TEAM 2025103

INDEX

01 Introduction

02 Literature Review

03 Data Pre-processing

04 Exploratory Data Analysis

05 Key Performance Indicators

06 Hypotheses Testing

07 Model and Budget/Dashboard

08 Inventory Optimisation

09 Annexure Title

10 Annexure

1. INTRODUCTION

ElectroMart is a leading e-commerce company based in Ontario that specializes in selling electronic products. Over the past year, the company has invested heavily in various marketing channels, including commercials, online campaigns, and promotional events. With Canada's e-commerce market expected to surpass $130 billion by 2027, competition is growing, and businesses must continuously refine their marketing strategies to stay ahead.

As the digital marketplace evolves, ElectroMart wants to ensure that its marketing budget is being spent wisely. The goal is to understand which marketing efforts are driving revenue growth and how to optimize future investments for maximum returns. To do this, data analytics will play a key role in uncovering insights from past performance.

This report will focus on:

- Identifying Key Performance Indicators (KPIs): Understanding which factors (e.g., order frequency, product categories, promotional timing) have the most significant impact on revenue.
- Measuring the Effectiveness of Marketing Channels: Analyzing how different marketing levers—such as TV ads, digital campaigns, and seasonal promotions—affect overall sales.
- Optimizing the Marketing Budget: Recommending the best way to allocate the marketing budget to maximize revenue while minimizing wasted spend.

By applying Exploratory Data Analysis (EDA), predictive modeling, and correlation analysis, hypothesis testing, this study will provide clear, data-driven recommendations to help ElectroMart make smarter marketing decisions. The ultimate goal is to enhance revenue growth, improve marketing efficiency, and strengthen ElectroMart's competitive position in the fast-growing e-commerce industry.

PROBLEM OVERVIEW

ElectroMart's CFO has raised concerns regarding the effectiveness of the company's recent marketing expenditures. Despite substantial spending on various marketing initiatives over the past year, the CFO believes these efforts did not yield sufficient returns. Consequently, there is a need to either reduce the overall marketing budget or allocate it more optimally across different marketing channels to enhance revenue performance. The primary objectives identified for addressing this problem include:

Performance Driver Analysis: Understanding which KPIs significantly influence revenue.

Impact Analysis on Marketing ROI: Quantifying how each commercial lever affects revenue.

Optimising Marketing Spending: Determining how to allocate the budget across various marketing channels to maximize revenue outcomes.

2. LITERATURE REVIEW

Data Analytics in E-Commerce

ElectroMart's CFO has raised concerns regarding the effectiveness of the company's recent marketing expenditures. Despite substantial spending on various marketing initiatives over the past year, the CFO believes these efforts did not yield sufficient returns. Consequently, there is a need to either reduce the overall marketing budget or allocate it more optimally across different marketing channels to enhance revenue performance. The primary objectives identified for addressing this problem include:

Optimising Marketing Budget with Non-Linear Programming

Explored nonlinear optimization to optimise marketing budgets, addressing the inefficiencies of traditional linear approaches. It highlights how diminishing returns impact spending efficiency and demonstrates how businesses can allocate budgets effectively across multiple channels using constrained optimization techniques. The author provides a Python-based implementation using the *scipy.optimize* library to maximize Return on Investment (ROI) while adhering to budget constraints. Key concepts covered include efficiency curves, marginal returns, and budget allocation strategies. By leveraging data-driven optimization, businesses can make smarter spending decisions, ensuring maximum impact while avoiding wasteful investments in less effective marketing channels.

Multi-Channel Marketing with Budget Complementarities

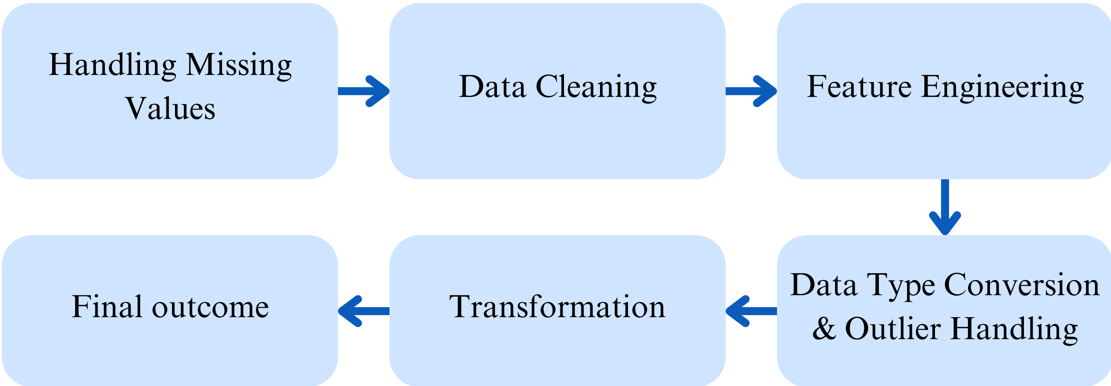
"Multi-Channel Marketing with Budget Complementarities" discusses how to best allocate the budget over many marketing channels while accounting for budget complementarities (for which some investment thresholds are necessary for substantial impact). The authors frame the problem as a multi-choice knapsack problem (MCKP) and suggest an approximation approach to produce nearly optimal budget allocation. They develop two new query algorithms—Generalised Binary Query (GBQ) and Heuristic Binary Query (HBQ)—to calculate efficiently investment thresholds. The research shows with simulations that HBQ provides nearly optimal outcomes at much reduced computation time, presenting a realistic marketing budget optimization solution for practice.

Inventory Management

Efficient warehousing and inventory management are essential for achieving high service levels at low costs. In previous research, companies in industries like food distribution and pharmaceuticals have been studied, with the result that high SKU-complexity companies utilise automation and sophisticated warehousing for quicker deliveries, while low SKU-complexity companies emphasize inventory optimization and demand forecasting. Inventory control and warehousing must be balanced to achieve smooth supply chain operations and efficient services.

3. DATA PREPROCESSING

To ensure data consistency, accuracy, and usability for analysis, multiple data preprocessing steps were performed. These included handling missing values, transforming variables, reducing complexity, and integrating external datasets to enhance insights. Given the nature of ElectroMart's business problem—evaluating marketing effectiveness, sales trends, and operational efficiencies—data preprocessing played a crucial role in preparing a structured and reliable dataset for further analysis.



3.1 HANDLING MISSING VALUES

Dealing with missing values, particularly in the weather dataset, was imperative for understanding the effect of environmental factors on delivery performance. Several imputation methods, such as K-Nearest Neighbors (KNN) Imputer and Multiple Imputation by Chained Equations (MICE) Imputer, were experimented with. Though these yielded fair approximations, the ultimate solution employed a localized imputation strategy, replacing missing values with the average of the previous three and next three observations to ensure temporal consistency.

For order and transactional data where missing values did occur, it was handled through the validation of key identifiers such as customer numbers and location data. Critical rows with missing values were deleted while numerical columns had median-based imputation for preserving data integrity.

3.2 DATA CLEANING

To improve data quality, multiple cleaning steps were performed:

Date standardization: Ensured that all timestamps were converted into a uniform format to facilitate time-based analysis.

Handling categorical variables: Text-based categorical fields were transformed into numerical representations for seamless integration into analytical models.

Replacing null indicators: Missing values encoded as placeholders (e.g., \N, empty strings) were replaced with appropriate defaults or NaN values for better processing.

3.3 FEATURE ENGINEERING

In order to enrich the dataset for more meaningful analysis, some new features were added. Time-based features were created by summing up monthly data and monitoring order frequency patterns to capture trends over time. Holiday and marketing indicators were added to analyze their influence on sales, aiding in measuring the effectiveness of promotional campaigns and seasonal demand variations. Product classification was further enriched by associating SKU information with upper-level categories to facilitate a more organised examination of category-specific revenue contributions. Payment channels were also numerically coded to gain insights into customer affinity and determine their impact on order completion rates. These developments delivered richer insights into sales patterns, marketing efficiency, and customer buying behavior.

3.4 DATA TYPE CONVERSION & OUTLIER HANDLING

To ensure the dataset was optimised for analytical models:

- Numerical data conversion: Columns stored as strings but representing numeric values (e.g., revenue, units sold) were converted to appropriate data types.
- Geographical corrections: Location-based information, such as postal codes, was validated to eliminate inconsistencies.
- Outlier detection: Extreme values in key financial and operational metrics (e.g., revenue, delivery time) were identified and handled using interquartile range (IQR) filtering to prevent distortions in analysis.

3.5 TRANSFORMATION

To improve computational efficiency and streamline the dataset for analysis:

- Unnecessary fields were dropped: Irrelevant or redundant variables that did not contribute to the problem statement were removed.
- Aggregation techniques were applied: Sales data was aggregated at different time levels (daily, monthly) to identify trends.
- Normalization and encoding: Numerical values were normalised for comparability, and categorical variables were transformed for better integration with analytical models.

3.6 FINAL OUTCOME

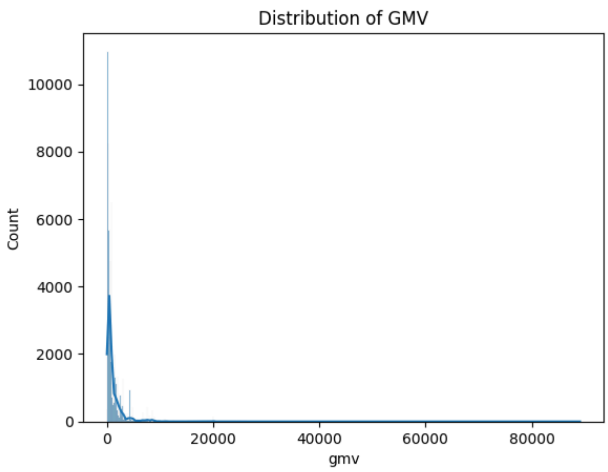
These preprocessing activities made sure that the dataset was formatted, cleaned, and optimised for assessing the effectiveness of marketing efforts, delivery performance analysis, and identifying business insights. Through combining several datasets and using sophisticated imputation, transformation, and cleaning methods, the data was set up for valuable analysis, supporting data-driven decision-making for ElectroMart.

4. EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) assists ElectroMart in discovering patterns and relationships for informed decision-making. It compares revenue (total GMV) with ad expenses to determine marketing budget optimization and analyzes the effect of promotional discounts on revenues. Customer satisfaction (NPS) and payment methods are also compared with their effects on revenue and order fulfillment.

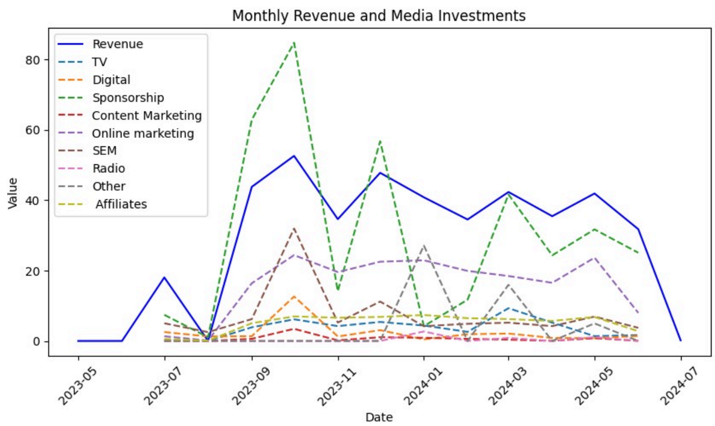
Furthermore, EDA measures delivery performance, seasonal patterns, and category contribution to revenues to determine top-performing products and streamline procurement. By combining marketing, sales, and operations data, it offers strategic insights to improve business strategies, increase efficiency, and accelerate growth.

Distribution of Gross Merchandise Value (GMV)



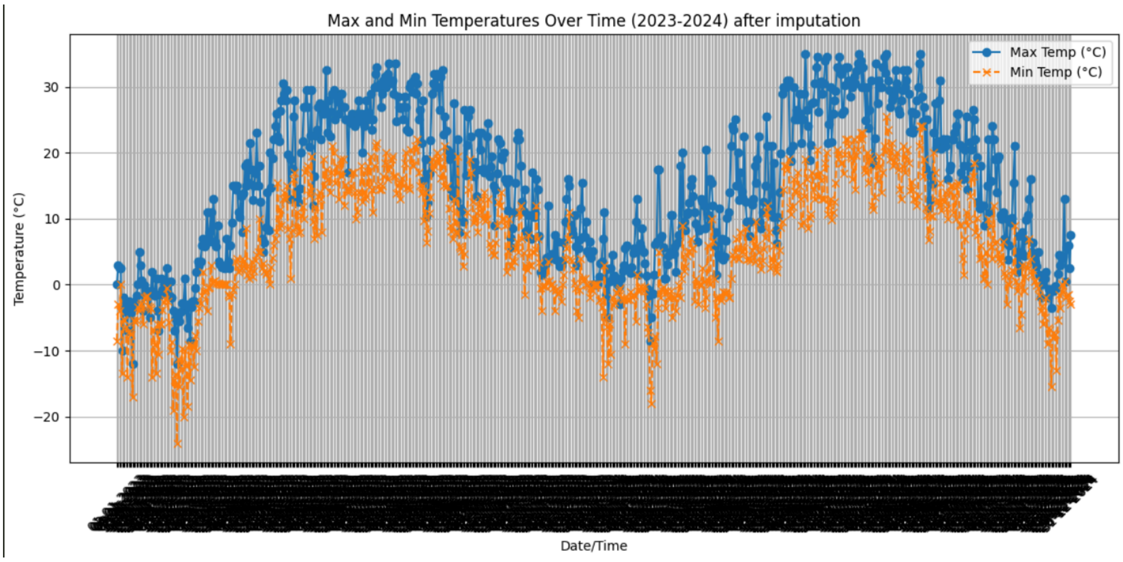
The GMV distribution is highly right-skewed, with most values concentrated near zero, indicating a few high-value transactions driving overall revenue.

Monthly Revenue and Media Investments



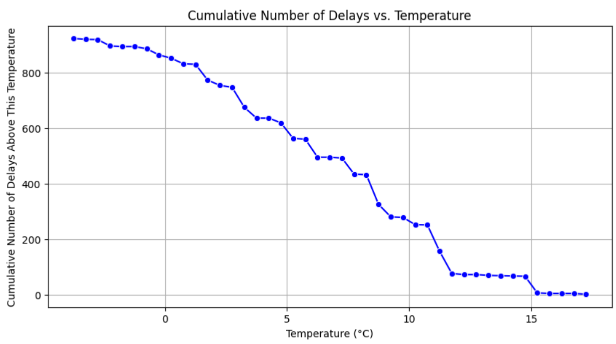
The chart illustrates monthly revenue patterns along with media expenditures on different channels. Revenue highs correlate with increases in sponsorship, online advertising, and SEM expenditure, showing their likely contribution. Decreases in both revenue and expenditure after the initial months of 2024 imply seasonality or market change.

Max and Min Temperature Over Time(2023-2024) after Imputation



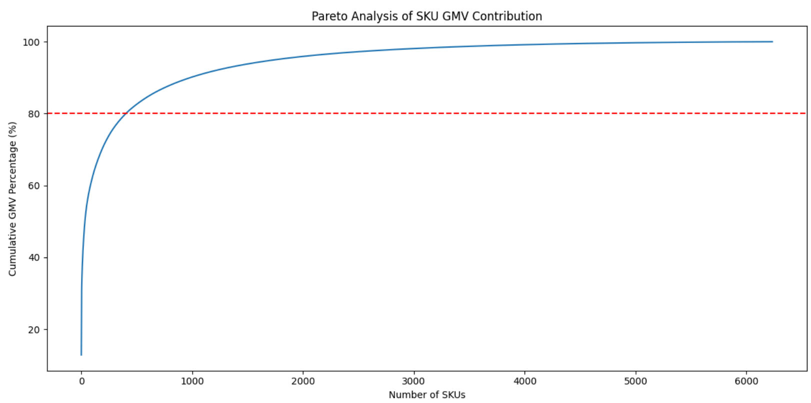
The graph displays maximum and minimum temperatures over time (2023-2024) after MICE imputation, showing clear seasonal trends. Temperature peaks mid-year and drops towards the end, reflecting typical annual weather patterns.

Cumulative Number of Delays vs. Temperature



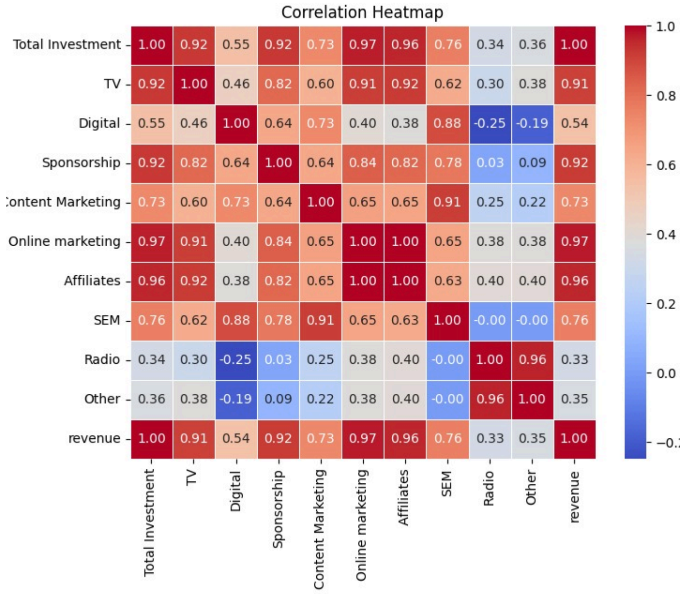
This graph demonstrates that there is a threshold temperature of around 3.5°C, above which the majority of delays tend to occur.

Pareto Analysis of SKU- total GMV Contribution



This Pareto chart shows SKU-total GMV contribution, depicting how a small percentage of SKUs drive most of the total GMV. The red dashed line marks the 80% total GMV threshold, highlighting key revenue-contributing SKUs.

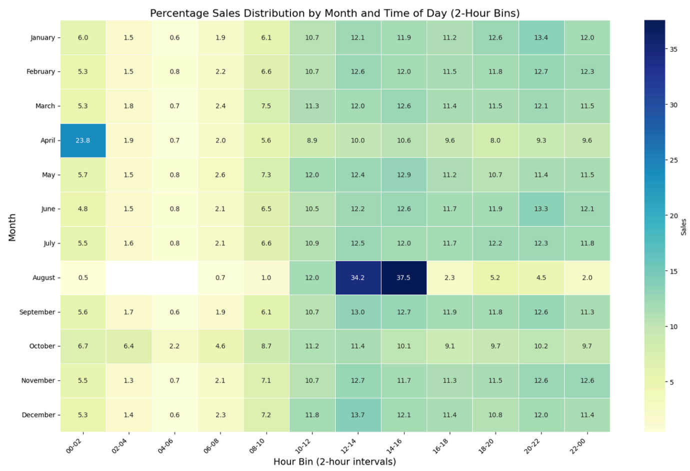
Correlation Heatmap of Revenue with various marketing channels



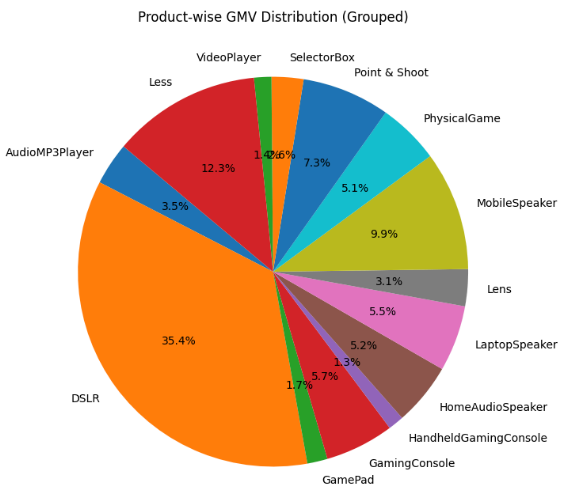
This correlation heatmap displays the correlations among various marketing investments and revenue. Strong positive relationship is represented by higher correlations (nearer to 1), whereas weak or inverse relationships are represented by lower or negative values. Total investment, sponsorship, and online marketing have high correlations with revenue, which means that they have strong influences on sales.

Percentage sales Distribution by Month and Time of Day (2-hour Bins)

The heatmap shows sales distribution by month and time. Sales rise late morning and stabilize in the evening, we also see lower sales (relatively) during sleeping hours and slightly relative higher rates during typical work lunch hours. Using this information allows us to improve our marketing strategy and load management



Product-wise Revenue Distribution



This pie chart shows product-wise distribution of Revenue. DSLRs account for the lion's share (35.4%), followed by Speakers (9.9%) and Point & Shoot devices (7.3%). "Less" refers to products with share of less than 1% (12.3%).

5. KEY PERFORMANCE INDICATORS

Sales and Revenue KPIs

- **Delivery Timeliness:**
- **Formula:**(Number of on-time deliveries / Total number of deliveries) × 100%.
- **First Order Rate :**
- **Formula:** (Number of first-time orders / Total number of orders) × 100%.
- **Repeat Order Rate:**
- **Formula:** (Number of repeat orders / Total number of orders) × 100%.
- **Total Spends per Customer (CLTV - Customer Lifetime Value):**
- **Formula:** Average Order Value × Number of Orders × Retention Period.
- **High Spender:**
- **Formula :** Customers with spending >75th percentile of all customer spending.
- **Low Spender:**
- **Formula :** Customers with spending <=75th percentile of all customer spending

Discount and Pricing KPIs

- **Discount Seeker:**
- **Formula:** (Number of discounted purchases / Total purchases by customer) × 100%. Helps in targeting price-sensitive segments.
- **Discount Rate Threshold:**
- **Formula:** Discount Rate Threshold = min(discount_percentage) where cumulative_gmv ≥ (0.90 × total_gmv)
- **Customer Retention Rate:** 7.7% of customers continue to make purchases.
- **Formula:**Repeat Customers/Total Unique Customers × 100%. This measures loyalty and satisfaction.
- **Warehouse Efficiency:**
- **Formula:** deliverydays=deliverybdays+deliverycdays

Logistics Inefficiency=∑delivery_days/Total Orders

Risk Metrics (KRIs)

- **Delivery Risk:**
- **Formula:** (Delayed deliveries / Total deliveries) × 100%. This identifies potential operational weaknesses.
- **Low NPS Risk:**
- **Formula:** (Detractors / Total respondents) × 100%. This predicts customer dissatisfaction issues.
- **High Churn Risk:** 95.3% of customers stopped making purchases within the given period
- **Formula:** Historical churn rate or predictive model based on behavior patterns. This helps identify at-risk customers for intervention.

Environmental Factors

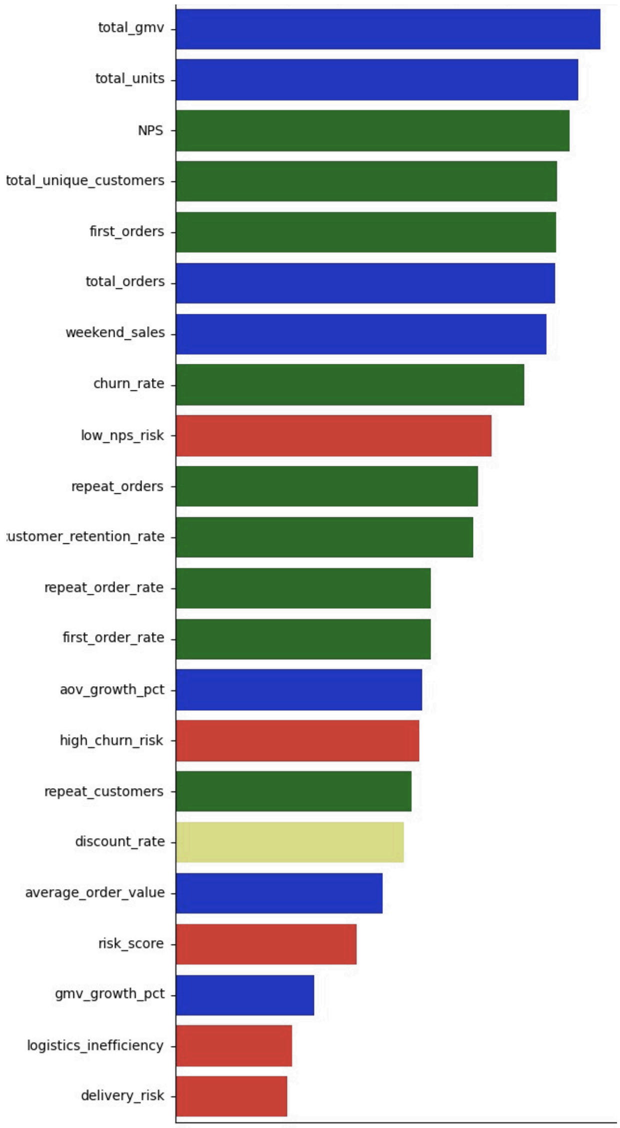
- **Correlation of Weather and Sales:**
- **Formula:** Pearson correlation coefficient between weather metrics and sales figures.
- **Impact of Rain/Snow on Delay:**
- **Formula:** ((Sales on precipitation days - Sales on non-precipitation days) / Sales on non-precipitation days) × 100%.
- **Change in Sales on Weekends : S**ales on weekends are 19% higher campared to weekdays.
- **Formula:** ((Weekend avg. sales - overall_avg_gmv) / overall_avg_gmv) × 100%.
- **Average Mean Temperature Impact:** The relationship between mean temperature and purchasing behavior. Measured through correlation coefficients and regression analysis. This supports seasonal merchandising strategies.

Operational KPIs

- **Customer Order Frequency:**
- **Formula:** (Total Orders in a Month)/(Total Unique Customers in that Month)
- **SLA Compliance:**
- **Formula:** (Number of deliveries meeting SLA / Total deliveries) × 100%. This measures operational reliability.
- **Order Cancellation Rate:**
- **Formula:** Number of canceled orders / Total number of orders
- **Variance of Procurement SLA :** Standard deviation of 53.46.

KPI IMPORTANCE

Relative Importance of KPIs for Revenue using Distance Correlation (Grouped by color)



Distance correlation (dCor) is a statistical measure that quantifies both linear and nonlinear dependencies between two random variables. Unlike Pearson correlation, which only captures linear relationships, distance correlation can detect any kind of dependency, including nonlinear ones.

7. OPTIMISING MARKETING BUDGET

7.1 OBJECTIVE

This project focuses on optimizing marketing budget allocation using a logarithmic model to maximize revenue. By analyzing expenditures across 10 marketing channels over 12 months, we developed a revenue estimation framework that accounts for diminishing marginal returns, ensuring efficient budget utilization. Our optimization model reallocates constrained budgets strategically, enhancing profitability while maintaining spending diversification. Rooted in microeconomic theory, our approach mitigates audience saturation and ad fatigue. ElectroMart, previously profitable in only one month, achieved a 107% ROI and a 7% net profit after implementation. This data-driven model enables dynamic budget adjustments based on empirical insights rather than intuition.

7.2 METHODOLOGY

Logarithmic Modeling

Theoretical Justification:

Marketing expenditure often exhibits decreasing marginal effectiveness, making linear regression unsuitable for modeling returns. The logarithmic specification provides a tractable approximation of this effect:

$$R_i = a_i + b_i \cdot \log(S_i)$$

Aggregate Revenue Representation:

$$R_{total} = \sum_{i=1}^{10} (a_i + b_i \cdot \log(S_i))$$

R_i : Revenue generated by marketing channel

S_i : Spend allocated to channel

a_i, b_i : Coefficients estimated via regression

To enhance model accuracy and interpretability, we employed several analytical techniques. A log transformation was applied to linearize the relationship between marketing expenditures and revenue, addressing heteroscedasticity and improving model clarity. **Ridge regression** was used to counteract multicollinearity by introducing a penalty term, ensuring stable and reliable coefficient estimates even when input variables were highly correlated. Additionally, cross-validation with a k-fold approach was implemented to assess model robustness and prevent overfitting. By testing performance across multiple data subsets, this technique ensured that the model generalises well to new data, leading to more reliable budget allocation recommendations.

HYPOTHESIS TESTING

The t-value in regression analysis helps determine whether a predictor variable has a statistically significant impact on the dependent variable. It is calculated as:

$$t=\beta/SE$$

Hessian Matrix Computation

$H=\partial 2J(w)/\partial w2$ $H^{-1}=Cov(w)$ $\sigma^2=diag(H-1)$

It measures the curvature of the function, helping assess weight uncertainty. Higher variance indicates more uncertainty in that weight.

t values for model coefficients

t test([7807.44560602, 260.01897534, 9774.986566 , 9800.46350299, 8307.18594449, 7899.81739084, 10080.42236126, 4577.1492911 , 4049.71012565])

Higher t-value → Greater confidence that the predictor significantly affects the outcome.

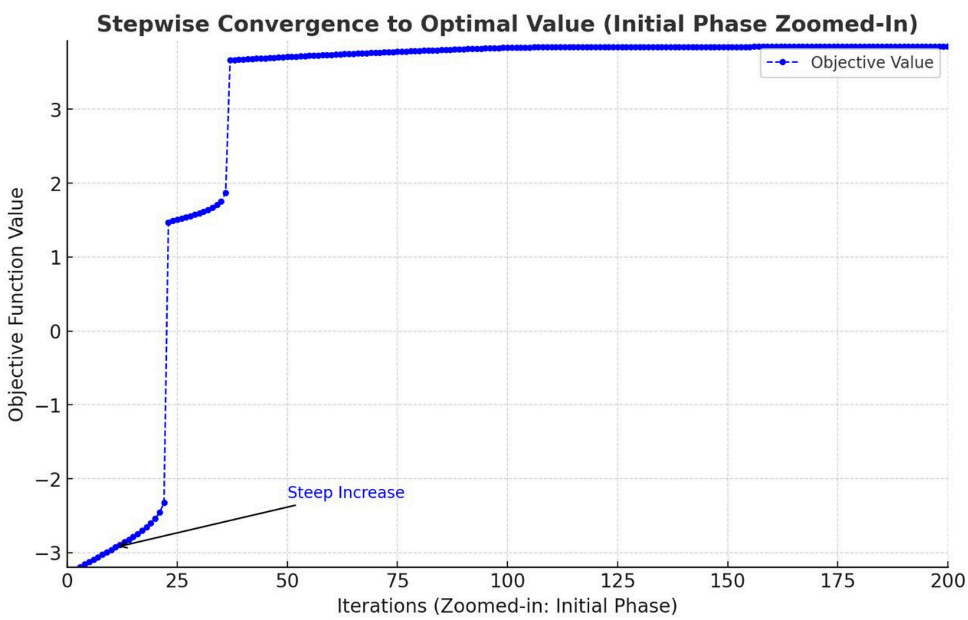
T critical for 5% significance level: 2.2621571628540993

F value for 5% significance level: 19.38

The F-test determines whether a regression model significantly explains the variation in the dependent variable (GMV) compared to a simpler model

f value : 6.700588438906192

$$F=\frac{\text{Explained Variance (Regression Mean Square)}}{\text{Unexplained Variance (Residual Mean Square)}}$$



7.3 OPTIMIZATION FRAMEWORK

Maximise total revenue subject to fixed budget constraint

$$\text{Maximize} \sum_{i=1}^{10} (a_i + b_i \cdot \log(S_i)) \quad \text{Subject to} \quad \sum_{i=1}^{10} S_i = \text{Total Budget}$$

We employed **gradient ascent** to iteratively adjust budget allocations, maximizing revenue based on estimated elasticities. Normalization constraints ensured that the total spend remained within budget while maintaining minimum and maximum allocation limits per channel to reflect operational constraints. Post-optimization, sensitivity analysis was conducted to evaluate the robustness of results under different budget scenarios and spending variations. This approach offered key benefits, including computational efficiency, achieving rapid convergence with minimal iterations, making it ideal for real-time planning. Additionally, it demonstrated adaptability by accommodating business constraints and channel-specific limits, ensuring practical applicability in dynamic marketing environments.

7.4 RESULTS AND INSIGHTS

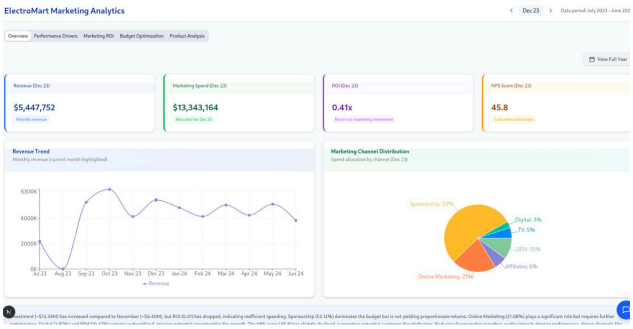
Ridge Regression demonstrated strong predictive reliability, yielding high R^2 values and effectively capturing diminishing returns. The log-linear relationship identified key inflection points where additional spending became inefficient, enabling targeted budget reductions. Implementing the optimization model led to a 128% ROI, with a 28% net profit, a significant improvement from the prior scenario where profitability was achieved in only one out of 12 months. Analysis revealed that previous budgets were overallocated to saturated channels, while the model successfully redirected funds to higher-elasticity channels, enhancing overall marketing efficiency. Sensitivity analysis confirmed the model's robustness under varying budget constraints.

7.5 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The logarithmic model effectively captures diminishing returns in marketing expenditures, providing a mathematically sound framework for budget allocation. Ridge Regression proved valuable in mitigating overfitting and ensuring stable coefficient estimation despite collinearity. The gradient-based optimization approach improved marketing efficiency by reallocating resources to high-ROI channels based on data-driven insights rather than heuristics.

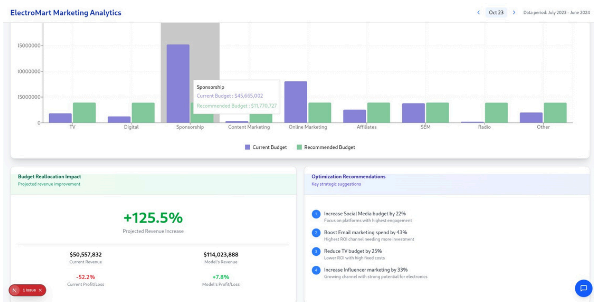
For future research, integrating macroeconomic indicators such as inflation and GDP trends could refine spending elasticity estimates. Developing real-time decision systems with interactive dashboards would enable dynamic budget adjustments. Additionally, applying reinforcement learning techniques could automate budget allocation in response to real-time market changes.

DASHBOARD



This dashboard provides insights into marketing performance, spending efficiency, and customer engagement. It highlights revenue trends, ROI, budget allocation, and NPS, helping identify inefficiencies, optimize spending, and improve ROI through data-driven decisions.

The dashboard shows major revenue drivers such as Marketing Channels and Sales, as well as Customer Engagement, Pricing, and External Factors. It monitors KPIs such as GMV Growth, Order Value, and Conversion Rate to look for optimization areas.



The dashboard also examines marketing expenditure, highlighting inefficiencies and budget optimization. It compares current vs. optimal allocation, focusing on high-ROI channels such as social media and email, while minimizing inefficient investments to drive revenue and profitability.

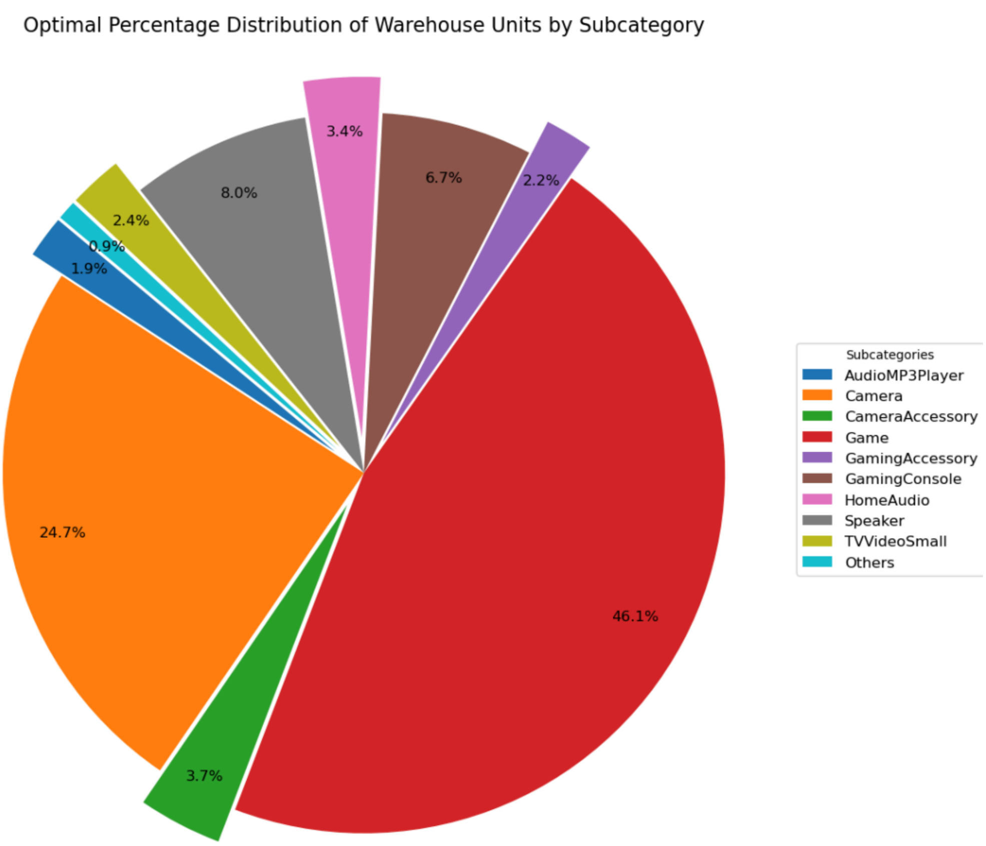
This dashboard breaks down product subcategory performance, indicating cameras as the leading revenue driver, followed by speakers. Small TVs and camera accessories show significant growth. Gaming accessories and audio storage have little contribution. Leading revenue-generating products are DSLRs and mobile speakers, which indicate market demand.

8. INVENTORY OPTIMISATION

We optimized inventory allocation by focusing on two key factors: procurement efficiency and sales contribution. First, we calculated the weighted average procurement Service Level Agreement (SLA) for each product subcategory by considering SLA values across different products. This provided a representative procurement SLA for each subcategory. Second, we analyzed the percentage contribution of each subcategory to total sales. By multiplying these two factors—weighted SLA and sales contribution—we derived coefficients for each subcategory. These coefficients were then used to adjust the allocated units to maximize overall revenue. The optimization objective was to maximize the sum of these weighted unit allocations across all subcategories. To achieve this, we employed a gradient descent optimization algorithm, ensuring an efficient and data-driven approach to inventory management.

$$Maximize \sum_{i=1}^{n} (WeightedAverageSLA_i \times PercentageSales_i \times Units_i)$$

where n is the number of subcategories, the optimal percentage distribution of warehouse units for each subcategory is represented in the pie chart below, with the weights provided as:



ANNEXURE

SALES AND REVENUE KPIS

- **1)Delivery Timeliness:** Measures how frequently orders are delivered within promised timeframes. Calculated as the percentage of orders delivered on or before the promised delivery date. This KPI indicates operational efficiency and customer satisfaction with the fulfillment process. Formula: (Number of orders delivered on time / Total number of orders) × 100%. Tracking this helps identify bottlenecks in the delivery process and improve customer experience.
- **First Order Rate :** Measures the percentage of orders coming from new customers. Calculated by dividing the number of first-time orders by total orders. This KPI indicates how effectively the business is acquiring new customers. Formula: (Number of first orders / Total number of orders) × 100%. A healthy business typically balances new customer acquisition with repeat purchases.
- **Repeat Order Rate:** Indicates the percentage of orders placed by existing customers. Calculated by dividing repeat orders by total orders in a given period. This metric reflects customer loyalty and satisfaction with previous purchases. Formula: (Number of repeat orders / Total number of orders) × 100%. Higher repeat rates generally indicate stronger customer retention and brand loyalty.
- **Total Spends per Customer (CLTV - Customer Lifetime Value):** Represents the total revenue a business can expect from a single customer throughout their relationship. Calculated by multiplying average purchase value by average purchase frequency and average customer lifespan. Formula: Average Purchase Value × Average Purchase Frequency × Average Customer Lifespan. CLTV helps determine how much to invest in customer acquisition and retention strategies.
- **High Spender:** A customer spending more than the 75th percentile of spending.Formula : Customers with spending >75th percentile of all customer spending.
- **Low Spender:** A customer spending less than or equal to the 75th percentile of spending.Formula : Customers with spending <=75th percentile of all customer spending

Discount and Pricing KPIs

- **Discount Seeker:** Identifies customers who primarily purchase products when discounts are offered. Typically measured by the percentage of a customer's orders that include discounted items. This KPI helps segment customers based on price sensitivity. Formula: (Number of discounted orders by customer / Total orders by customer) × 100%. Understanding discount-seeking behavior helps optimize pricing and promotional strategies.

- **Discount Rate Threshold:**Represents the minimum discount percentage that effectively stimulates purchase behavior for different customer segments. This KPI helps determine optimal discount levels for promotions. Analysis typically involves testing different discount rates against conversion rates. Finding the ideal threshold prevents excessive margin erosion while maintaining sales velocity.
- **Warehouse Efficiency:** Measures how effectively warehouse operations utilize resources to fulfill orders. Key metrics include picking accuracy, order processing time, and inventory turnover. Formula : (Number of accurately picked orders / Total orders) × 100% or (Number of orders processed / Labor hours). Efficient warehousing reduces operational costs and improves delivery times.
- **Customer Retention Rate:**Measures the percentage of customers who continue to purchase from the business over a specific period. Calculated by dividing the number of customers retained by the total number at the start of the period. Formula: ((Number of customers at end of period - New customers acquired during period) / Number of customers at start of period) × 100%. Higher retention rates typically correlate with profitability.

Risk Metrics (KRIs)

- **1)Delivery Risk:** Identifies orders with elevated risk of delivery issues, delays, or failures. Calculated using predictive models that consider factors like destination, carrier performance, and weather conditions. This KPI helps prioritize preventive actions for high-risk shipments. Monitoring delivery risk allows for proactive customer communication and contingency planning.Formula: (Delayed deliveries / Total deliveries) × 100%. This identifies potential operational weaknesses.
- **2)Low NPS Risk:** Identifies customers likely to give low Net Promoter Scores based on their experience. Predictive models use factors like delivery issues, customer service interactions, and product quality complaints. Formula: (Detractors / Total respondents) × 100%. This predicts customer dissatisfaction issues.
- **3)High Churn Risk:** Identifies customers with an elevated probability of not making future purchases. Uses behavioral indicators like decreasing purchase frequency, reduced engagement, and browsing competitors. Typically expressed as a probability score from predictive models. Early identification of churn risk enables targeted retention efforts before customers defect.Formula: Historical churn rate or predictive model based on behavior patterns. This helps identify at-risk customers for intervention.

Environmental Factors KPIs

- **Correlation of Weather and Sales:** Analyzes the statistical relationship between weather conditions and sales performance. Typically measured using correlation coefficients between weather variables (temperature, precipitation) and sales metrics. Values range from -1 (strong negative correlation) to +1 (strong positive correlation). Understanding these relationships enables weather-based sales forecasting and inventory planning.
- **Impact of Rain/Snow on Delay:** Quantifies how precipitation events affect sales volumes and patterns. Typically calculated by comparing sales on rainy/snowy days to similar non-precipitation days. Formula: ((Sales on precipitation days - Sales on non-precipitation days) / Sales on non-precipitation days) × 100%. This analysis helps optimize staffing, inventory, and marketing during weather events.
- **%Change in Sales on Weekends:** Measures the difference in sales performance between weekends and weekdays. Calculated by comparing average weekend daily sales to average weekday daily sales. Formula: ((Weekend average daily sales - Weekday average daily sales) / Weekday average daily sales) × 100%. Understanding this pattern helps optimize staffing, inventory, and promotional timing.
- **Average Mean Temperature Impact:** Analyzes how ambient temperature affects sales volumes across different product categories. Typically measured by correlating temperature ranges with sales performance. This KPI helps predict demand fluctuations based on weather forecasts. Temperature sensitivity analysis enables proactive inventory and marketing adjustments for seasonal transitions.

Operational KPIs

- **Customer Order Frequency:** Tracks how often individual customers place orders within a specific time period. Calculated as the average number of orders per customer per month. Formula: (Total Number of Orders / Total Number of Unique Customers). Higher frequency typically indicates stronger customer engagement and loyalty to the platform.
- **SLA Compliance:** Tracks the percentage of deliveries that meet promised Service Level Agreements. Calculated by dividing the number of compliant deliveries by total deliveries. Formula: (Number of deliveries meeting SLA / Total number of deliveries) × 100%. High compliance rates indicate operational reliability and customer satisfaction.
- **Order Cancellation Rate:**Measures the percentage of orders that are canceled before fulfillment, either by customers or the business. Calculated by dividing the number of canceled orders by the total number of orders placed in a specific period. Formula: (Number of Canceled Orders / Total Number of Orders) × 100%. High cancellation rates may indicate problems with product descriptions, checkout experience, inventory management, or competitive pricing. Monitoring trends in cancellation reasons provides insights for operational and website improvements.

- **A high Variance of Procurement SLA:** It indicates significant inconsistencyin the time suppliers take to fulfill orders. This inconsistency can lead to operational inefficiencies, such as delays in inventory restocking, disrupted order fulfillment, and increased uncertainty in supply chain management. Unpredictable procurement times may result in stockouts or overstocking, affecting both customer satisfaction and business performance. To improve supply chain stability, companies should focus on streamlining supplier processes, negotiating better procurement terms, and implementing data-driven forecasting methods. Reducing this variance enhances efficiency, reliability, and overall operational predictability.

REFERENCES

- https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/ea6310e7-8ef8-43359625b82b11042797/content#:~:text=Using%20a%20parameter%2Dbased%20nonlinear,the%20advertising%20budget%20and%20performance.
- https://medium.com/data-science/the-art-of-spending-optimizing-your-marketing-budget-with-nonlinear-optimization-6c8a39afb3c2
- https://www.sciencedirect.com/science/article/pii/S0957417425002714
- https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/ea6310e7-8ef8-4335-9625-b82b11042797/content#:~:text=Using%20a%20parameter%2Dbased%20nonlinear,the%20advertising%20budget%20and%20performance
- https://www.nature.com/articles/s41598-022-19728-x
- https://dl.acm.org/doi/abs/10.1504/ijdats.2023.133019
- https://link.springer.com/chapter/10.1007/978-3-031-42455-7_31
- https://www.sciencedirect.com/science/article/pii/S0148296323006136
- https://www.researchgate.net/publication/354651302_MARKETING_BUDGET_TOWARDS_AN_EFFECTIVE_MARKETING_PLAN
- https://optimization-online.org/wp-content/uploads/2021/03/8303.pdf
- https://github.com/aakashdhruva/Marketing-Budget-Optimization
- https://aamas.csc.liv.ac.uk/Proceedings/aamas2017/pdfs/p1232.pdf
- https://ceur-ws.org/Vol-3179/Paper_9.pdf
- https://prasadnaik.faculty.ucdavis.edu/wp-content/uploads/sites/422/2016/11/jmr2011.pdf
- https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9377156
- https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10742814
- https://www.sciencedirect.com/topics/computer-science/shapley-additive-explanation
- https://ijcrt.org/papers/IJCRT2307594.pdf

- https://www.researchgate.net/publication/353685938_Evaluation_of_the_Performance_of_E Commerce_Using_the_Analytic_Hierarchy_Process_AHP_Business_Perspectives_on_E-Commerce
- https://www.researchgate.net/publication/341139608_The_Importance_of_Weather_for_E-Commerce_Orders_Forecasting
- https://www.researchgate.net/publication/312247476_A_retail_store_SKU_promotions_optimization_model_for_category_multi-period_profit_maximization
- https://arxiv.org/pdf/2409.13847v1
- https://arxiv.org/pdf/1902.01128
- https://arxiv.org/pdf/2211.01160
- https://arxiv.org/pdf/1705.07874
- https://github.com/shap/shap
- https://arxiv.org/pdf/1205.0638
- https://www.researchgate.net/publication/353685938_Evaluation_of_the_Performance_of_E-Commerce_Using_the_Analytic_Hierarchy_Process_AHP_Business_Perspectives_on_E-Commerce
- https://business.adobe.com/content/dam/dx/us/en/blogs/pdfs/adobe-dei-weather-report_2023.pdf
- https://arxiv.org/pdf/2412.09142
- https://www.mayainsights.com/blog/business-intelligence-for-marketing/
- https://arxiv.org/pdf/1601.07235