Customer Segmentation Analysis

Project Overview

This project aims to segment e-commerce customers based on their purchase history and browsing behavior to optimize marketing strategies. Using the Online Retail II dataset, the project applies clustering techniques to identify distinct groups of customers, providing insights into their buying patterns and preferences.

Dataset

The dataset used is the Online Retail II data set from the UCI Machine Learning Repository, which contains all the transactions occurring for a UK-based online retail between 01/12/2009 and 09/12/2011.

Steps and Analysis Performed:

Step 1: Data Preprocessing

- Data cleaning: Handling missing values, removing duplicates, and filtering out cancellation transactions.

- Feature engineering: Added calculated fields such as TotalPrice, TotalOrders, and DaysSinceLastPurchase.

Step 2: Exploratory Data Analysis (EDA)

- Visualized the distribution of various features such as total spend, total orders, and days since the last purchase.

- Identified outliers and analyzed the correlation matrix to understand the relationships between different features.

Step 3: Customer Segmentation

- Applied K-means clustering with an optimal number of clusters determined using the Elbow method.

- Evaluated clusters using silhouette scores and analyzed cluster centroids.

Step 4: Cluster Profiling and Interpretation

- Detailed analysis of each cluster to identify and interpret customer segments.

- Provided insights into the behavior and characteristics of each segment.

Step 5: Visualization

- Used PCA for dimensionality reduction to visualize the customer segments.

- Created scatter plots to show the distribution of clusters.

Key Findings

- **Cluster 0**: High-value customers who make purchases frequently.

- **Cluster 1**: Customers with moderate spending but high order counts.

- **Cluster 2**: Infrequent shoppers with lower total spend.