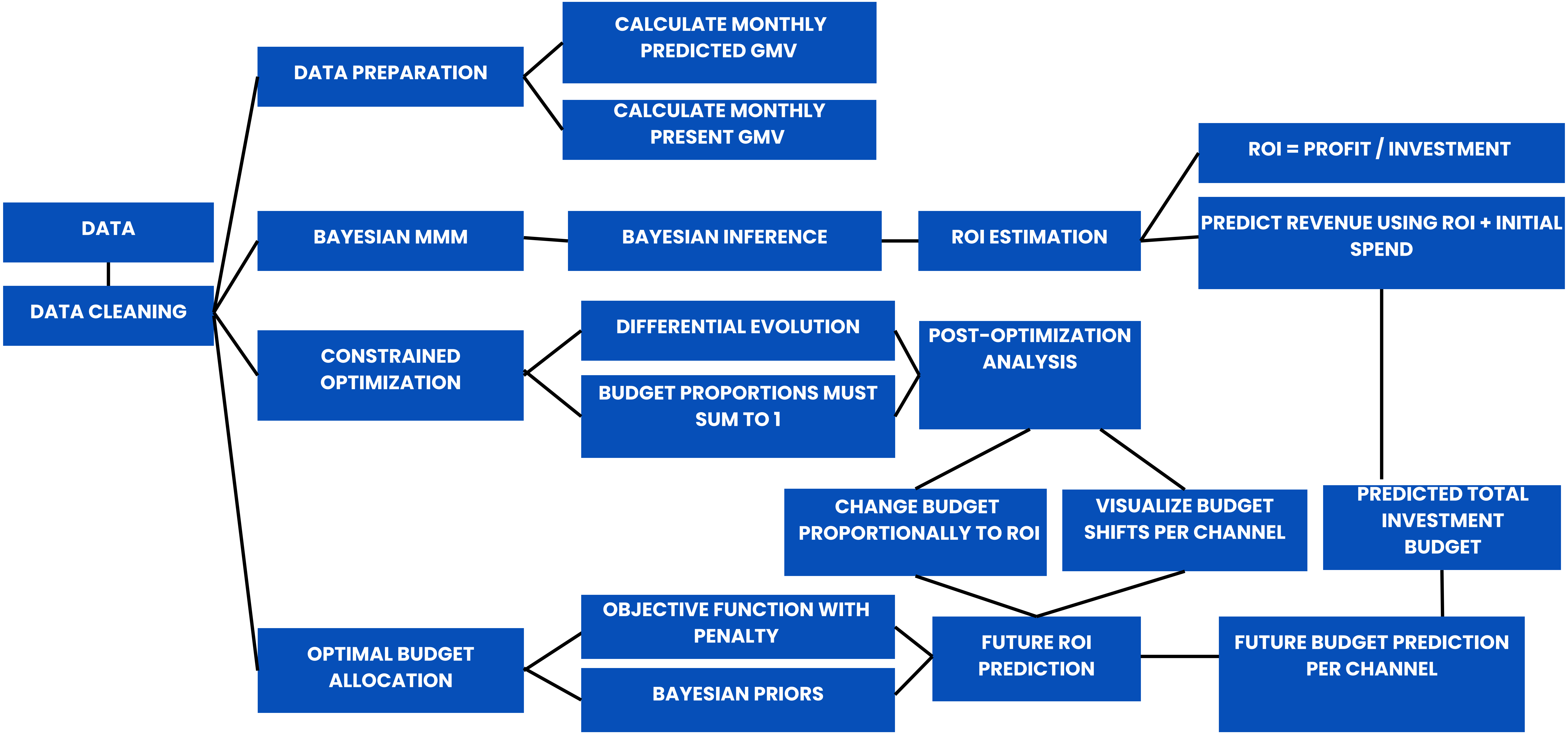
- Objective: Maximize profit under ±5% revenue constraint.
- Penalty applied for constraint violation
- Optimized spend using Bayesian optimization.
- Spend Bounds: 80%–120% of original per channel.

4 Post-Optimization Insights

- Compared original vs. optimized spend, revenue, and profit.
- Visualized budget shifts across channels.
- Reallocated budget proportionally based on ROI.

Channel	Prev ROI (%)	Optimized ROI (%)	ROI Improvement (%)	Previous Investment (in Cr.)	Predicted Investment (in Cr.)	Profit in Investment (in Cr.)
Online Marketing	-84.41	-99.45	-15.04	61.36	15.65	+45.71
SEM	-90.30	-75.59	14.71	4.67	4.47	+0.20
Affiliates	-93.35	-72.81	20.53	91.20	13.51	+77.69
Content Marketing	-99.06	-73.25	25.81	193.64	39.35	+154.29
Sponsorship	-78.57	-81.80	-3.23	8.03	7.99	+0.04
Other	-93.67	-86.53	7.14	353.59	105.44	+248.15
TV	-93.52	-65.89	27.62	29.66	3.22	+26.44
Radio	-99.35	-87.48	11.87	48.02	6.76	+41.26
Digital	-95.64	--91.39	4.25	365.37	183.45	+181.92

SOLUTION 2: BUDGET REALLOCATION



BUDGET REALLOCATION STRATEGY

1 Bayesian ROI Estimation

- Calculated predicted revenue using time series model
- Calculated Bayesian Priors from model
- Modeled revenue using PyMC; sampled with NUTS (3,000 samples, 4 chains).

2 Bayesian ROI Estimation

- Estimated channel-wise ROI via Bayesian inference.
- ROI = Profit / Investment per channel
- Predicted original revenue from estimated ROIs.

3 Optimization Approach

- Differential Evolution: A constrained optimization problem by keeping budget proportions sum as 1
- Objective function: Balances revenue maximization with a penalty for violating budget constraints.

4 Post-Optimization Insights

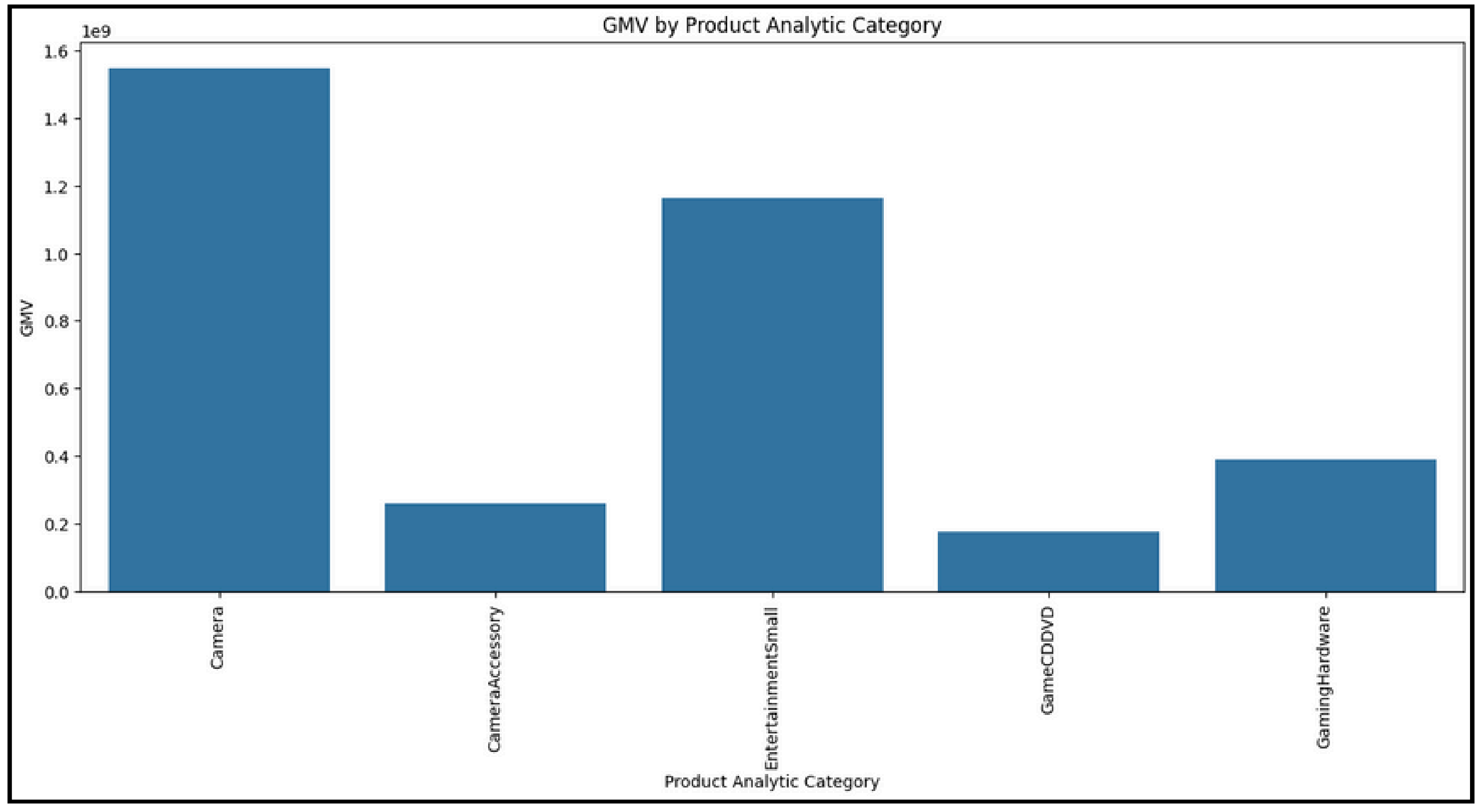
- Compared original vs. optimized spend, revenue, and profit.
- Visualized budget shifts across channels.
- Reallocated budget proportionally based on predicted coefficients.

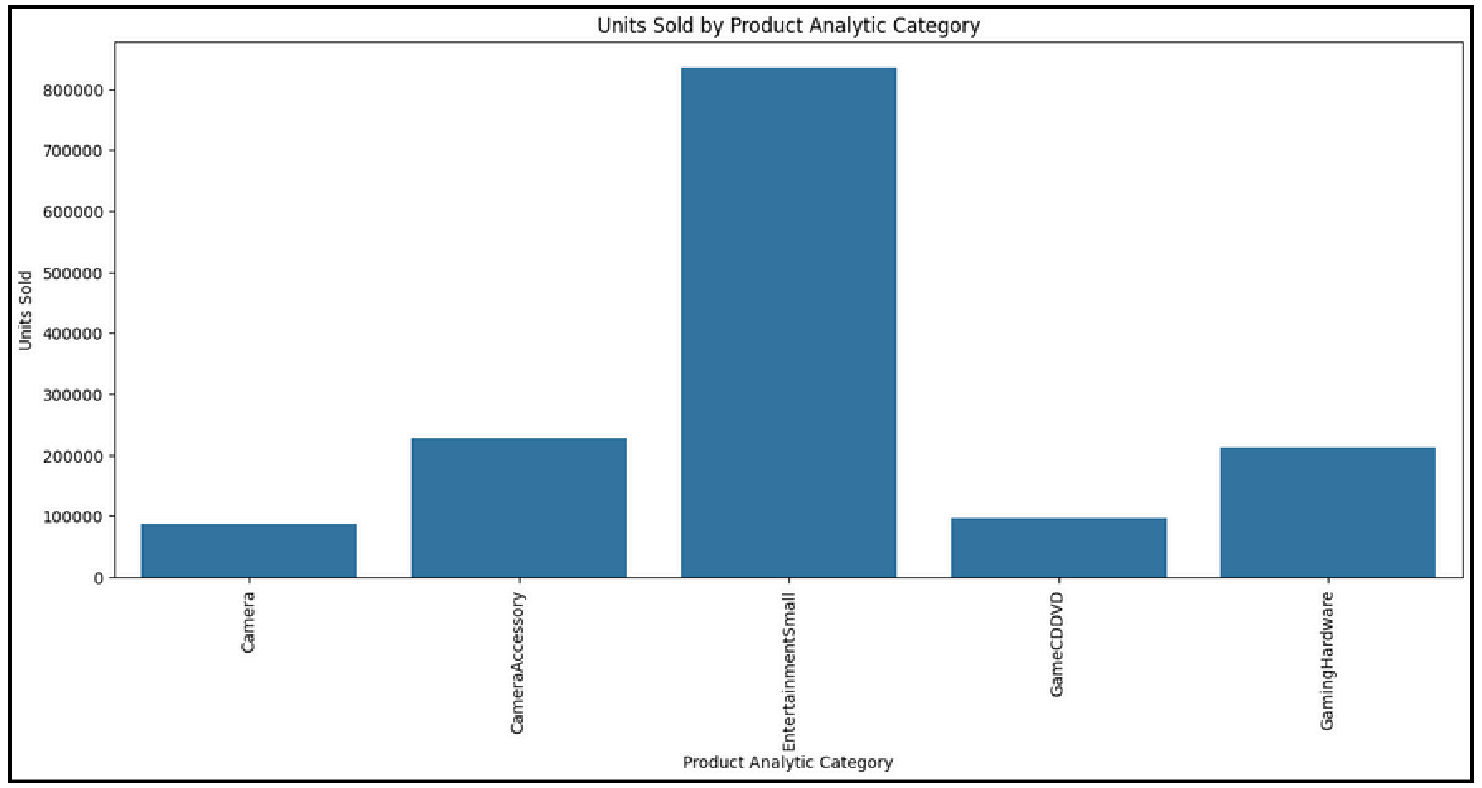
Channel	Prev ROI (%)	Optimized ROI (%)	ROI Improvement (%)	Predicted Investment (in Crores)	Previous Investment (in Crores)	Increase in revenue (in Crores)
TV	-84.41	-82.72	1.69	17.03	61.36	0.53
Content Marketing	-90.30	-82.17	8.13	77.52	4.67	2.40
Affiliates	-93.35	-88.66	4.69	34.88	91.20	1.08
SEM	-99.06	-96.68	2.38	20.17	193.64	0.63
Radio	-78.57	-76.29	2.28	23.72	8.03	0.73
Other	-93.67	-59.48	34.19	309.41	353.59	9.60
Digital	-93.52	-59.85	33.67	303.98	29.66	9.43
Sponsorship	-99.35	-97.56	1.79	14.68	48.02	0.46
Online Marketing	-95.64	-90.57	5.07	45.12	365.37	1.40

PRODUCT CATEGORIES TO TARGET FOR CAMPAIGN

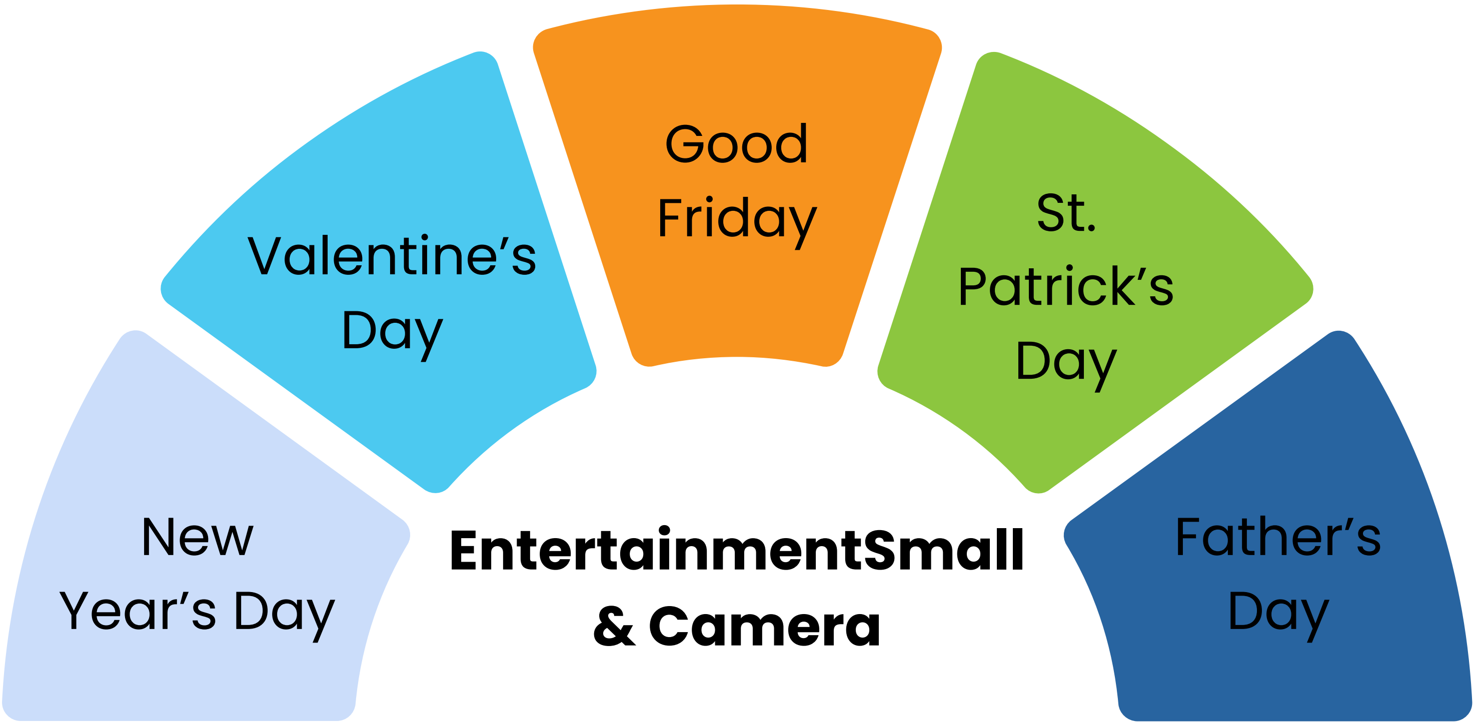
The product category which has both **higher revenue** and **higher units sold** should be prioritized for campaign. In our case it is **EntertainmentSmall Product Category**.

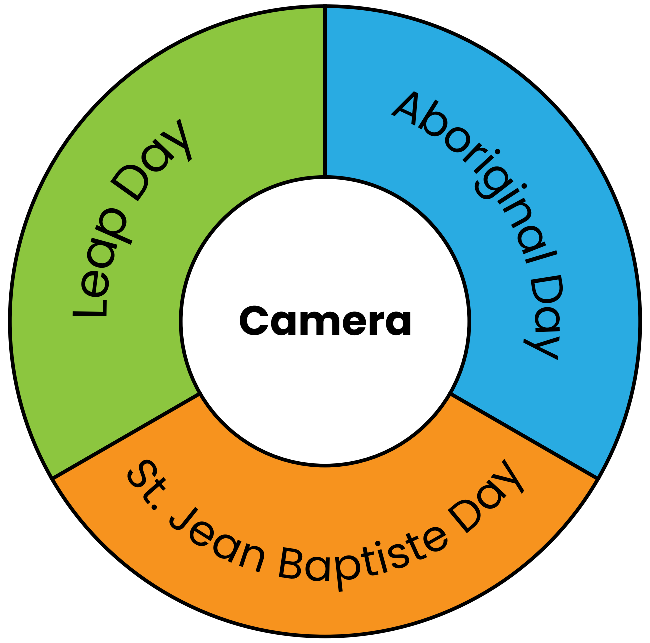
Product Analytic Category	gmv	units	revenue percentage	units percentage
Camera	12544.93	86695	43.783577	5.938514
CameraAccessory	7720.62	228514	7.324695	15.652963
EntertainmentSmall	9428.10	835530	32.905401	57.232904
GameCDDVD	5257.75	95824	4.988128	6.563841
GamingHardware	11592.69	213314	10.998200	14.611779

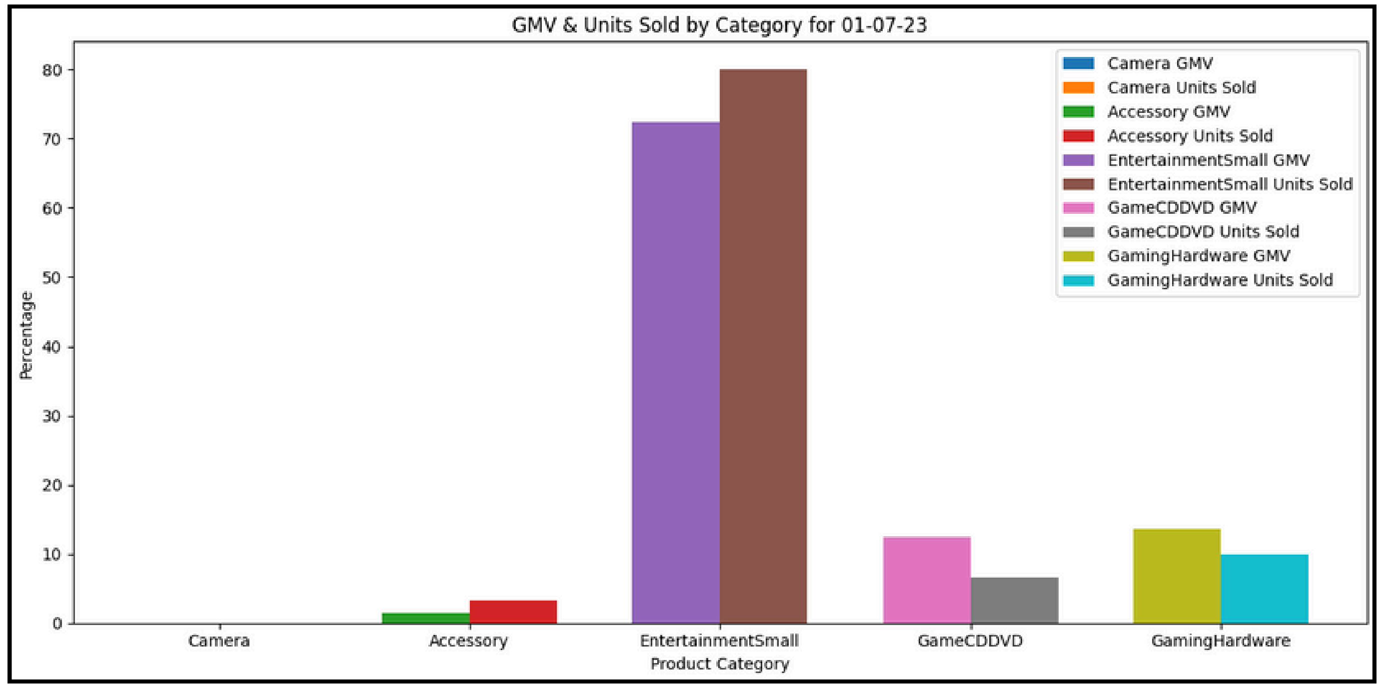


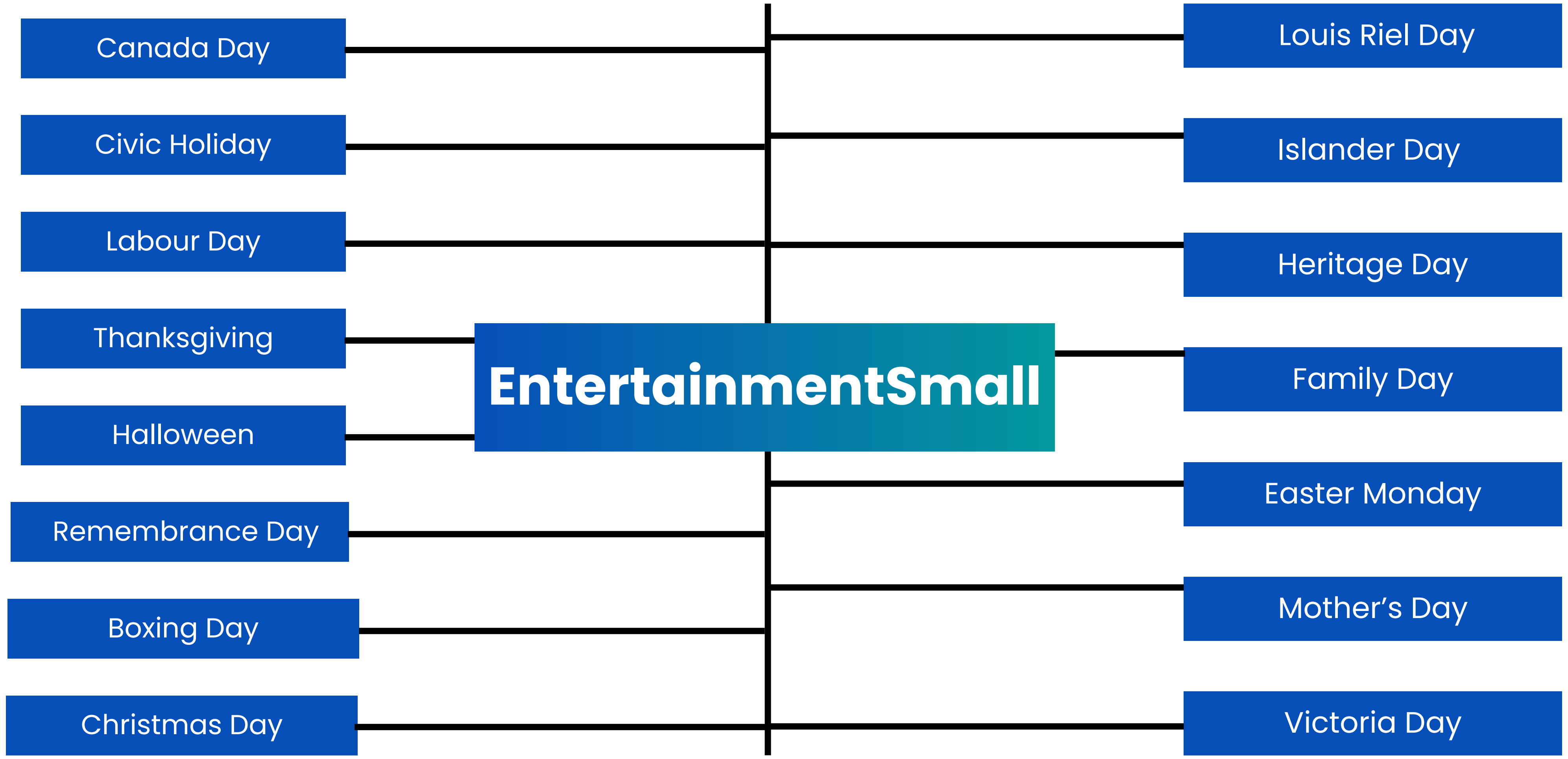


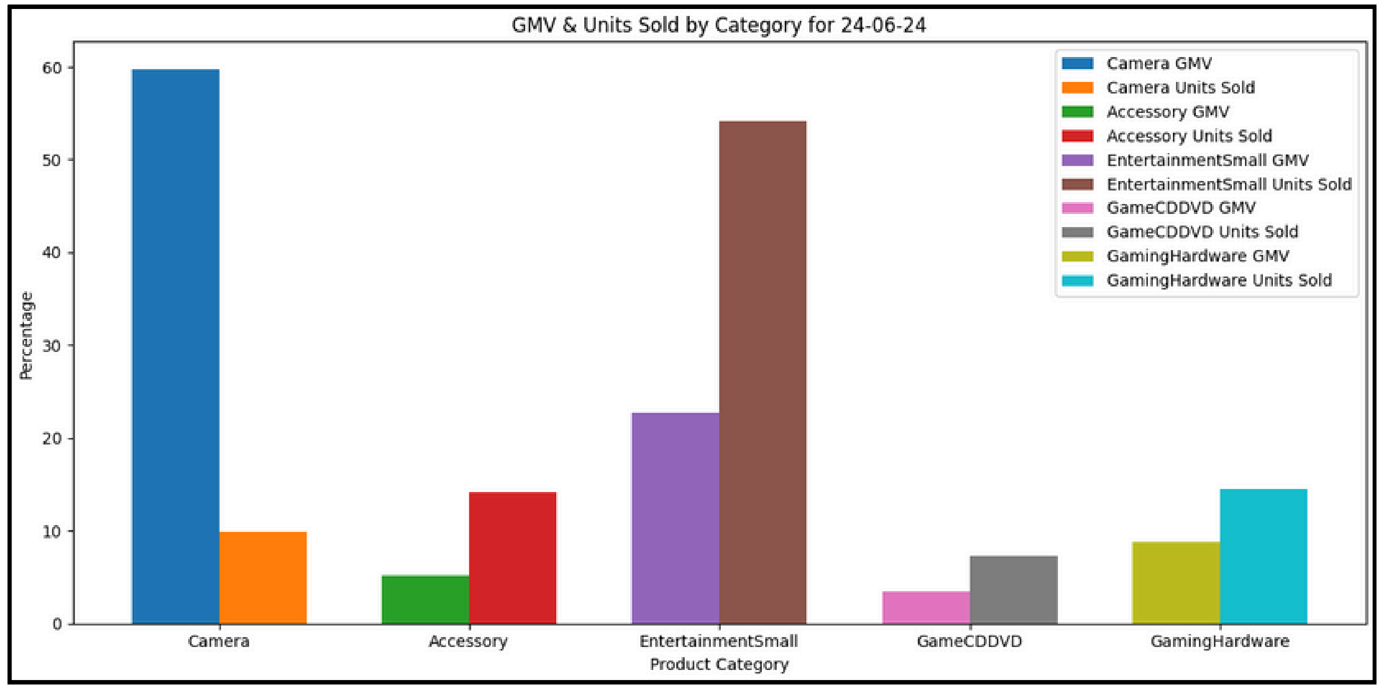
TOP-SELLING PRODUCT CATEGORIES DURING HOLIDAYS

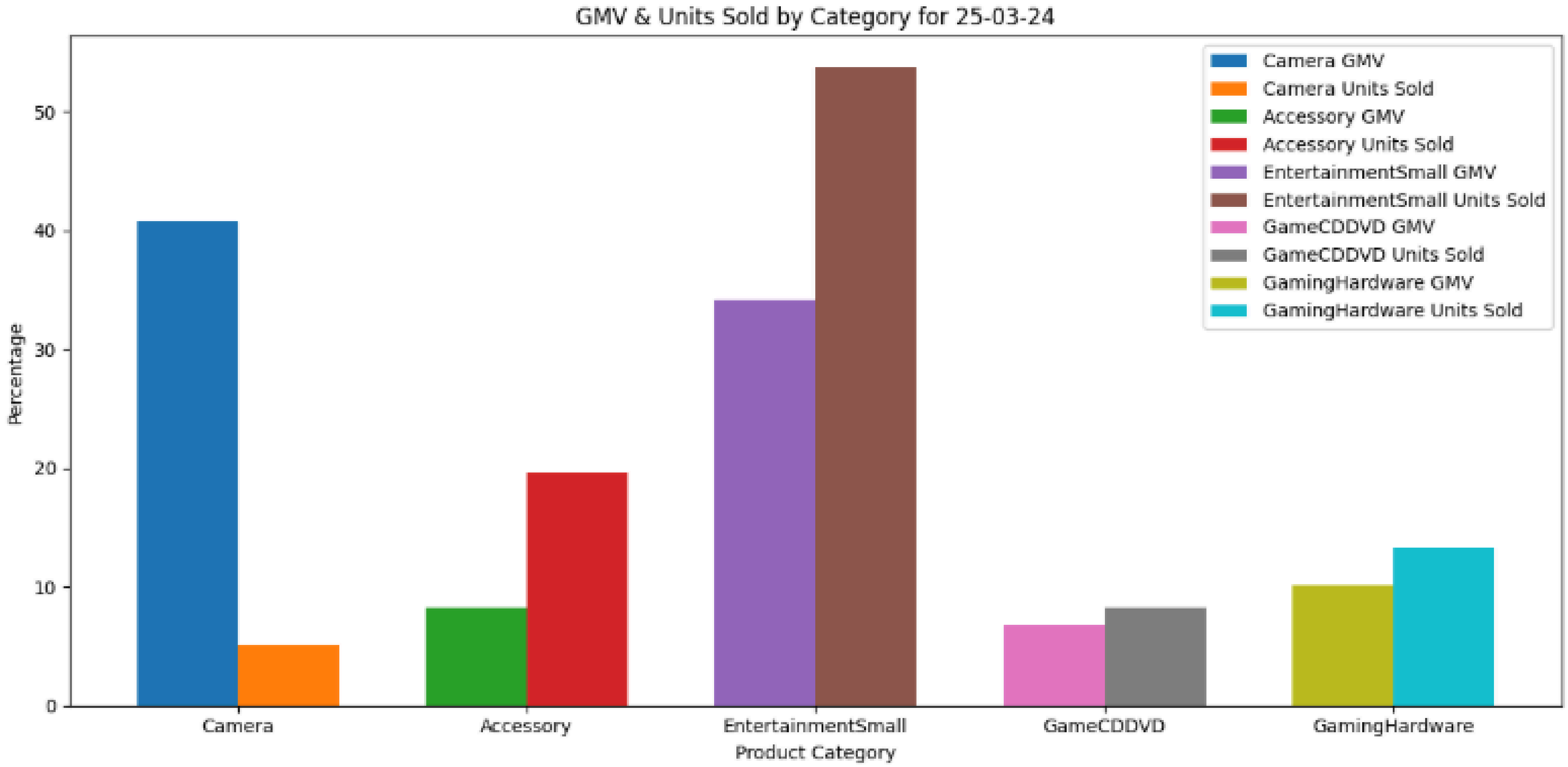










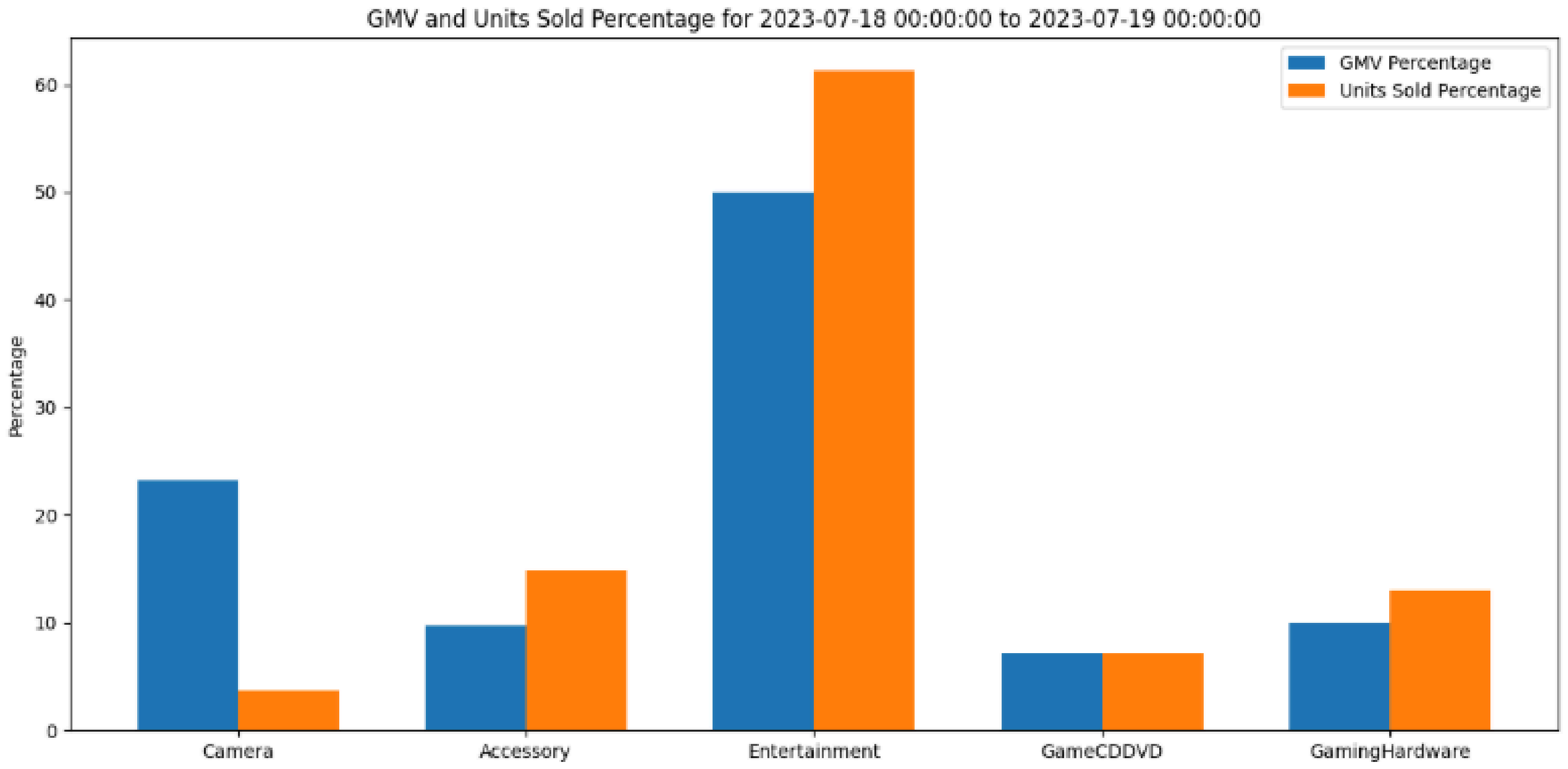


LOGIC OF CAMPAIGN SELECTION

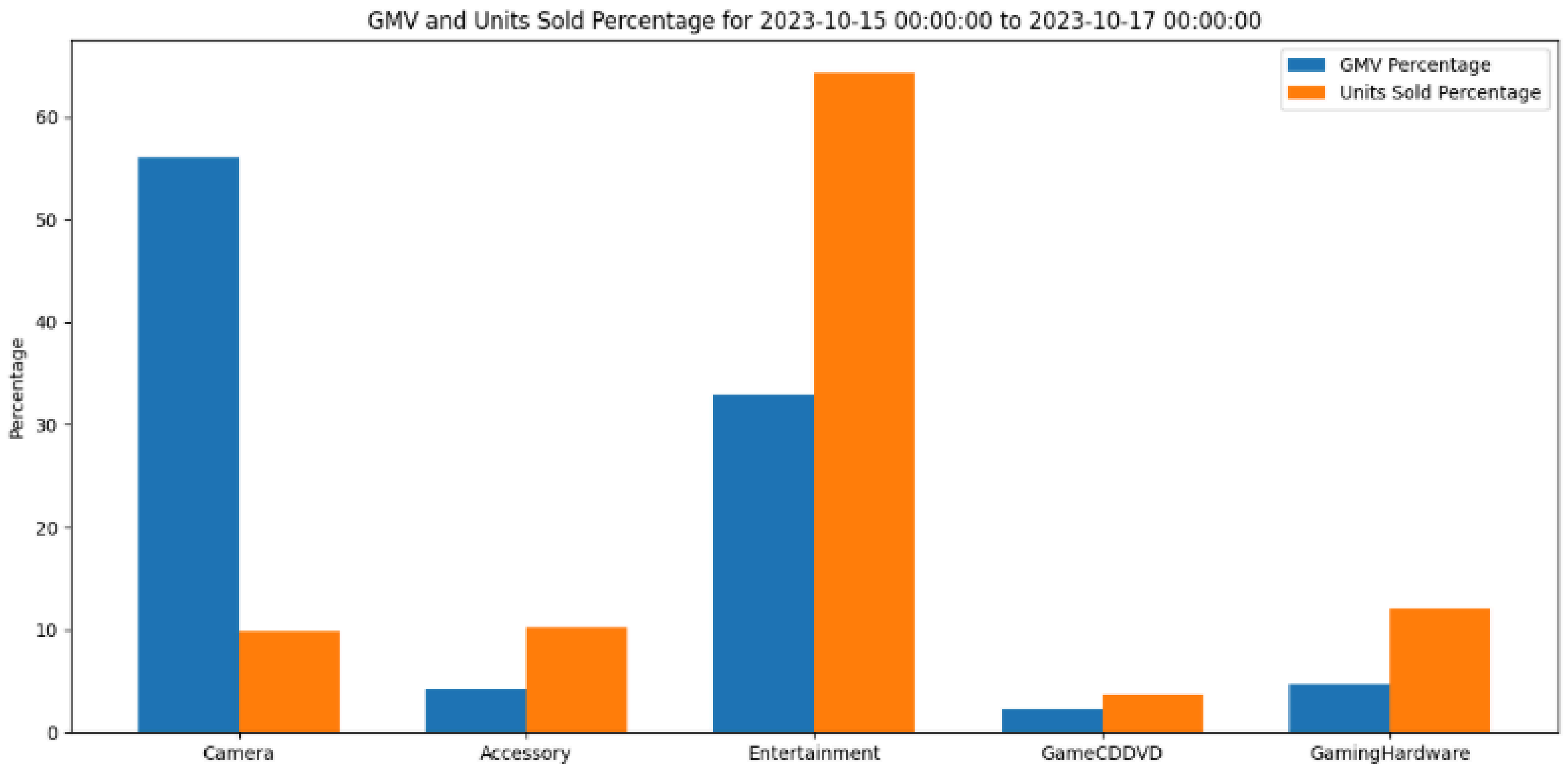
- Wherever EntertainmentSmall - Revenue and Units Sold both are high
- Wherever EntertainmentSmall & Camera - Revenue of Camera is high so we can't neglect it.
- Wherever Camera - Revenue of EntertainmentSmall isn't comparable to Camera although EntertainmentSmall has higher Units Sold.

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

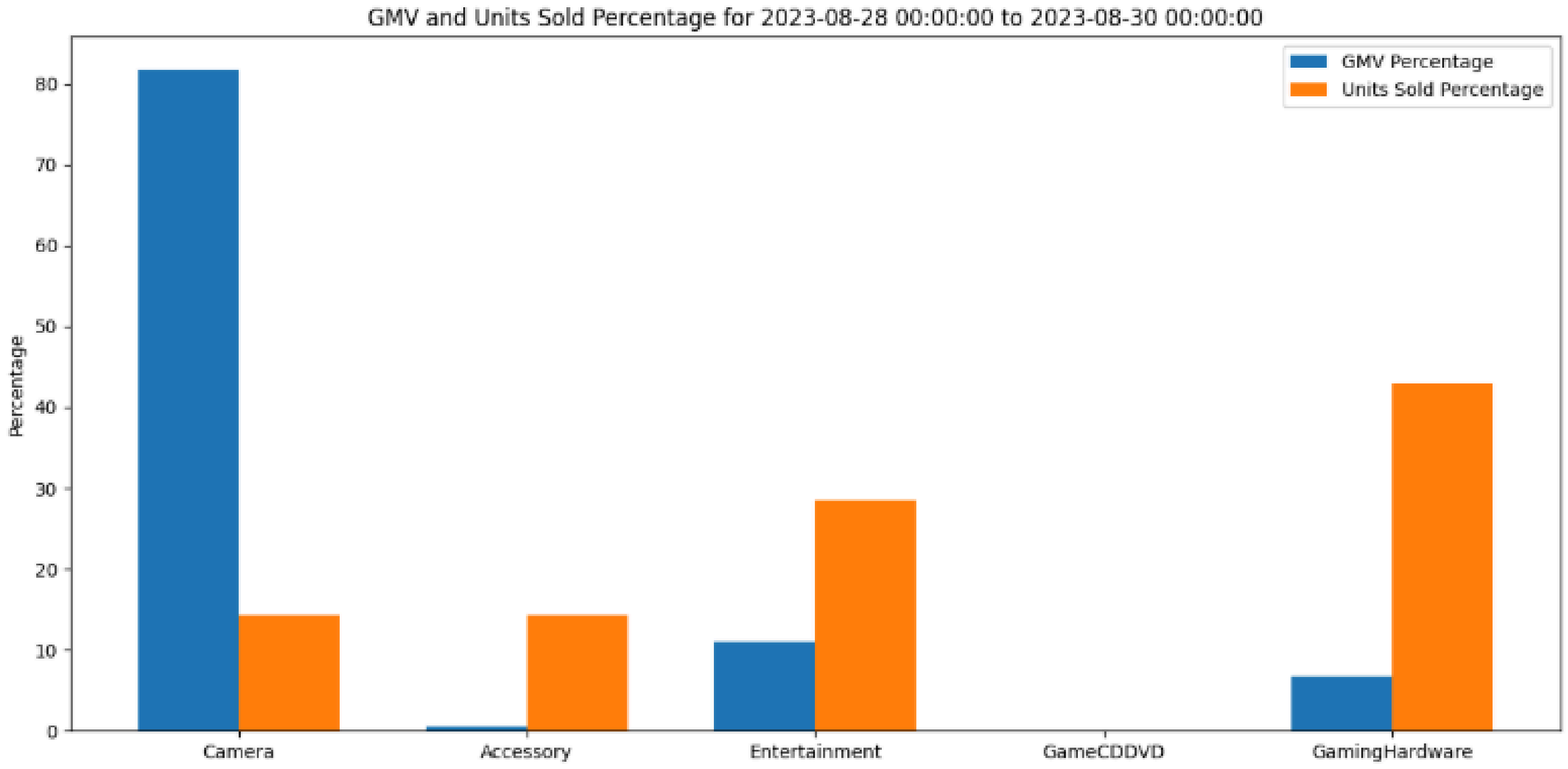
CAMPAIGN TO BE DONE IN EACH DURATION



Campaign Should be done for Entertainment category

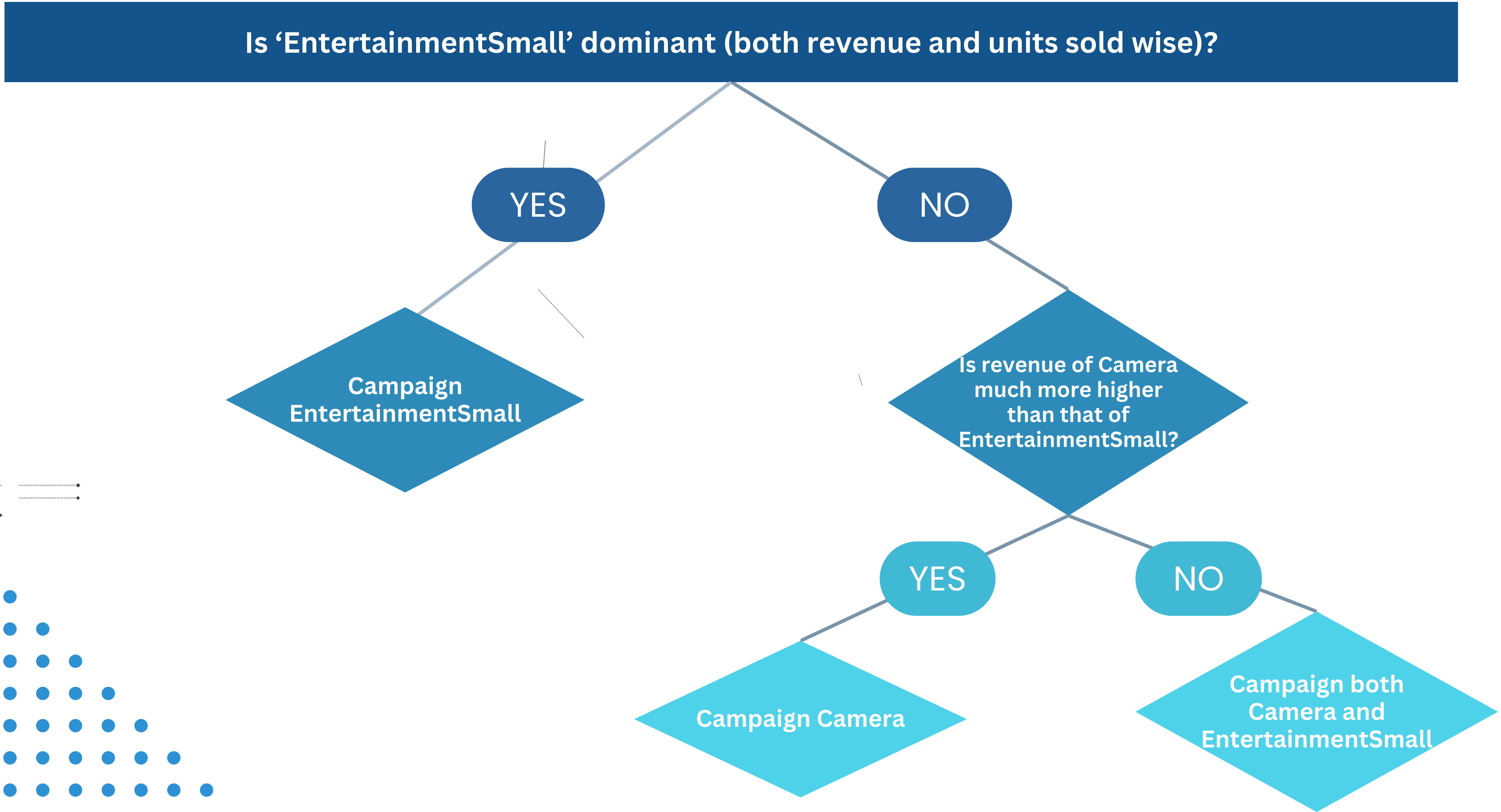


Campaign Should be done for Entertainment & Camera category



Campaign Should be done for Camera category

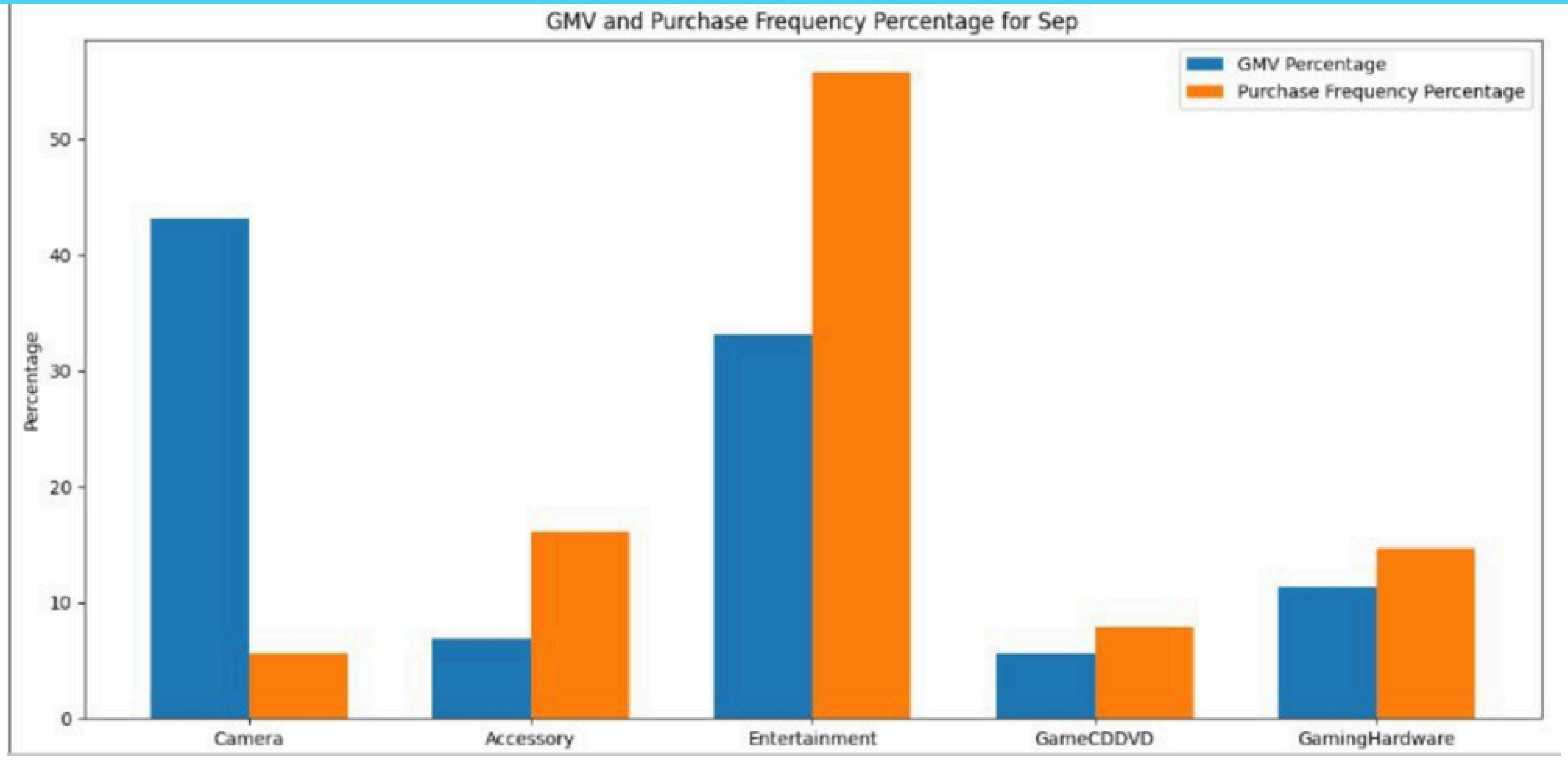
PRODUCT/ PRODUCT CATEGORIES TO TARGET FOR UPCOMING CAMPAIGN

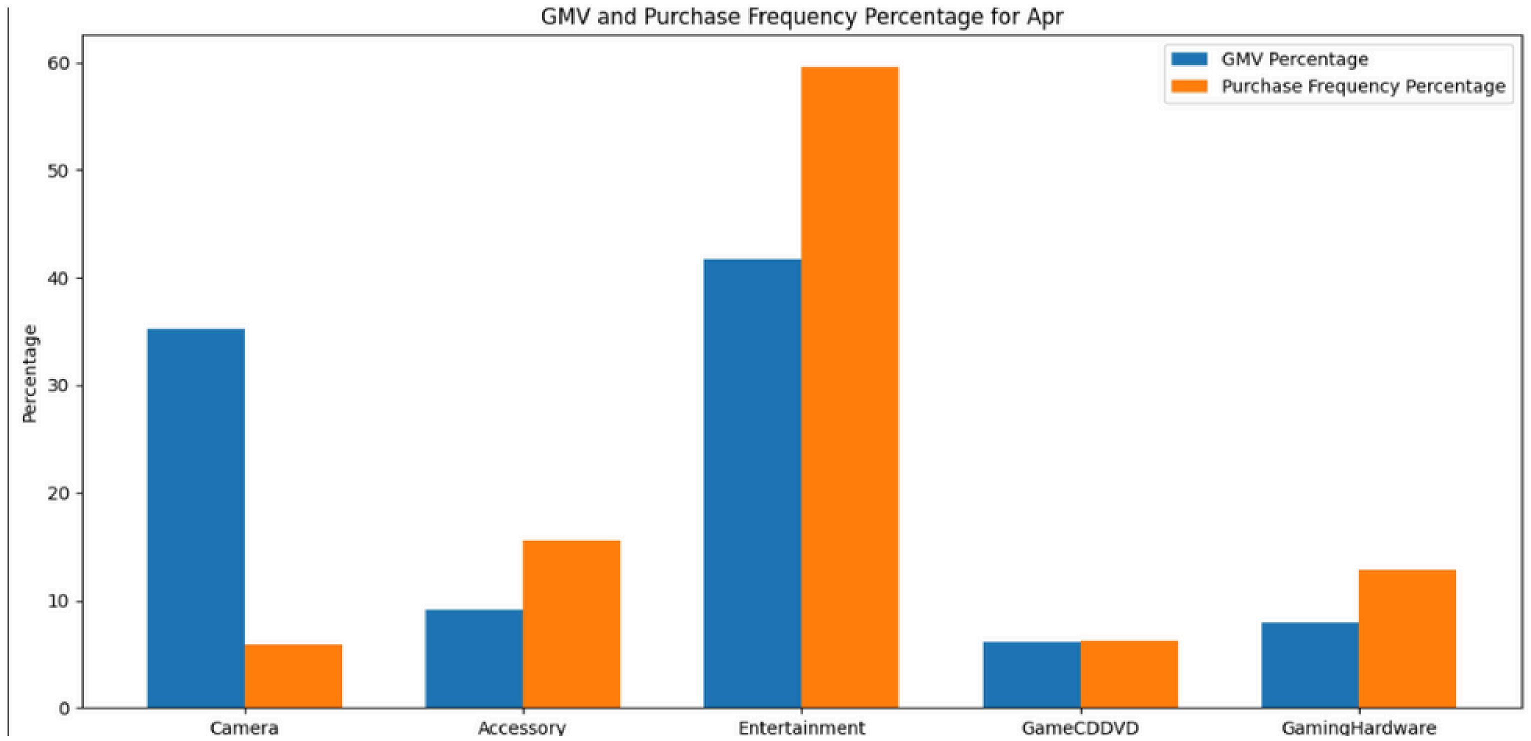


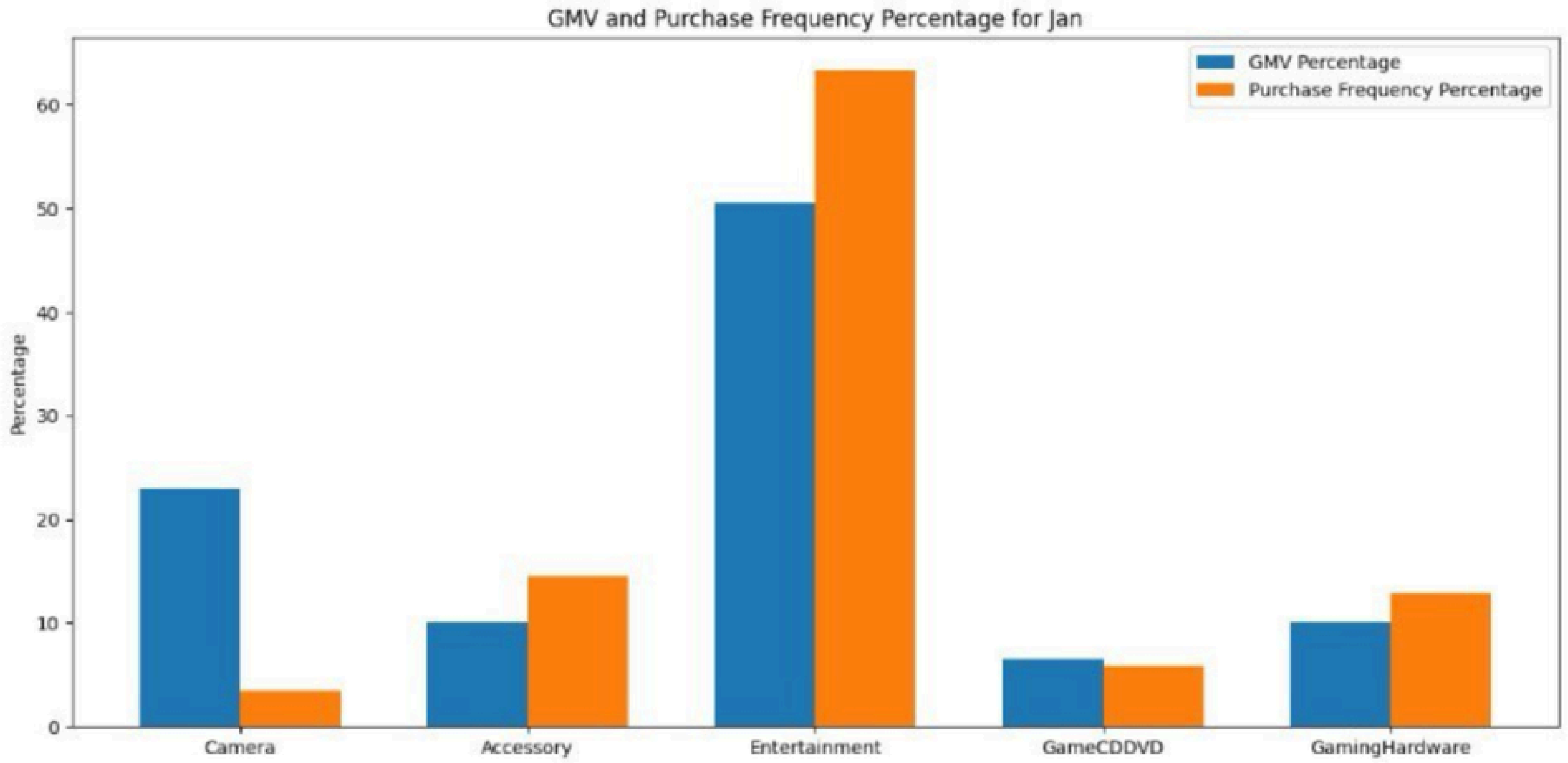
OPTIMAL MARKETING CHANNEL ALLOCATION FOR EACH PRODUCT CATEGORY

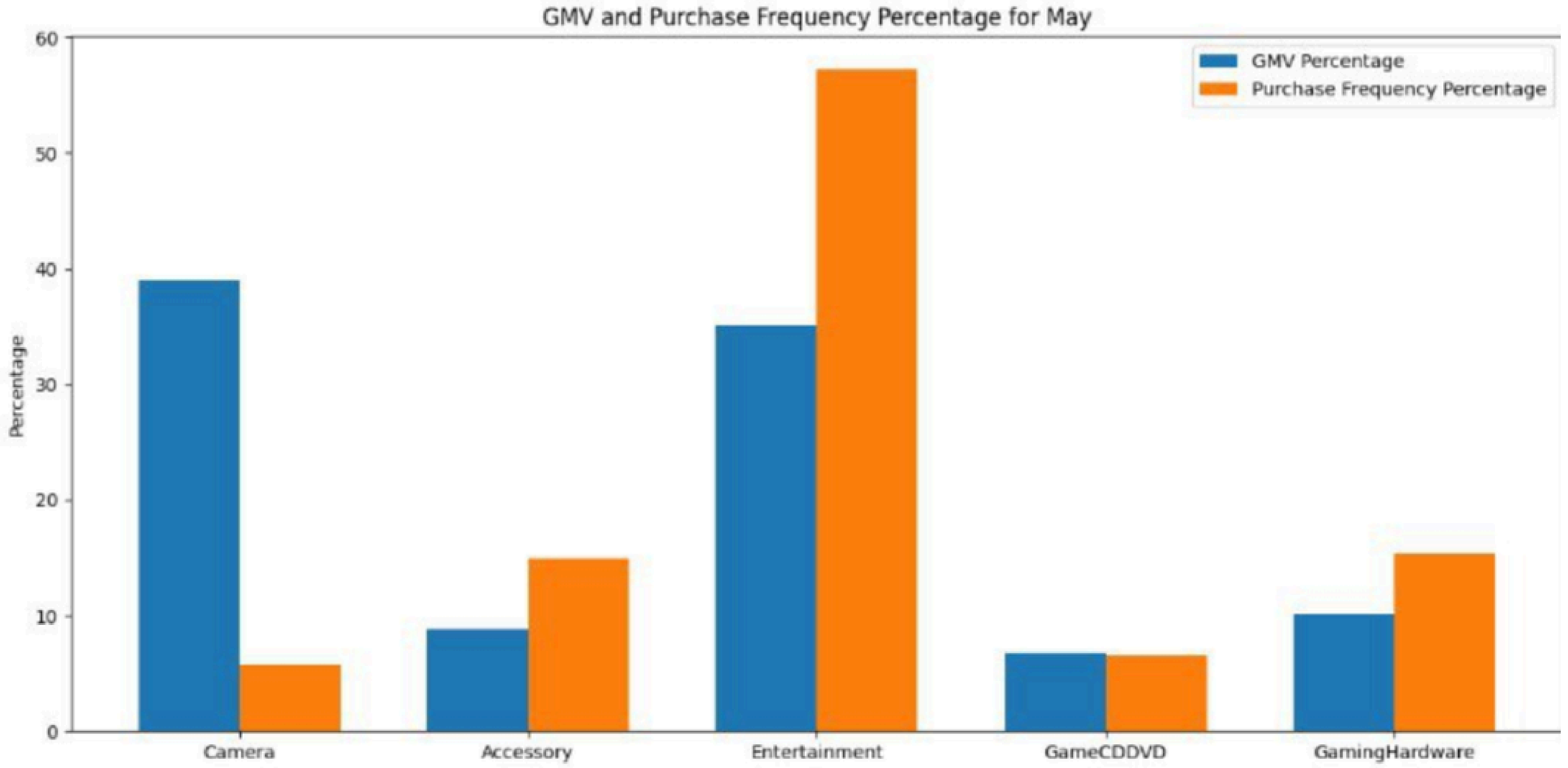
New Dataframe

- Rows: Month
- Columns: total GMV and the total units sold in that month for a specific product type.









OPTIMAL MARKETING CHANNEL ALLOCATION FOR EACH PRODUCT CATEGORY

RESULT ANALYSIS

Revenue Trend:

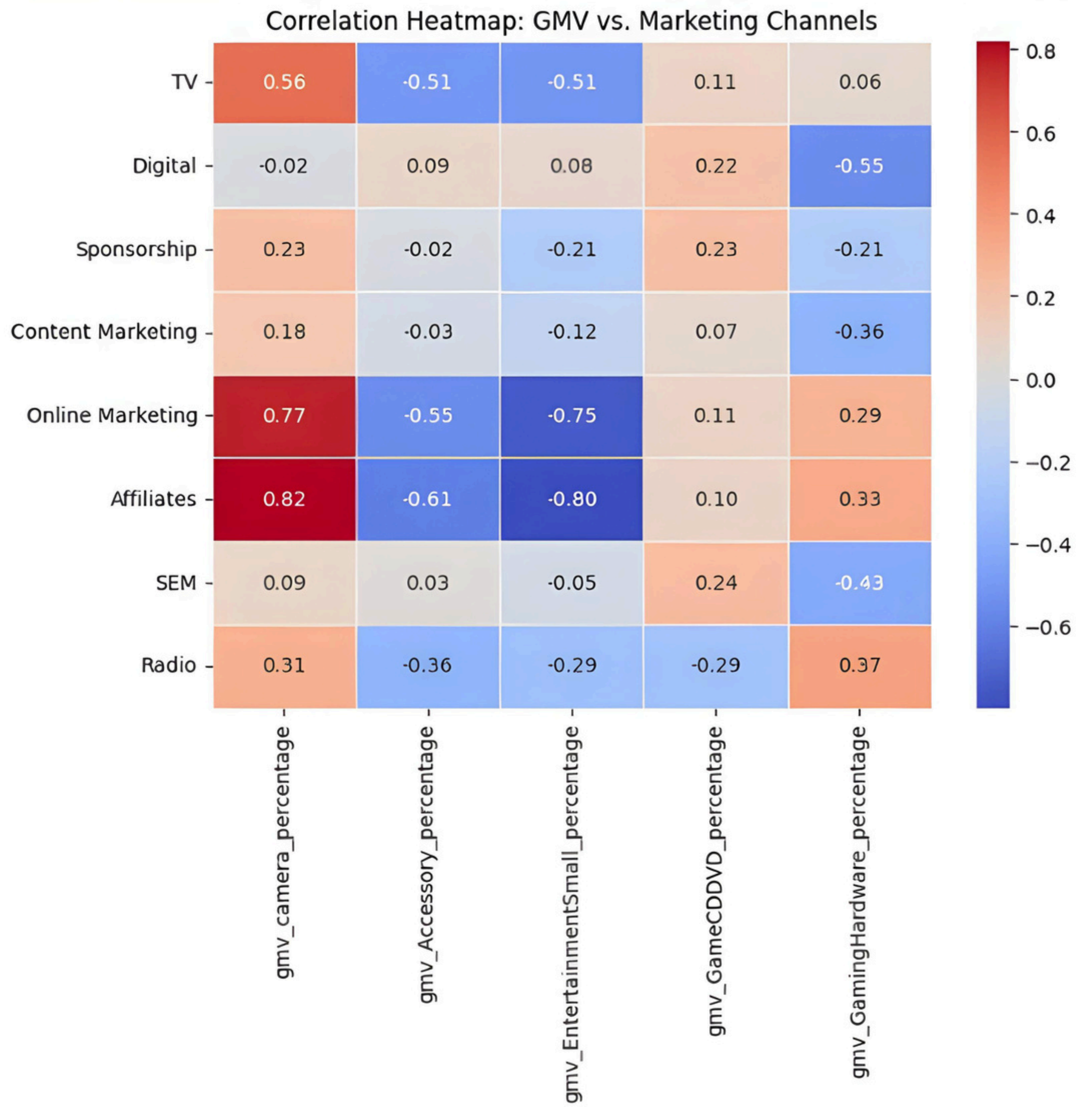
- Entertainment products generate the highest monthly revenue across all months

Purchase Frequency Trend:

- From January to April, entertainment products have the highest purchase frequency
- From May onward, camera products exhibit the highest purchase frequency

KEY INSIGHTS

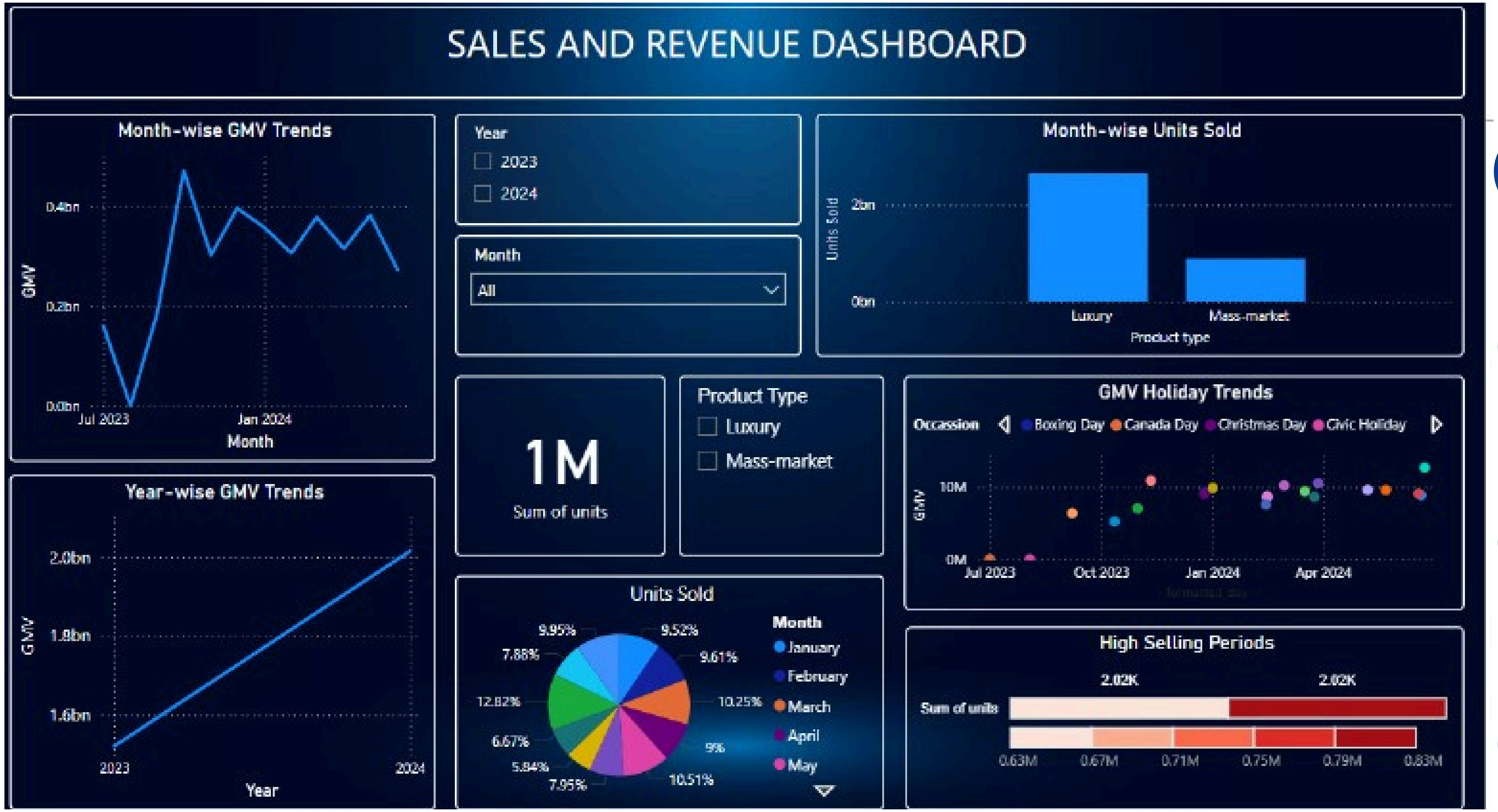
- The dominance of entertainment products in revenue suggests higher-priced items or higher sales volume in this category
- The shift in purchase frequency from entertainment to cameras after April may indicate seasonal demand, marketing influences, or product launches
- Further investigation into average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



Product Type	Marketing Channel to be Leverage
Camera	Affiliates
Camera Accessory	Digital
Entertainment small	Digital
Game CD DVD	Sponsorship
Gaming Hardware	Radio

DASHBOARDS

SALES AND REVENUE DASHBOARD

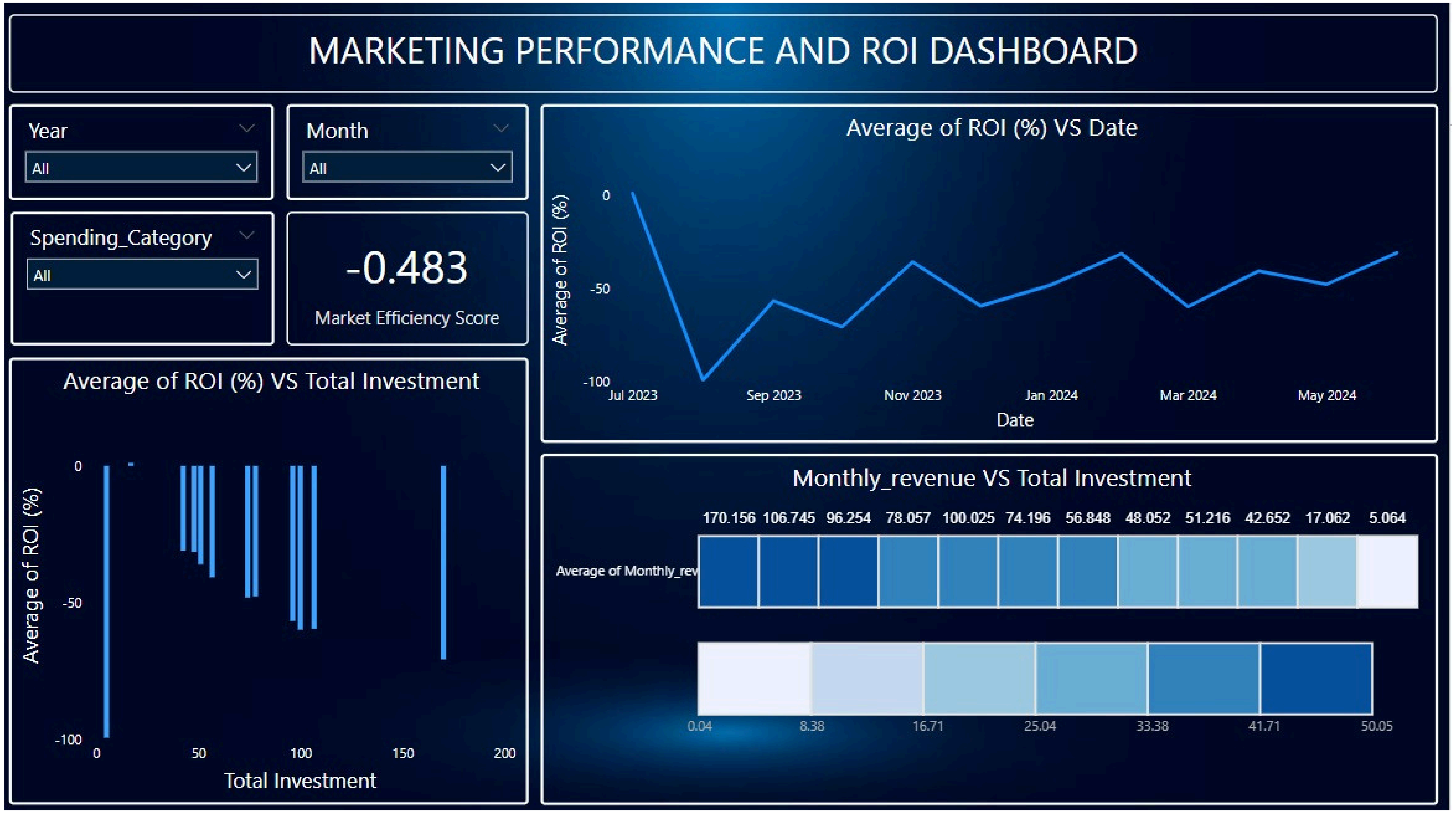


INFERENCES

- GMV shows strong growth from 2023 to 2024
- Luxury products dominate unit sales, surpassing mass-market items
- Holiday events drive GMV spikes, with notable peaks on Boxing Day & Christmas
- May and April are high-selling months, contributing significantly to total units sold

SALES AND REVENUE

MARKETING PERFORMANCE AND ROI DASHBOARD

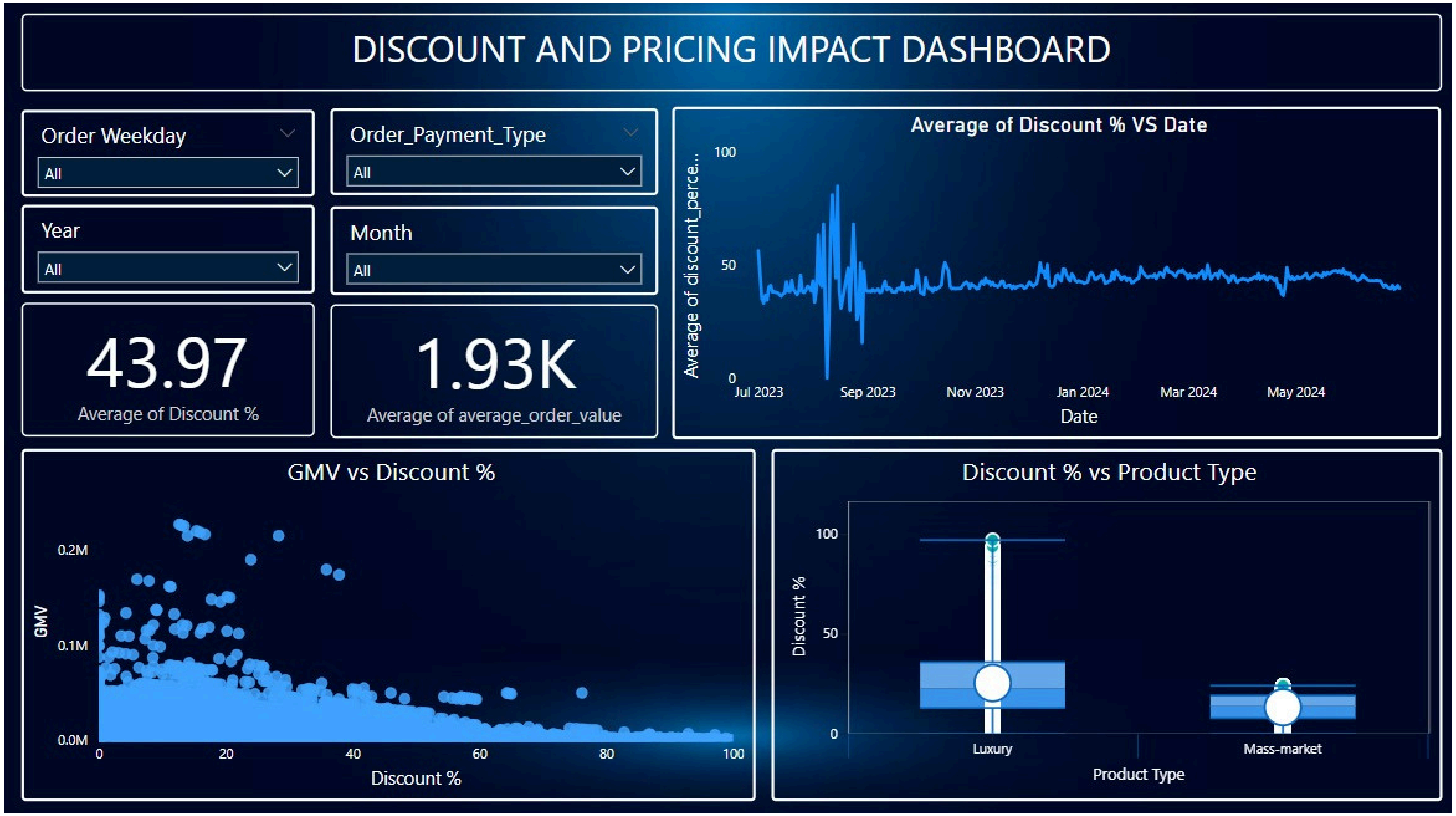


INFERENCES

- -0.483, indicating inefficient marketing spend
- Fluctuations suggest inconsistent performance
- Higher investments don’t always yield better returns
- Some lower investments generate higher revenues
- Optimize budget allocation for better ROI efficiency

MARKETING PERFORMANCE AND ROI

DISCOUNT AND PRICING IMPACT DASHBOARD

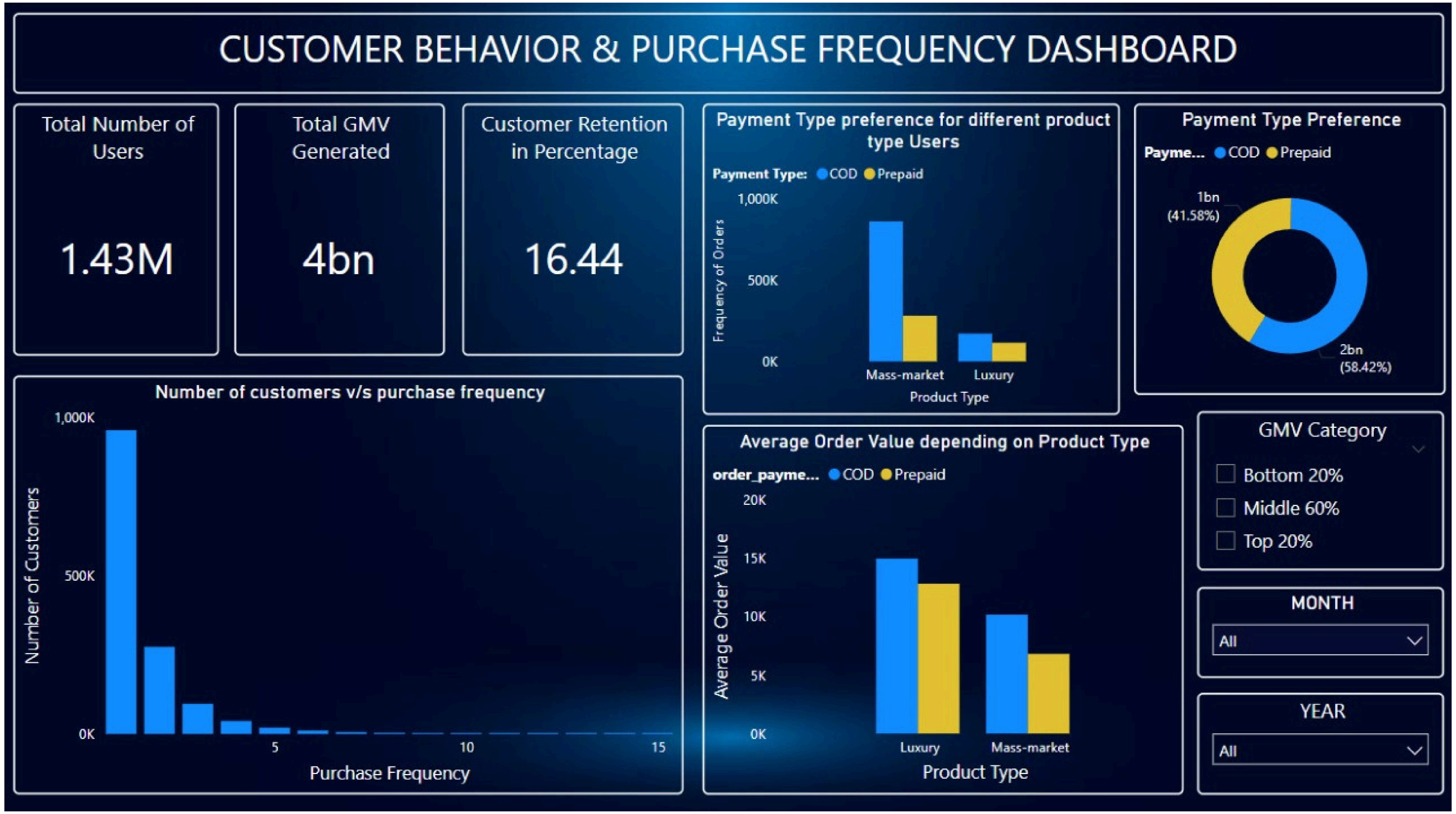


INFERENCES

- Average discount: 43.97%, with an average order value of 1.93K
- Trend: Discounts fluctuated but stabilized recently
- GMV vs Discount: High discounts don’t always drive GMV growth
- Product Type: Luxury items get higher discounts than mass-market
- Takeaway: Optimize discounts for profitability

DISCOUNT AND PRICING IMPACT

CUSTOMER BEHAVIOR & PURCHASE FREQUENCY DASHBOARD

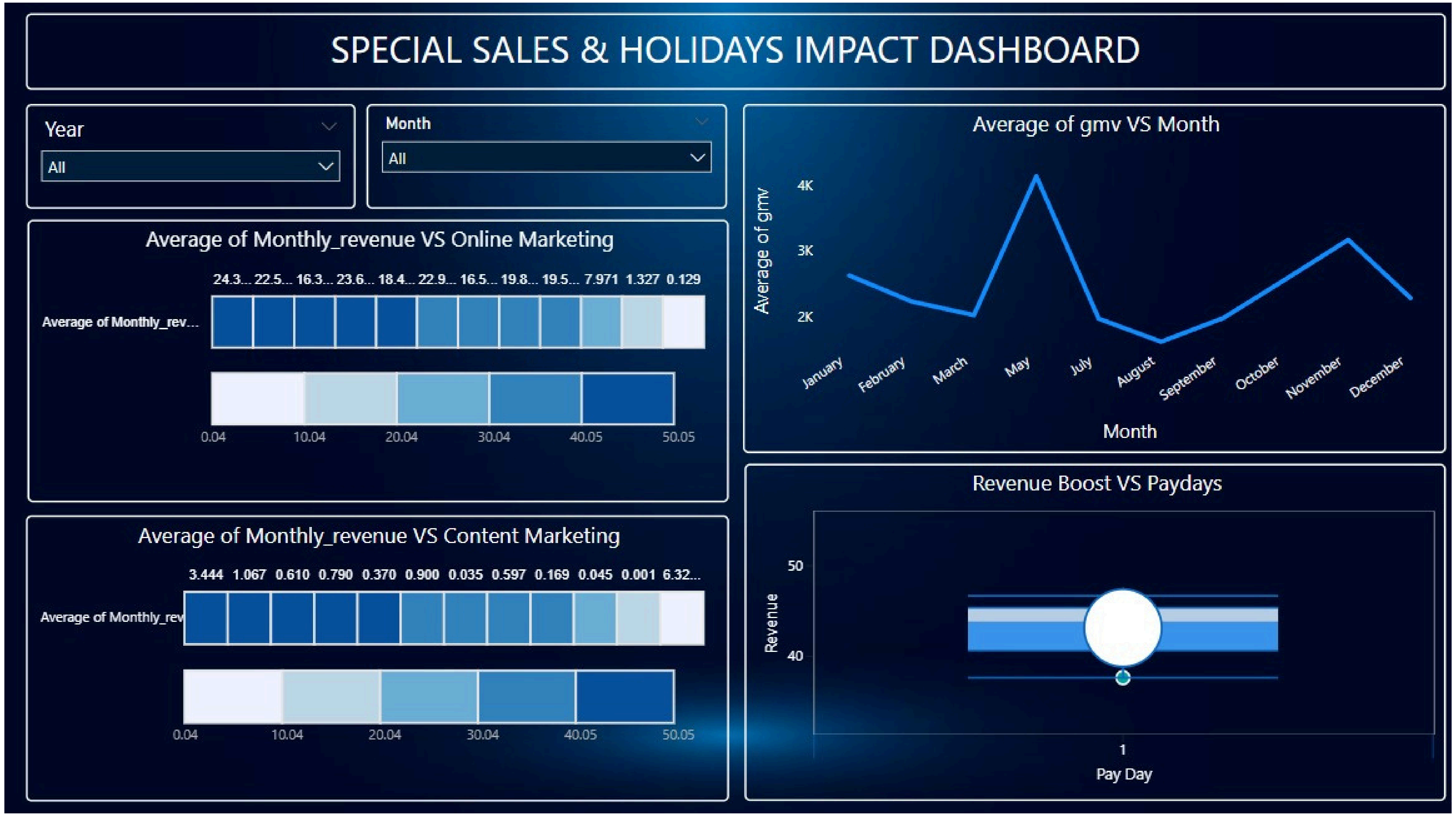


INFERENCES

- Users: 1.43M total users, 4bn GMV, 16.44% retention
- Purchase Frequency: Majority are one-time buyers
- Payment Preference: COD (58.42%) > Prepaid (41.58%)
- Product Type & Payment: Mass-market users prefer COD, luxury users show a mix
- AOV: Luxury items have a higher average order value

CUSTOMER BEHAVIOR & PURCHASE FREQUENCY

SPECIAL SALES & HOLIDAYS IMPACT DASHBOARD



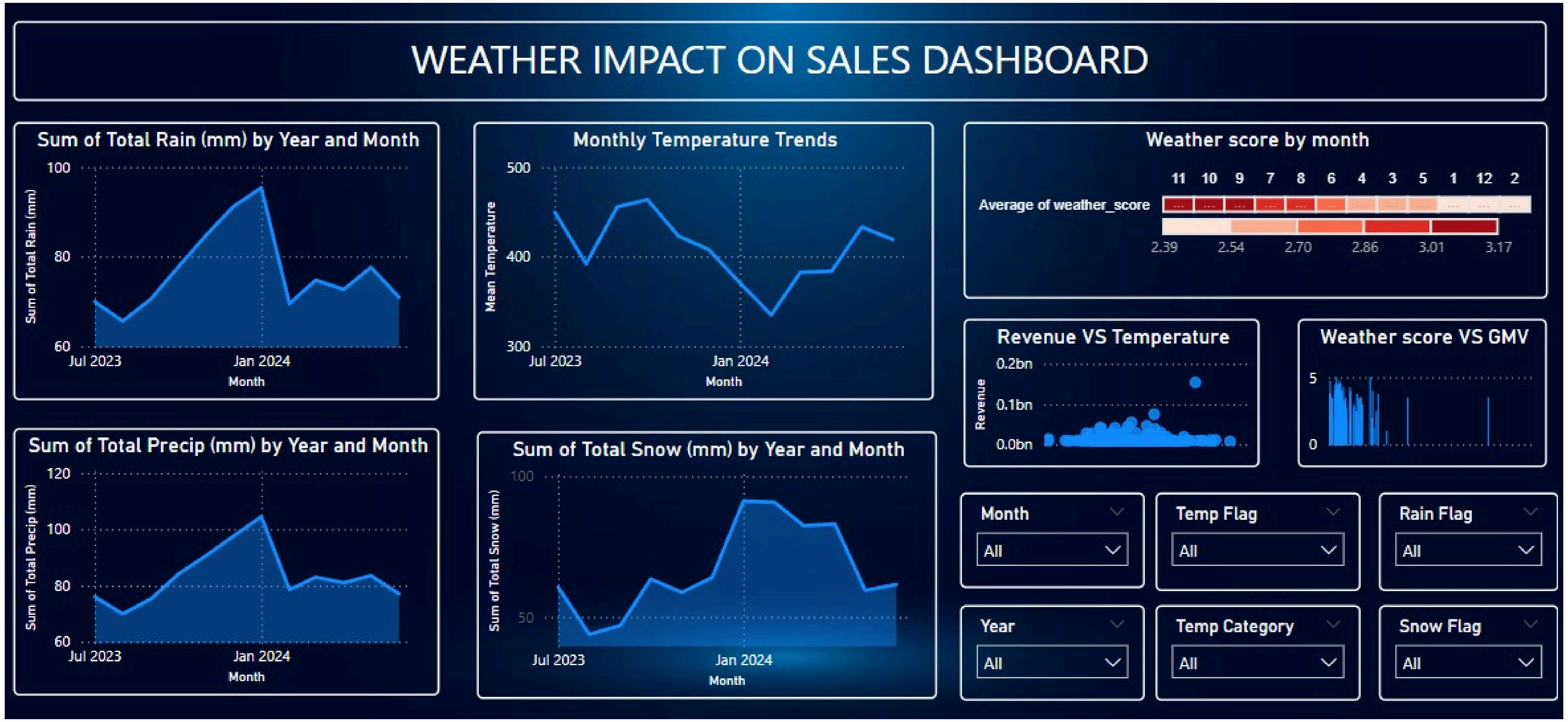
INFERENCES

- GMV Peaks: Significant rise in May, November, and December (holiday & sale periods)
- Marketing Impact: Online marketing drives higher monthly revenue than content marketing
- Payday Effect: Revenue shows a boost on paydays

SPECIAL SALES & HOLIDAYS IMPACT

WEATHER IMPACT ON SALES DASHBOARD

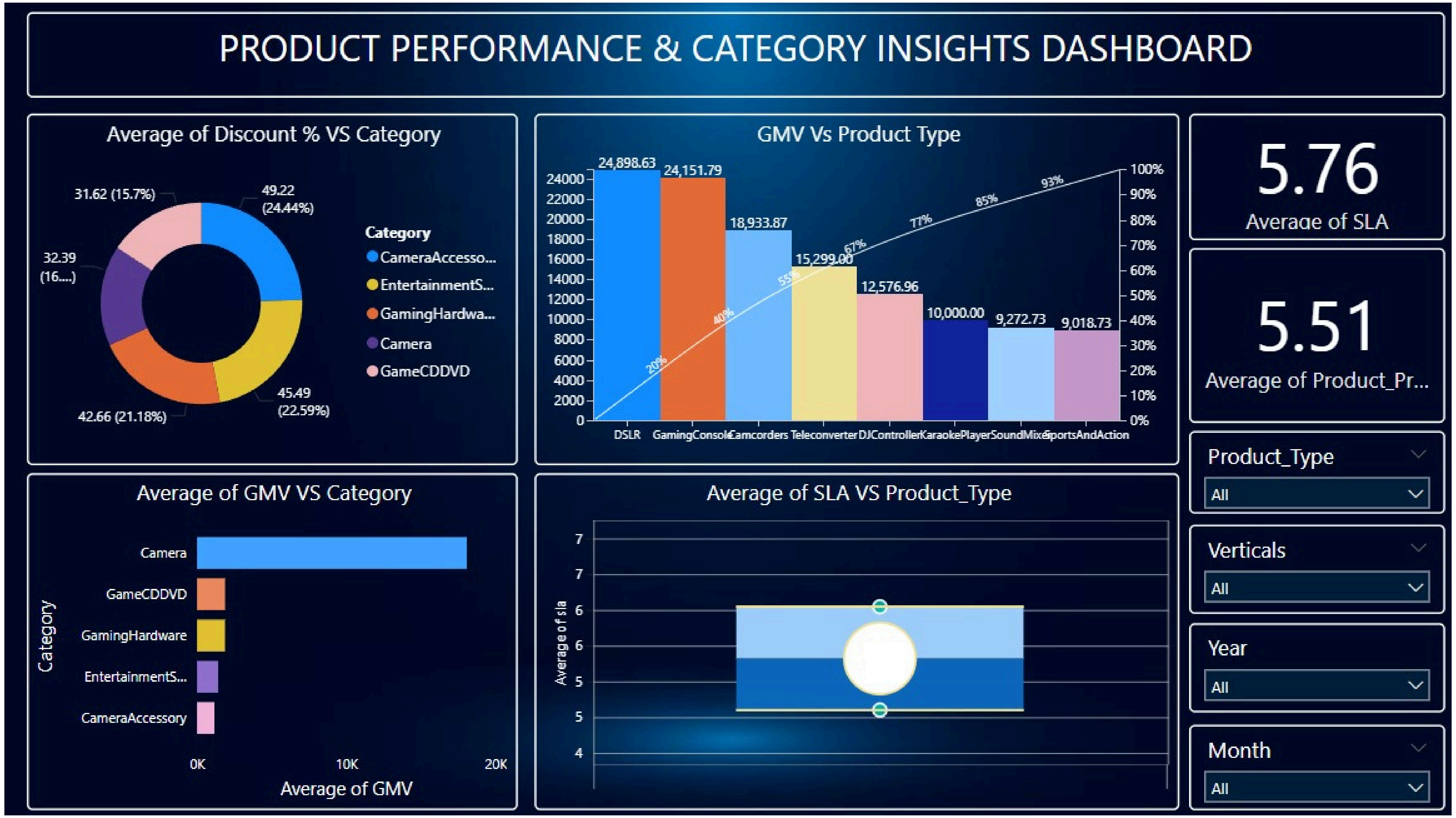
INFERENCES



- Rain, snow, and temperature fluctuate across months, impacting sales
- Higher temperatures show a positive correlation with revenue
- Weather conditions influence GMV, with varying impact across months

WEATHER IMPACT ON SALES

PRODUCT PERFORMANCE & CATEGORY DASHBOARD



INFERENCES

- Camera category leads in GMV, followed by GameCD/DVD
- DSLRs and gaming consoles generate the highest GMV among product types
- Discounts vary across categories, with some receiving over 45%
- SLA remains stable, averaging around 5.76

PRODUCT PERFORMANCE & CATEGORY

THANK YOU