

Marketing Campaigns – Course-End Project Write-Up

Project Title :-

- Project Name: *Marketing Campaigns — EDA & Hypothesis Testing*
- Course: *Applied Data Science with Python*
- Your Name -: Moholkar Anil Vikrant
- Date of Submission -: 14-09-2025

2. Introduction

- Explain the **Marketing Mix (4 Ps: Product, Price, Place, Promotion)**.
- State that the dataset includes customer demographics, product spend, sales channels, and campaign responses.
- Objective: *Perform exploratory data analysis and hypothesis testing to identify factors contributing to customer acquisition*

3. Problem Statement

- Restate from project brief:

We aim to explore how customer characteristics, spending behaviour, and promotional channels impact acquisition and campaign success.

4. Data Description

- Mention sources:
 - Demographics: *Birth Year, Education, Marital Status, Income (People)*
 - Product spends: *Wine, Fruits, Meat, Fish, Sweets, Gold (Product)*
 - Channels: *Web, Catalog, Store Purchases (Place)*
 - Promotions: *Responses to Campaigns, Discounts (Promotion)*

5. Data Preprocessing

- Checked import formats (`Dt_Customer` → `datetime`, `Income` → numeric).
- Cleaned categorical columns (`Education`, `Marital_Status`) for consistency.
- **Missing value imputation:**
 - Assumption: Customers with same *Education* × *Marital Status* have similar Income.
 - Filled missing Income with group mean; fallback = median.

6. Feature Engineering

- New derived variables:
 - `Age` = `Dt_Customer year` – `Year_Birth`
 - `Total_Children` = `Kidhome` + `Teenhome`
 - `Total_Spending` = sum of all product spends
 - `Total_Purchases` = sum of purchases via Web + Catalog + Store

7. Exploratory Data Analysis

- Visualizations used:
 - **Histograms** for `Income`, `Spending`, `Age` → to view distribution.
 - **Boxplots** → to detect outliers.
 - **Correlation heatmap** → to see relationships (e.g., spending vs income).
- Observations:
 - `Income` is right-skewed, with a few extreme outliers.
 - `Total Spending` is strongly correlated with `Income`.
 - Most customers are middle-aged with small household sizes.

8. Outlier Treatment

- Applied IQR method to cap extreme outliers in `Income` and `Total_Spending`.
- Result: Distributions became more normal, making hypothesis tests more reliable.

9. Hypothesis Testing

- **T-test:**
 - *Null hypothesis:* High-income and low-income groups spend the same.
 - *Result:* $p < 0.05 \rightarrow$ reject null \rightarrow higher-income customers do spend significantly more.
- **Chi-square test:**
 - Tested dependency between *Campaign Response* and *Education*.
 - $p < 0.05 \rightarrow$ reject null \rightarrow response is significantly related to education.

10. (Optional) Predictive Modeling

- Logistic Regression \rightarrow predicted campaign response.
- Features used: Income, Age, Total Spending, Total Purchases.
- Accuracy $\sim X\%$ (you'll fill in after running).
- Top positive features: [list results].

11. Insights & Recommendations

- **Target Segment:** Middle-aged, high-income, higher-spending households.
- **Channel Strategy:** Web and catalog purchases correlate with higher acquisition \rightarrow prioritize digital campaigns.
- **Promotion Personalization:** Education level and marital status affect response \rightarrow tailor promotions accordingly.
- **Spending Focus:** Wine and gold products drive majority of spending \rightarrow key categories for targeted campaigns.

12. Conclusion

- Summarize workflow: cleaning \rightarrow feature engineering \rightarrow EDA \rightarrow outlier handling \rightarrow hypothesis testing \rightarrow insights.
- Re-state the impact: *Findings help marketing teams allocate resources, improve targeting, and boost customer acquisition efficiency.*