Week 3: Unsupervised Learning and Feature Engineering

Task 3.1: Clustering with K-Means

Objective: Apply K-means clustering to segment customer data and analyze patterns without predefined labels.

Dataset: Mall Customer Segmentation Data from Kaggle. This dataset includes information like customer ID, gender, age, annual income, and spending score, which are used to segment customers based on shopping behaviour.

- **Dataset URL**: https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python

Activities:

1. **Data Exploration**:
    - Load the dataset and conduct basic exploratory data analysis to understand the features.
2. **Clustering Implementation**:
    - Use the K-means algorithm from scikit-learn to find clusters among customers.
    - Determine the optimal number of clusters using the elbow method.
3. **Analysis of Clusters**:
    - Analyze and profile the characteristics of each customer cluster.
    - Visualize the clusters using scatter plots by choosing appropriate feature pairs.

Expected Output:

- A Jupyter notebook that performs K-means clustering and includes both the elbow method visualization and cluster profiling.

Documentation:

- Document the complete process, the decision for the number of clusters, and insights from cluster analysis.