

Customer Segmentation using K-Means Clustering

FlexiSAF Generative AI & Data Science Internship Project

Overview

This project is part of the FlexiSAF Generative AI and Data Science Internship and focuses on customer segmentation using K-Means clustering. The goal is to categorize customers based on their purchase behavior, helping businesses tailor marketing strategies and improve customer experience.

Dataset

The dataset is sourced from Kaggle: Customer Purchasing Behaviors. It includes features such as:

- Annual Income
- Purchase Amount
- Loyalty Score
- Purchase Frequency

Key Features

- K-Means Clustering: Implemented to group customers into clusters based on similar characteristics.
- Data Preprocessing: Includes handling missing values, scaling numerical features, and cleaning the dataset.
- Cluster Visualization: Visualized the clusters and marked the centroids to understand the center of each group.

Cluster Interpretations

After applying K-Means clustering, we obtained three distinct clusters:

- Cluster 0: Represents low-income customers who also have low spending habits and loyalty scores.
- Cluster 1: Indicates high-income customers who are more loyal and spend more frequently.
- Cluster 2: Represents a balanced group that does not strongly lean towards high or low spending.

The Silhouette Score for the clustering was 0.56, indicating a moderately well-defined clustering structure.

Pain Points Addressed

- Generic marketing approach -> Enables personalized campaigns.

Customer Segmentation using K-Means Clustering

FlexiSAF Generative AI & Data Science Internship Project

- Poor resource targeting -> Helps focus on high-value customers.
- Lack of insight into customer behavior -> Provides clear patterns based on income, loyalty, and frequency.
- Difficulty retaining customers -> Identifies loyal vs. at-risk segments.
- One-size-fits-all marketing strategy -> Enables tailored engagement.

Key Takeaways

- K-Means is an effective method for grouping customers with similar characteristics.
- Visualizing the clusters and their centroids helps to understand customer behavior patterns.
- The analysis can be used for customer segmentation in marketing, allowing for more personalized campaigns.