MARKETING CAMPAIGN ANALYSIS

1. Introduction

In today's competitive retail environment, data-driven marketing plays a critical role in improving customer engagement and maximizing return on marketing investments. Retail companies run multiple marketing campaigns across different channels; however, not all customers respond equally to these campaigns. Understanding customer behavior, spending patterns, and campaign responses is essential for effective targeting and personalization.

This project focuses on analyzing historical marketing campaign data of a retail company to identify high-value customers, understand factors influencing campaign acceptance, and derive actionable insights for future marketing strategies. The analysis leverages Python for data processing and exploratory analysis, SQL for structured analytics, and Streamlit for building an interactive dashboard.

Tools & Technologies Used

- Python (Pandas, NumPy, Matplotlib, Seaborn)
- SQL (Analytical queries)
- Streamlit (Dashboard)
- VS Code (Development environment)

2. Problem Statement

The retail company has collected extensive customer-level data, including demographics, spending across product categories, purchase channels, and campaign response indicators. However, the data exists in raw form and lacks a consolidated analytical view.

Management requires:

- Identification of the most valuable and responsive customer segments
- Understanding of spending behavior across demographics
- Insights into channel usage patterns
- Detection of under-served customer groups
- Clear recommendations for improving future marketing campaigns

3. Objectives of the Project

The key objectives of this project are:

1. To clean and preprocess raw marketing data
2. To perform exploratory data analysis (EDA) to uncover patterns
3. To create meaningful derived features and customer segments
4. To analyze campaign response drivers using SQL and Python
5. To design an interactive Streamlit dashboard for stakeholders
6. To provide actionable business recommendations based on insights

4. Data Description

4.1 Data Sources

- **marketing_campaign_data.csv** – Main dataset containing customer-level information
- **marketing_data_dictionary.csv** – Metadata explaining each variable

4.2 Key Variables

Demographics

- ID, Year_Birth, Education, Marital_Status, Income
- Kidhome, Teenhome, Country

Customer Relationship

- Dt_Customer (date of enrollment)
- Recency (days since last purchase)

Spending Variables

- MntWines, MntFruits, MntMeatProducts
- MntFishProducts, MntSweetProducts, MntGoldProds

Channel Usage

- NumWebPurchases, NumCatalogPurchases
- NumStorePurchases, NumDealsPurchases
- NumWebVisitsMonth

Campaign Indicators

- AcceptedCmp1 – AcceptedCmp5
- Response (latest campaign)

5. Methodology & Approach

The project was executed in a structured, step-by-step manner as shown below:

1. Data understanding and cleaning
2. Feature engineering
3. Exploratory data analysis
4. Rule-based customer segmentation
5. SQL-based analytical querying
6. Dashboard development
7. Insight generation and recommendations

6. Data Cleaning & Feature Engineering

6.1 Data Cleaning

The following preprocessing steps were applied:

- Converted `Dt_Customer` to date format
- Handled missing values in Income using median imputation
- Removed unrealistic ages (below 18 and above 90)
- Standardized categorical fields
- Verified numeric fields for consistency

6.2 Derived Features

To enhance analytical depth, several new features were created:

Feature	Description
Age	Current year – Year_Birth
Children	Kidhome + Teenhome
Total_Spend	Sum of all spending variables
Total_Purchases	Sum of all channel purchases
Customer_Tenure_Days	Days since customer enrollment

These derived metrics helped in identifying customer value and behavior patterns.

7. Exploratory Data Analysis (EDA)

7.1 Univariate Analysis

- Age distribution showed majority customers in middle-aged groups
- Income distribution revealed presence of high-income outliers
- Total spending varied significantly across customers
- Overall campaign response rate was relatively low, indicating scope for better targeting

7.2 Bivariate & Multivariate Analysis

Key observations include:

- Customers who accepted campaigns generally had higher income
- Responders showed higher total spending compared to non-responders
- High web visits did not always translate to campaign acceptance
- Family customers preferred store-based purchases

EDA helped identify behavioral differences among customers and guided segmentation rules.

8. Rule-Based Customer Segmentation

Based on business logic and statistical thresholds, customers were segmented as follows:

Segment	Definition
High Income	Income > ₹75,000
Young Customer	Age < 30
Campaign Responder	Response = 1
High Web Engagement	NumWebVisitsMonth > 5
Family Customer	Children > 0
High Spender	Total_Spend > 90th percentile

Each customer can belong to multiple segments, allowing flexible targeting strategies.

9. SQL Analytics

The cleaned and segmented dataset was structured for SQL-based analysis. Analytical queries were written to compute:

- Overall and segment-wise campaign response rates
- Average spending by income, age band, and segment
- Channel usage patterns by high-value customers
- Identification of under-served customers (high visits, low response)

SQL aggregations and CASE statements enabled efficient KPI computation to support dashboard requirements.

10. Dashboard Development (Streamlit)

An interactive dashboard was built using Streamlit to present insights to business stakeholders.

Key Features:

- KPI cards displaying customer count, response rate, average spend
- Sidebar filters for country, education, and marital status
- Tab-based navigation for overview, segmentation, and channel analysis
- Clear charts with business-friendly labels

The dashboard allows non-technical users to explore data interactively and understand campaign performance visually.

11. Key Insights

1. High-income customers have a significantly higher campaign response rate
2. High spenders contribute a disproportionate share of revenue
3. Web channel is heavily used but under-optimized for conversion
4. Family customers prefer in-store and catalog purchases
5. Certain segments show high engagement but low response, indicating missed opportunities

12. Business Recommendations

1. Focus future campaigns on high-income and high-spender segments
2. Improve personalization for frequent website visitors to increase conversion
3. Design family-oriented offers promoted via store and catalog channels
4. Reduce marketing spend on consistently low-response segments
5. Use segmented campaigns instead of mass marketing to improve ROI
6. Monitor under-served segments and test targeted incentives

13. Project Evaluation Mapping

Evaluation Area	How Addressed
Python & EDA	Complete cleaning, feature engineering, deep EDA
SQL	Structured queries, KPIs, aggregations
Dashboard	Interactive Streamlit app with filters & KPIs
Business Insight	Clear insights and actionable recommendations
Documentation	Well-structured, reproducible report

14. Conclusion

This project successfully demonstrates how customer-level marketing data can be transformed into actionable insights using Python, SQL, and interactive visualization tools. By combining data cleaning, exploratory analysis, segmentation, and dashboarding, the solution provides a comprehensive view of customer behavior and campaign effectiveness.

The approach can be extended to predictive modeling and real-time campaign optimization, making it highly relevant for real-world marketing analytics applications.