

Customer Personality Analysis - EDA & Clustering Report

1. Dataset Introduction

- Dataset: Customer Personality Analysis (marketing_campaign.csv).
- Purpose: To understand customer behavior and perform segmentation using clustering.
- Attributes: Demographics, product purchases, promotions, complaints, and shopping channels.

2. Data Preprocessing

- Performed initial dataset checks (shape, info, unique values).
- Dropped constant columns: Z_CostContact and Z_Revenue.
- Handled missing values in Income column using imputation.
- Created new features: Age, Total_Children, Total_Spent, Customer_For.
- Removed outliers in income and spending.

3. Exploratory Data Analysis (EDA)

- Univariate analysis of demographics (age, income, marital status, education).
- Spending analysis across product categories.
- Bivariate analysis: income vs spending, purchases by channel, campaign responses.
- Used Seaborn, Matplotlib, and Plotly for visualizations (histograms, bar charts, heatmaps).

4. Feature Scaling & Clustering

- Standardized numerical features for clustering.
- Applied K-Means clustering with Elbow Method and Silhouette Score to decide cluster number.
- Performed Hierarchical Clustering and plotted dendrograms for comparison.
- Cluster profiling revealed distinct customer groups (e.g., high-income spenders, budget-conscious families, low-spending young customers).

5. Insights & Conclusions

- High-income customers spend heavily on wines and luxury products.
- Families with children prefer catalog and store purchases.
- Younger customers tend to use the web channel more frequently.
- Some clusters respond better to campaigns, guiding marketing strategies.
- Segmentation allows businesses to personalize campaigns and increase ROI.

Summary in 3 Points

- 1. Data Cleaning & Feature Engineering: Handled missing values, removed irrelevant columns, created new features.
- 2. EDA: Analyzed demographics, spending habits, and channel preferences using visualizations.
- 3. Clustering & Segmentation: Applied K-Means (with Silhouette Score for optimal K) & Hierarchical clustering, profiled clusters, and extracted actionable insights.

