

Customer Personality Analysis

Comprehensive Clustering & Segmentation Report

Generated: December 28, 2025

Project Scope: This report presents a comprehensive analysis of customer personality types using advanced clustering techniques including Hierarchical Clustering, K-Medoids Clustering, Fuzzy Logic, and Genetic Algorithm optimization.

Table of Contents

- 1. Executive Summary
- 2. Project Overview & Objectives
- 3. Dataset Information
- 4. Project Lifecycle Framework
- 5. Phase A: Business Understanding
- 6. Phase B: Data Understanding
- 7. Phase C: Data Preparation
- 8. Phase D: Modeling Implementation
- 9. Key Code Sections
- 10. Results & Insights

1. Executive Summary

This project implements a comprehensive customer segmentation analysis on the Customer Personality Dataset from Kaggle. The analysis applies four distinct clustering methodologies to identify natural customer groupings based on demographics, income levels, spending patterns, and marketing campaign responses.

Key Findings: The analysis identified distinct customer segments with high-value premium customers and family-oriented budget-conscious segments, enabling targeted marketing strategies and personalized campaigns.

2. Project Overview & Objectives

Primary Objective:

Perform customer segmentation to enable targeted marketing strategies and personalized campaign approaches by identifying natural groupings within the customer base.

Selected Analytical Methods:

- **Hierarchical Clustering:** For dendrogram-based segmentation and cluster hierarchy visualization
- **K-Medoids Clustering:** Robust centroid-based clustering with superior outlier handling
- **Fuzzy Logic Clustering:** For probabilistic segment membership and overlapping classifications
- **Genetic Algorithm:** For automated optimal cluster count determination

3. Dataset Information

Source: Kaggle - Customer Personality Analysis Dataset

Data Content:

- Demographics (Age, Income, Education, Marital Status)
- Children Information (Kidhome, Teenhome)
- Spending Patterns (Wines, Fruits, Meat, Fish, Sweets, Gold Products)
- Marketing Campaign Responses (5 Campaigns + Response)
- Purchase Channels (Web, Catalog, Store)
- Recency and Customer Tenure Metrics

4. Project Lifecycle Framework

The project follows a structured data mining lifecycle consisting of four phases: Business Understanding, Data Understanding, Data Preparation, and Modeling Implementation.

5. Phase A: Business Understanding

Strategic Business Objectives:

Customer Segmentation: Identify natural groupings within the customer base based on behavioral and demographic characteristics

Targeted Marketing: Determine which customer segments represent the most valuable targets for specific marketing campaigns

Response Prediction: Develop insights that can help predict customer responsiveness to various marketing initiatives

6. Phase B: Data Understanding

Data Exploration Procedures:

1. Initial Data Assessment

- Load and examine dataset structure and dimensions
- Display preliminary data samples (head, tail)
- Generate comprehensive dataset information (info())
- Calculate descriptive statistics (describe())

2. Data Quality Assessment

- Identify and quantify missing values across all attributes
- Assess data types and format consistency

3. Exploratory Visualization

- Distribution analysis of key demographic features (Age, Income)
- Spending pattern visualization across customer categories
- Correlation heatmap generation for attribute relationships

7. Phase C: Data Preparation

Data Preprocessing Steps:

1. Missing Value Treatment

Implement appropriate strategies for handling missing data and assess impact on analytical outcomes.

2. Feature Engineering

- Total_Spending: Summation of all product category expenditures
- Children: Combined count from Kidhome and Teenhome variables
- Premium_Products: Sum of wines and meat products
- Is_Parent: Binary indicator of having children
- Total_Accepted_Campaigns: Sum of campaign responses
- Tenure: Days since customer joined

3. Data Transformation

- Standardize numerical features using StandardScaler for equal weighting in clustering algorithms
- Encode categorical variables using OneHotEncoder for algorithm compatibility

8. Phase D: Modeling Implementation

1. Hierarchical Clustering

- **Methodology:** Agglomerative clustering with multiple linkage methods
- **Visualization:** Dendrograms (Single, Complete, Average, Ward Linkage)
- **Cluster Determination:** Analysis of dendrogram structure for optimal cluster count
- **Output:** Hierarchical customer segments with relationship structure

2. K-Medoids Clustering

- **Implementation:** Using pyclustering library with optimization
- **Advantage:** Medoid-based robustness to outliers and real cluster centers
- **Validation:** Silhouette score analysis for cluster quality metrics
- **Results:** Identified optimal k with business-interpretable clusters

3. Fuzzy Logic Clustering

- **Approach:** Fuzzy Inference System (FIS) with membership functions
- **Membership:** Probabilistic segment affiliations (0-100 score)
- **Flexibility:** Customers can belong to multiple segments with varying degrees
- **Segments:** Budget Conscious, Standard, Premium/Platinum

4. Genetic Algorithm Optimization

- **Purpose:** Automatically determine optimal number of clusters (k)
- **Methodology:** Evolutionary algorithm for cluster configuration
- **Fitness Function:** Optimization based on cohesion and separation metrics

9. Key Code Sections

9.1 Data Loading & Exploration

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv(r'marketing_campaign.csv', sep='\t')
print(df.head())
print(df.info())
print(df.describe(include='all'))
```

9.2 Feature Engineering

```
df['Total_Spending'] = df['MntWines'] + df['MntFruits'] + df['MntMeatProducts'] +
df['MntFishProducts'] + df['MntSweetProducts'] + df['MntGoldProds']

df['Children'] = df['Kidhome'] + df['Teenhome']
df['Premium_Products'] = df['MntWines'] + df['MntMeatProducts']
df['Age'] = 2025 - df['Year_Birth']
df['Is_Parent'] = (df['Children'] > 0).astype(int)
df['Total_Accepted_Campaigns'] = df['AcceptedCmp1'] + df['AcceptedCmp2'] + ... + df['Response']
```

9.3 Data Preprocessing & Scaling

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler, OneHotEncoder

numeric_cols = df.select_dtypes(include=['int64', 'float64']).columns.tolist()
categorical_cols = df.select_dtypes(include='object').columns.tolist()

preprocessor = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), numeric_cols),
        ('cat', OneHotEncoder(drop='first'), categorical_cols)
    ]
)
X = preprocessor.fit_transform(df)
```

9.4 K-Medoids Clustering with Silhouette Analysis

```
from pyclustering.cluster.kmedoids import kmedoids
from sklearn.metrics import silhouette_score

k_range = range(2, 10)
for k in k_range:
    kmedoids_instance = kmedoids(X_for_clustering.tolist(), initial_medoids)
    kmedoids_instance.process()
    clusters = kmedoids_instance.get_clusters()
    sil = silhouette_score(X_for_clustering, labels)
    print(f'k = {k} → Silhouette Score = {sil:.4f}')
```

9.5 Fuzzy Logic Clustering

```
import skfuzzy as fuzz
from skfuzzy import control as ctrl

income = ctrl.Antecedent(income_range, 'income')
spending = ctrl.Antecedent(spending_range, 'spending')
segment_score = ctrl.Consequent(segment_range, 'segment_score')

income['low'] = fuzz.trapmf(income.universe, [0, 0, 20000, 45000])
income['high'] = fuzz.trapmf(income.universe, [60000, 90000, max_income, max_income])

rule1 = ctrl.Rule(income['high'] & spending['high'], segment_score['premium'])
segment_ctrl = ctrl.ControlSystem([rule1, rule2, ...])
```

9.6 Hierarchical Clustering with Dendrograms

```
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.cluster import AgglomerativeClustering

Z_ward = linkage(X_for_clustering, method='ward', metric='euclidean')

plt.figure(figsize=(12, 8))
dendrogram(Z_ward)
plt.title('Hierarchical Clustering - Ward Linkage')
plt.show()

H1 = AgglomerativeClustering(n_clusters=4, linkage='ward')
cluster_labels = H1.fit_predict(X_for_clustering)
df['Cluster_Hierarchical'] = cluster_labels
```

10. Results & Insights

10.1 K-Medoids Clustering Results

Optimal Clusters: K=2 (identified through Silhouette Score analysis)

Cluster 1 - High-Value Premium Customers (34.4%):

- Characterized by high income (above average) and high spending on premium products
- Minimal number of children (below average)
- Ideal for luxury campaigns, VIP offers, and premium product promotions
- Expected high response rate to exclusive marketing initiatives

Cluster 2 - Family-Oriented Budget Customers (65.6%):

- Larger segment with average to below-average income
- Presence of children (higher than average)
- Price-sensitive behavior with preference for deals and discounts
- Responsive to family bundles and promotional offers

10.2 Marketing Recommendations

Premium Segment:

- Target with luxury wine and meat product campaigns
- Offer personalized VIP memberships and exclusive access
- Use sophisticated marketing channels (email, premium platforms)

Family Segment:

- Emphasize family bundles and bulk discounts
- Promote balanced products (fruits, sweets, fish)
- Utilize cost-effective channels and loyalty programs

10.3 Fuzzy Logic Segment Distribution

The Fuzzy Logic approach identified three overlapping customer personality types with probabilistic memberships, allowing for more nuanced customer profiling and transition-based marketing strategies.

11. Conclusion

This comprehensive customer personality analysis successfully implemented four advanced clustering methodologies to segment the customer base into actionable groups. The K-Medoids clustering revealed two distinct segments with clear business implications: high-value premium customers and family-oriented budget-conscious customers.

The analysis demonstrates that by leveraging these distinct customer segments, organizations can:

- Design targeted marketing campaigns with higher ROI
- Personalize customer communication based on segment characteristics
- Optimize product placement and promotional strategies
- Improve customer lifetime value through segment-specific engagement
- Reduce marketing waste by focusing resources on high-value targets

The combination of multiple clustering approaches (Hierarchical, K-Medoids, Fuzzy Logic, and Genetic Algorithm) provides robust validation of segmentation results and enables comprehensive customer personality profiling.