Customer Personality Analysis

Project Overview

Dataset Information:

- **Download Link:** https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis
- **Data Content:** Comprehensive customer data including demographics, income levels, spending patterns, marketing campaign responses, children information, and recency metrics.
- **Primary Objective:** Perform customer segmentation to enable targeted marketing strategies and personalized campaign approaches

  Selected Analytical Methods.

The project will implement and compare between Two approaches:

1. Hierarchical Clustering - For dendrogram-based segmentation
2. K-Medoids Clustering - Robust centroid-based clustering
3. Fuzzy Logic Clustering - For probabilistic segment membership
4. Genetic Algorithm - For optimal cluster count determination

Project Lifecycle Framework

Phase A: Business Understanding

Strategic Business Objectives:

1. **Customer Segmentation:** Identify natural groupings within the customer base based on behavioral and demographic characteristics
2. **Targeted Marketing:** Determine which customer segments represent the most valuable targets for specific marketing campaigns
3. **Response Prediction:** Develop insights that can help predict customer responsiveness to various marketing initiatives

Phase B: Data Understanding

Data Exploration Procedures:

1. **Initial Data Assessment**
   - Load and examine dataset structure and dimensions
   - Display preliminary data samples.
   - Generate comprehensive dataset information.
   - Calculate descriptive statistics.
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 different customer categories
   - Correlation heatmap generation to identify relationships between key attributes

Phase C: Data Preparation

Data Preprocessing Steps:

1. **Missing Value Treatment**
   - Implement appropriate strategies for handling missing data
   - Assess impact of missing values on analytical outcomes
2. **Feature Engineering**
   - Create Total_Spending feature: Summation of all product category expenditures
   - Create Children feature: Combined count from Kidhome and Teenhome variables
   - Develop additional derived features relevant to customer behavior analysis
3. **Data Transformation**
   - Standardize numerical features to ensure equal weighting in clustering algorithms
   - Apply appropriate scaling techniques (StandardScaler, MinMaxScaler) based on algorithm requirements
   - Encode categorical variables where necessary

Phase D: Modeling Implementation

1. Hierarchical Clustering

- Methodology: Agglomerative clustering approach
- Visualization: Dendrogram construction to illustrate cluster hierarchy
- Cluster Determination: Analysis of dendrogram structure to identify optimal cluster count
- Output: Hierarchical customer segments with clear relationship structure

2. K-Medoids Clustering

- Implementation: Using sklearn library
- Comparative Analysis: Contrast with traditional K-means results
- Advantage Utilization: Leverage medoid-based robustness to outliers
- Validation: Silhouette analysis and cluster quality metrics

3. Fuzzy Logic Clustering

- Approach: Fuzzy Logic implementation
- Membership Probability: Assign probabilistic segment affiliations
- Flexibility: Allow customers to belong to multiple segments with varying degrees
- Interpretation: Analyze overlapping segment characteristics

4. Genetic Algorithm Optimization

- Purpose: Automatically determine optimal number of clusters (k)
- Methodology: Evolutionary algorithm searching for optimal cluster configuration
- Comparison: Validate results against traditional Elbow method findings
- Fitness Function: Optimization based on cluster cohesion and separation metrics

Submission:

- A written report explaining the analysis steps and findings (in English Not Arabic) [Printed in Discussions].
- An Python script file containing all codes and results.