

Customer Segmentation Using K-Means Clustering

Objective

The goal of this project is to segment customers of a shopping mall into distinct groups based on their annual income and spending scores. By identifying these segments, we aim to tailor marketing strategies and improve customer satisfaction.

Data Collection

- Dataset Description: The dataset used for this project contains information about customers of a shopping mall.
- Columns:
- CustomerID: Unique identifier for each customer.
- Gender: Gender of the customer.
- Age: Age of the customer.
- Annual Income (k\$): Annual income of the customer in thousands of dollars.
- Spending Score (1-100): A score assigned by the mall based on customer behavior and spending nature.

Data Preprocessing

Steps:

- Loading Data: Read the dataset into a DataFrame using Pandas.
- Exploratory Data Analysis (EDA):
 - Check for missing values and data types.
 - Generate summary statistics.
 - Visualize data distribution using histograms and boxplots.
- Data Cleaning: Handle missing values, if any, and ensure data consistency.

Data Visualization

- Purpose: To understand the distribution of data and relationships between features.
- Visualizations:
- Histograms: Show the distribution of Annual Income and Spending Score.
- Scatter Plots: Visualize the relationship between Annual Income and Spending Score.

K-Means Clustering Algorithm

- K-Means is an unsupervised machine learning algorithm used to partition the data into k clusters.

Steps:

- Initialize: Randomly select k centroids.
- Assignment: Assign each data point to the nearest centroid.
- Update: Calculate new centroids as the mean of the assigned points.
- Repeat: Repeat the assignment and update steps until convergence.

Implementation:

- Import the KMeans class from scikit-learn.
- Fit the model to the data.
- Predict the cluster labels for each data point.

Cluster Interpretation

Clusters Identified:

- Cluster 1 (Green): High income, low spending.
- Cluster 2 (Red): Average income, average spending.
- Cluster 3 (Yellow): High income, high spending.
- Cluster 4 (Blue): Low income, low spending.
- Cluster 5 (Purple): Low income, high spending.

Insights:

- Cluster 3 (Yellow): Represents the most valuable customers.
- Cluster 4 (Blue): Includes customers with the lowest income and spending.
- Each cluster has distinct characteristics that can guide marketing strategies.

Strategic Recommendations

Tailored Marketing Campaigns:

- Cluster 3 (Yellow): Offer exclusive deals and loyalty programs.
- Cluster 5 (Purple): Introduce affordable luxury items and promotions.

Enhancing Customer Experience:

- Cluster 2 (Red): Improve the shopping experience and introduce mid-range product lines.

Discounts and Promotions:

- Cluster 4 (Blue): Focus on budget-friendly products, discounts, and promotions.

Engaging High-Income, Low-Spending Customers (Cluster 1):

- Understand their needs and offer relevant products or services to increase engagement.

Conclusion

- By segmenting customers into distinct groups, the shopping mall can implement targeted marketing strategies, optimize product offerings, and enhance overall customer satisfaction.
- This project provides a data-driven approach to understanding customer behavior and improving business outcomes.