Unsupervised Learning Project

Problem: Customer Segmentation for an E-commerce Platform

Objective:

Segment customers based on their purchasing behavior using unsupervised 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.

Part 1:

- 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.

Part 2:

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.

Hierarchical Clustering

- Apply hierarchical clustering and evaluate the results using the silhouette score.

Principal Component Analysis (PCA):

- Apply PCA to reduce the dimensionality of the data.
- Analyze the explained variance ratio.

Visualization and Comparison:

- Visualize the clusters for K-Means and Hierarchical clustering in the PCA-reduced space.
- Compare the silhouette scores of K-Means and Hierarchical clustering.
- Create a comparison table and bar plot to visualize the results.
- Determine the best-performing algorithm based on silhouette scores.