K-Means Clustering

Problem: Customer Segmentation for an E-commerce Platform

Objective:

Segment customers based on their purchasing behavior using the K-Means clustering algorithm. The goal is to identify distinct groups of customers to tailor marketing strategies and improve customer engagement.

Dataset:

Use the Online Retail dataset from the UCI Machine Learning Repository. This dataset contains transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

Steps:

- Data Preparation:
  - Load the dataset and perform basic preprocessing to clean the data.
  - Create Recency, Frequency, and Monetary (RFM) metrics for each customer.
  -
- Standardize the Data:
  - Standardize the RFM metrics using `StandardScaler` to ensure each metric contributes equally to the clustering process.
  -
- Apply K-Means Clustering:
  - Apply the K-Means algorithm to the standardized data to segment customers into distinct clusters.
  - Use the elbow method to determine the optimal number of clusters.
  -
- Evaluate the Clustering Performance:
  - Calculate the silhouette score to evaluate the quality of the clusters.
  -
- Visualize the Clusters:
  - Visualize the clusters using appropriate plots (e.g., pair plots) to understand the characteristics of each customer segment.