Customer Segmentation Project

Overview

This project aims to perform customer segmentation using K-Means clustering. It involves analyzing a dataset of mall customers to identify distinct customer segments based on their annual income and spending score. This README provides an overview of the project, including problem statement, dataset description, data preprocessing, analysis techniques, findings, and recommendations.

Problem Statement

The primary objective of this project is to segment mall customers into distinct groups based on their annual income and spending score. This segmentation will help the mall understand its customer base better and tailor marketing and customer engagement strategies accordingly.

Phases of Development

The project consists of the following phases:

1. Data Collection: The dataset of mall customers is obtained.

2. Data Preprocessing: Data is cleaned and numeric features are standardized.

3. Feature Engineering: Relevant features for clustering are selected.

4. Clustering (K-Means): K-Means clustering is applied to create customer segments.

5. Visualization: The clusters are visualized for interpretation.

6. Interpretation: Characteristics of each cluster are analyzed.

7. Documentation & Submission: The project is documented and prepared for submission.

Dataset Description

The dataset used in this project is "Mall_Customers.csv." It contains information about mall customers, including age, gender, annual income, and spending score.

Data Preprocessing

- Numeric columns are selected for standardization.

- Data is standardized using the StandardScaler.

- Relevant features for clustering are selected (Annual Income and Spending Score).

Analysis Techniques

- K-Means clustering is employed to segment the customers into distinct clusters.

- Cluster characteristics are analyzed using descriptive statistics.

Findings & Recommendations

- Key findings and insights from the customer segments are presented.

- Recommendations for marketing and customer engagement strategies are provided based on the segments.