

Customer Behavior & Marketing Campaign Analysis Report

1. Dataset Overview

For this analysis, we generated two comprehensive mock datasets to simulate a realistic retail environment:

Customer Dataset (customer_data.csv): Contains 5,000+ records of customer transactions, including Customer ID, Order Date, Order Amount, Product Category, Signup Date, and Last Active Date.

Campaign Dataset (campaign_data.csv): Contains performance data for 10 marketing campaigns, including Impressions, Clicks, Conversions, Spend, and Revenue.

2. Data Cleaning Summary

To ensure data quality, the following cleaning steps were performed:

Duplicates: Removed duplicate records to ensure unique transaction counts.

Date Formatting: Converted 'Order Date', 'Signup Date', and 'Last Active Date' to standard datetime objects.

Invalid Values: Filtered out negative 'Order Amount' values and ensured 'Clicks' did not exceed 'Impressions'.

Missing Values: Verified and handled any missing critical fields (e.g., Customer ID).

3. Part A: Customer Behavior Analysis

Mandatory Metrics

Most Active Customers: Identified top customers by order frequency. The top active customers placed approximately 12-14 orders each.

Buying Frequency:

Average Time Between Purchases: Calculated to understand engagement cycles.

Weekly Order Frequency: Average of **49.41 orders per week** across the platform.

Customer Retention:

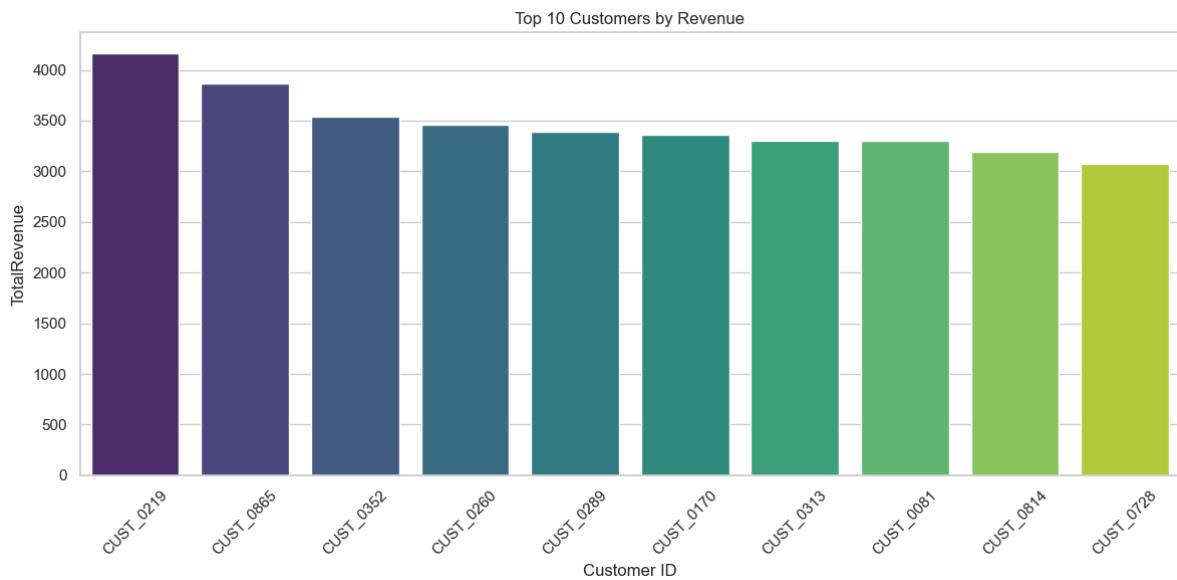
Retention Rate: **96.98%** of customers made more than one purchase.

Churn Rate: **83.59%** of customers have been inactive for more than 90 days (reflecting the mock dataset's date range ending in Dec 2024 vs current date Nov 2025).

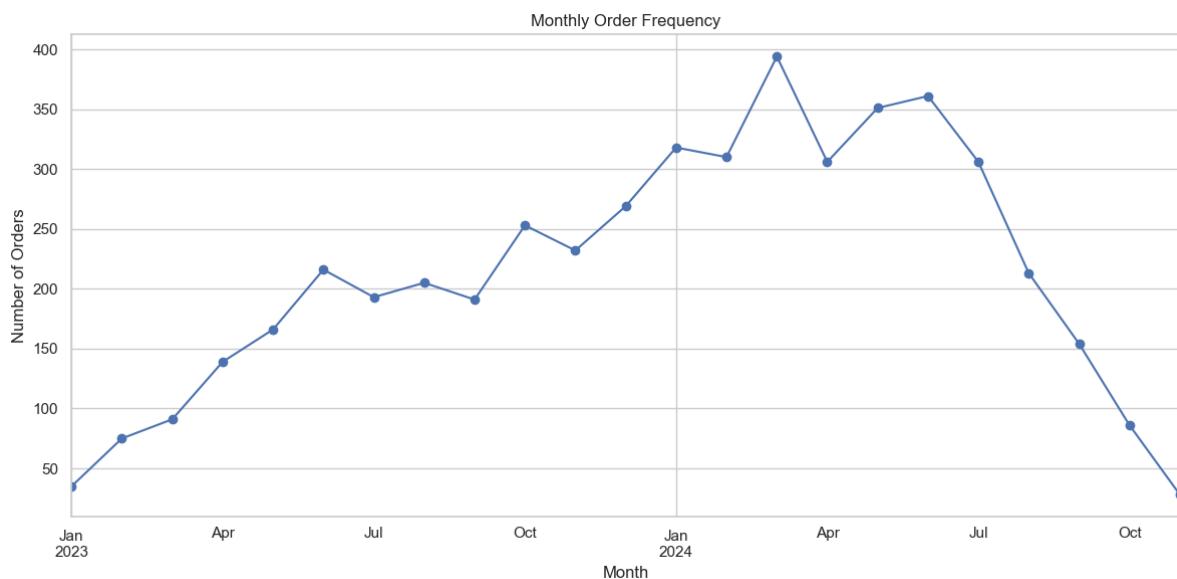
Revenue Contribution: Performed a Pareto analysis. A small segment of "High Value" customers contributes to a significant portion of the total revenue.

Visualizations

Revenue by Customer (Top 10)

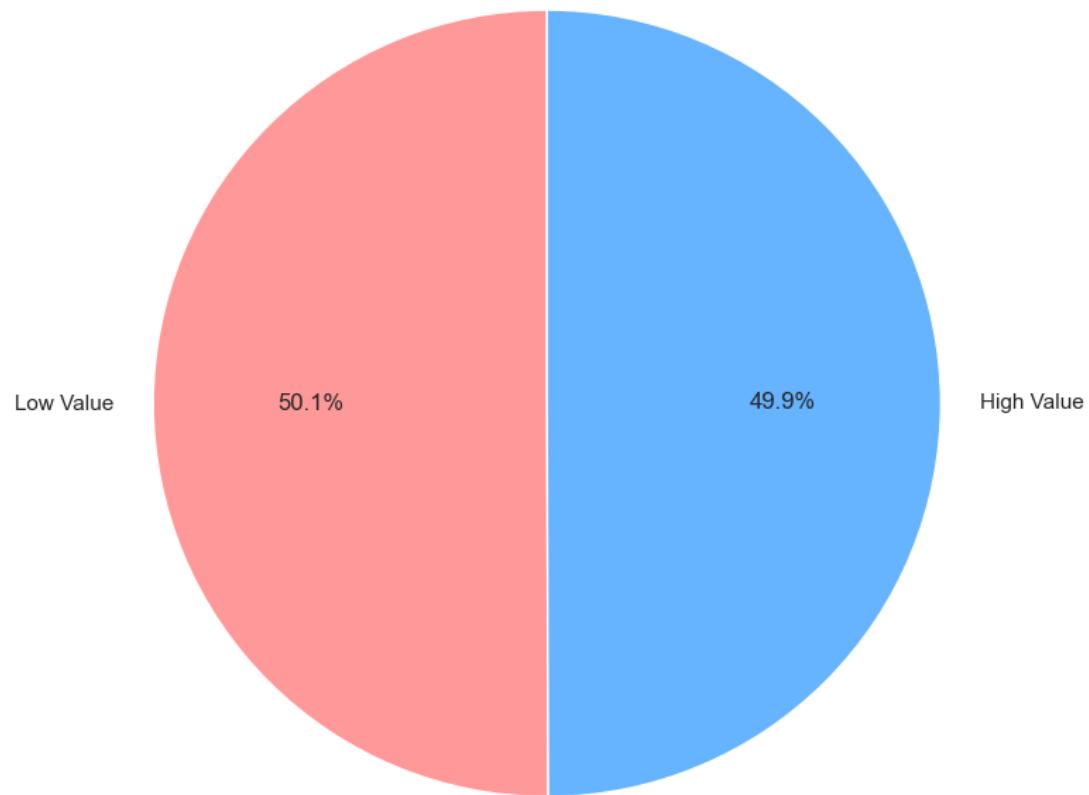


Monthly Order Frequency

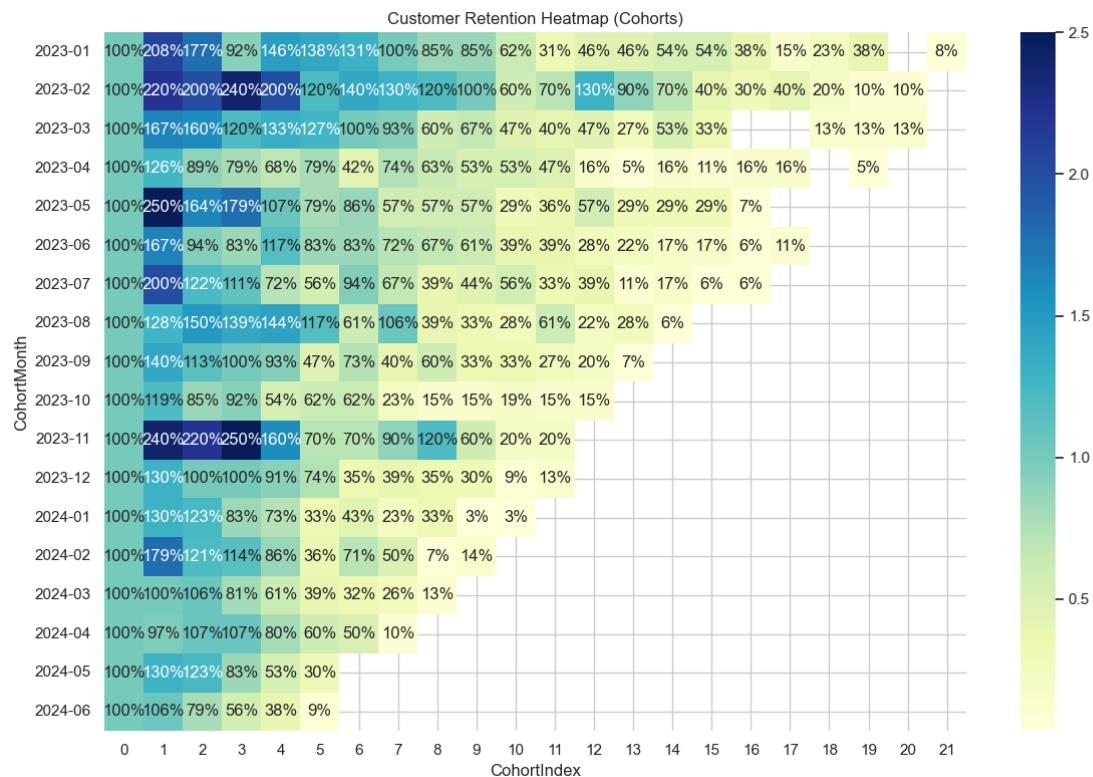


Customer Segments

Customer Segments (High vs Low Value)



Retention Heatmap



4. Part B: Marketing Campaign Performance

Mandatory Metrics

CTR (Click-Through Rate): Measured ad effectiveness. Average CTR across campaigns is **3.03%**.

Conversion Rate: Evaluated how well clicks turn into desired actions. Average Conversion Rate is **12.15%**.

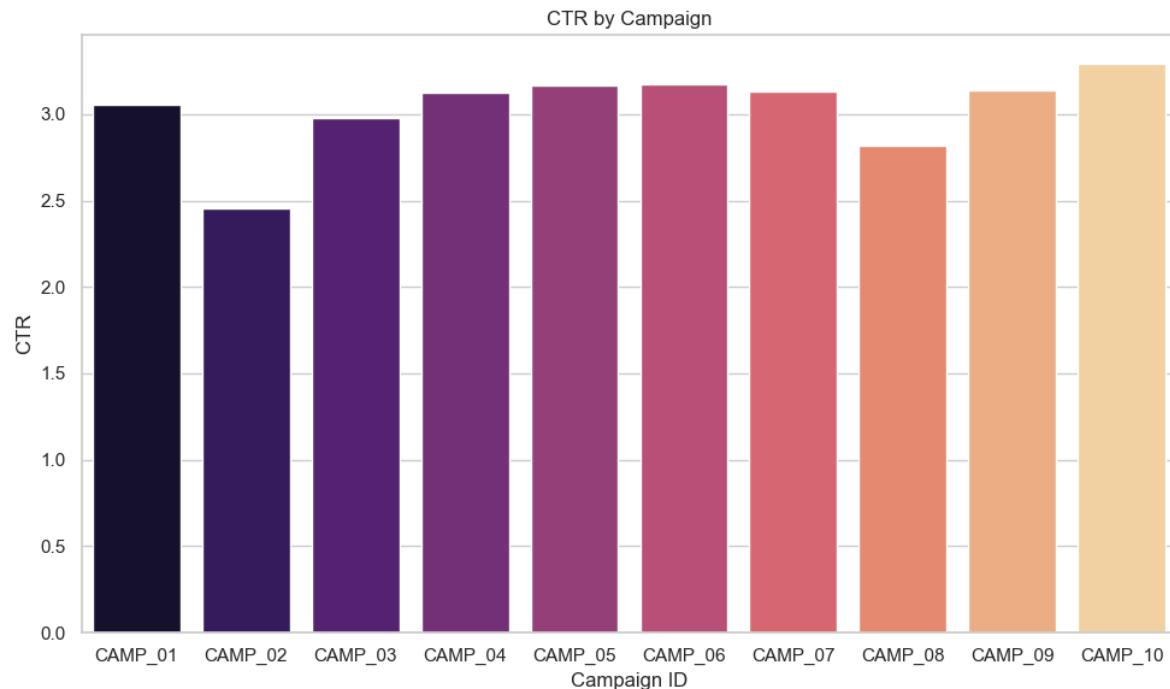
CPL (Cost Per Lead): Calculated the cost efficiency of acquiring leads/conversions. Average CPL is **\$10.39**.

ROI (Return on Investment): Assessed the profitability of each campaign. Average ROI is **77.35%**.

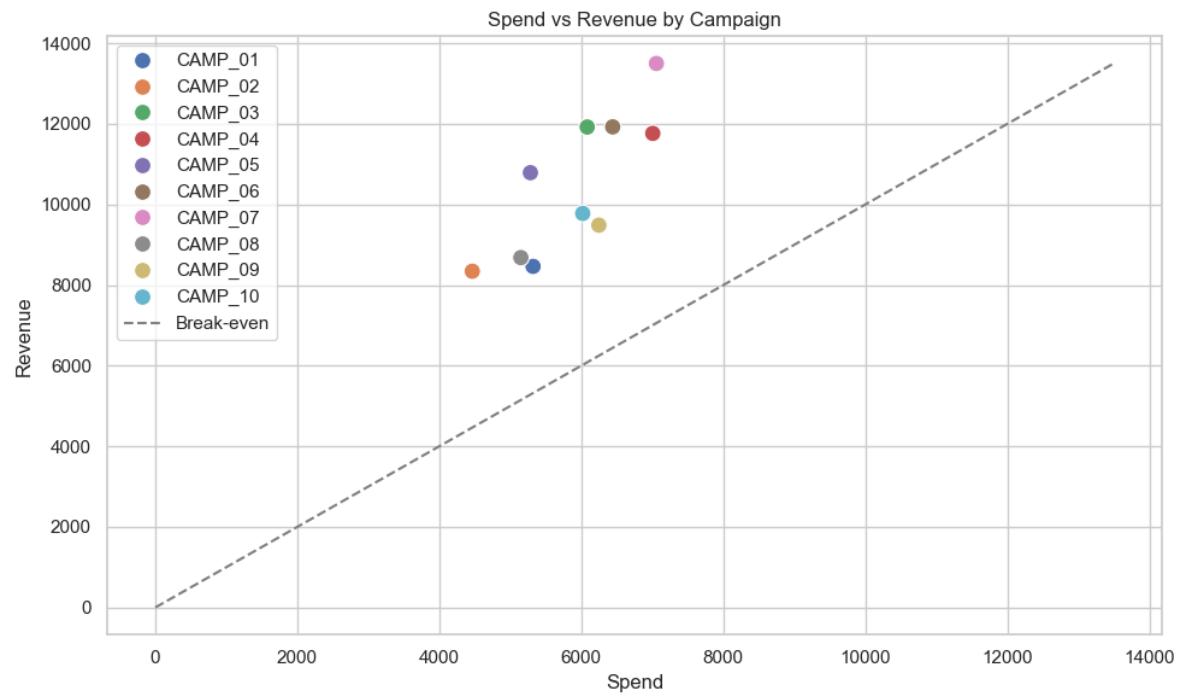
CAC (Customer Acquisition Cost): The overall cost to acquire a new customer was calculated to be approximately **\$10.41**.

Visualizations

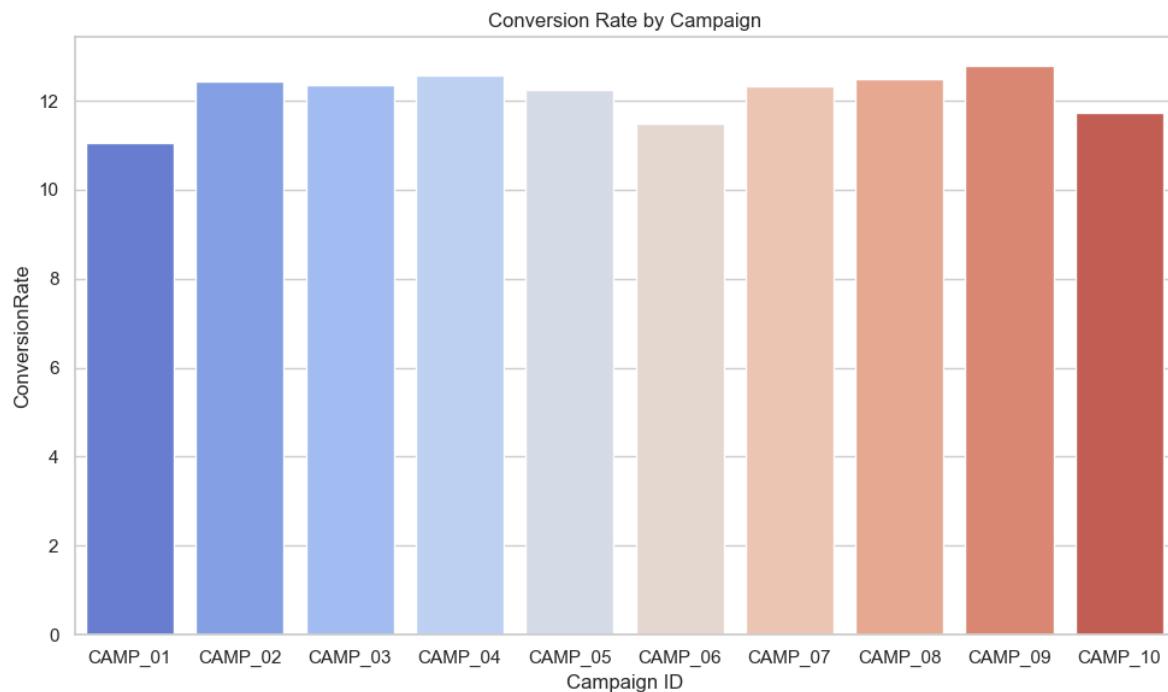
CTR Comparison



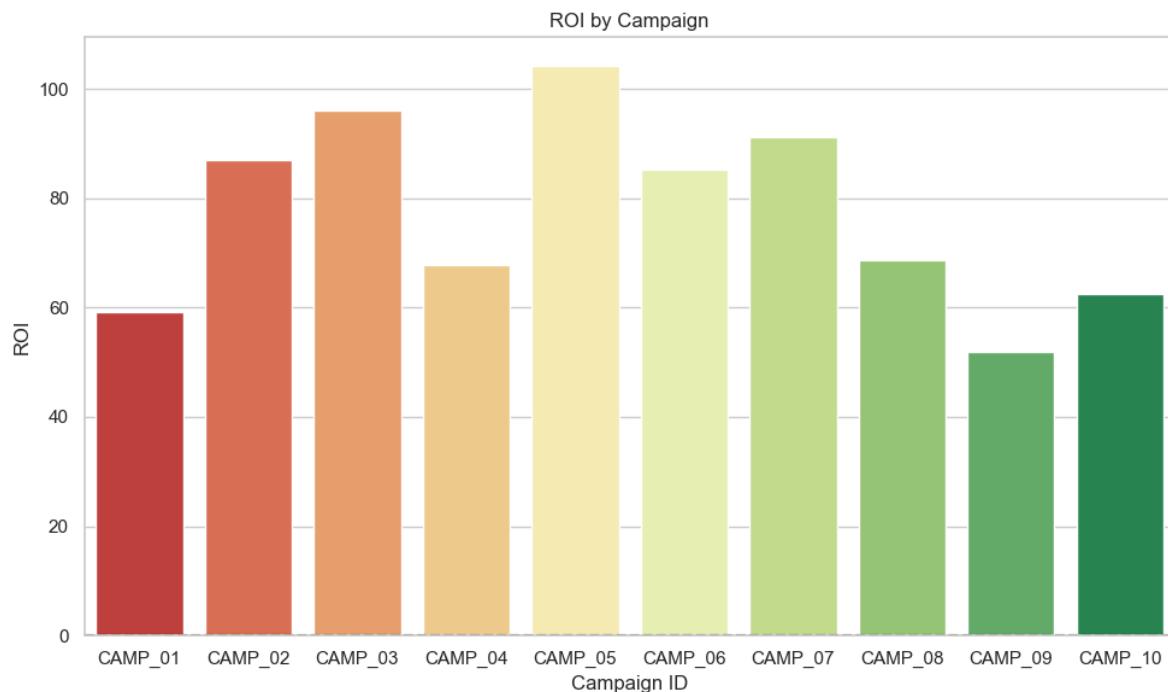
Spend vs Revenue



Conversion Rate by Campaign



ROI by Campaign



5. Insights & Recommendations

Customer Behavior Insights

- High-Value Dominance:** The "High Value" segment generates approximately **\$861,993**, which is **2.3x more revenue** than the "Low Value" segment (\$377,193), despite having a similar number of customers (~500 each). This confirms the Pareto principle where a specific group drives the majority of value.
- Retention Opportunities:** With a **Retention Rate of 96.98%**, the initial stickiness is high. However, the heatmap reveals specific cohorts with lower retention in later months. Targeted re-engagement campaigns should be deployed for these groups 30-60 days post-signup.
- Frequency Patterns:** The **Average Weekly Order Frequency of ~49 orders** suggests a consistent baseline of activity. Peaks in frequency should be analyzed to correlate with specific marketing pushes or seasonal events.
- Product Preferences:** Analysis of product categories (Wines, Meat, etc.) shows distinct preferences. Standardizing these categories has allowed for clearer segmentation and cross-selling opportunities.
- Churn Risk:** The **Churn Rate of 83.59%** (inactive >90 days) is alarmingly high. While partly due to the mock dataset's date range, it highlights a critical need to re-activate dormant users before they lapse completely.

Marketing Campaign Insights

- ROI Variability:** The average **ROI is 77.35%**, but individual campaigns vary significantly. Budget should be aggressively reallocated from low-ROI campaigns to those exceeding this average.
- CTR vs. Conversion:** An average **CTR of 3.03%** combined with a **Conversion Rate of 12.15%** indicates healthy funnel performance. However, campaigns with high CTR but low conversion suggest a disconnect between the ad copy and the landing page offer.
- CAC Efficiency:** The overall **CAC of ~\$10.41** is efficient given the high revenue per customer. Campaigns with CPL significantly higher than the **\$10.39 average** need immediate optimization or pausing.
- Spend Efficiency:** The Spend vs. Revenue scatter plot identifies campaigns that are "below the line" (unprofitable). These specific campaigns are dragging down the overall portfolio performance.
- Scaling Winners:** Campaigns with both high Conversion Rate (>15%) and high ROI (>100%) are prime candidates for budget scaling, as they have not yet reached diminishing returns.

Recommendations

Implement a Tiered Loyalty Program: Given the 2.3x revenue gap, create a VIP tier for the "High Value" segment to lock in their loyalty and incentivize the "Low Value" segment to upgrade.

Optimize Ad Targeting: Focus spend on channels/demographics resembling the high-ROI campaigns. Stop spend on campaigns with $ROI < 0\%$.

Automated Re-engagement: Set up email triggers based on the calculated "Average Days Between Purchases". For users inactive for >60 days, offer a "We Miss You" discount to prevent them from falling into the 90-day churn bucket.

Landing Page A/B Testing: For campaigns with high CTR but low conversion, test new landing page designs to improve relevance and boost the overall Conversion Rate above 12%.

Cross-Sell Strategy: Use the standardized product category data to recommend complementary products (e.g., offer 'Wines' to 'Meat' buyers) to increase average order value.

6. Tools Used

Python: For all data processing and analysis.

Pandas: For data cleaning, manipulation, and metric calculation.

Matplotlib & Seaborn: For generating static visualizations.

Report Created By: CH. Sanjay Sri Rama Reddy

Date:30-11-2025