

Task 3 Write-up

The goal of this project was to perform **customer segmentation** using unsupervised learning techniques on the *Mall Customers* dataset. This involved identifying distinct groups of customers based on their behavioral and demographic attributes — namely **Age**, **Annual Income (k\$)**, and **Spending Score (1–100)** — to help businesses understand different customer types and strategize targeted marketing.

Step-by-Step Workflow:

1. Data Loading & Exploration:

The dataset was loaded and briefly explored to understand the nature of the data. No missing values were found, making it suitable for direct preprocessing.

2. Preprocessing & Feature Scaling:

Relevant features were selected and standardized using `StandardScaler` to normalize the range of values. This ensures each feature contributes equally to the clustering process.

3. Finding Optimal Clusters (Elbow Method):

The **Elbow Method** was used to determine the optimal number of clusters (k) by plotting the **Within-Cluster Sum of Squares (WCSS)** for different values of k . The elbow point visually indicated the most appropriate number of clusters (usually $k = 5$ for this dataset).

4. Model Training with K-Means:

A **KMeans** clustering model was trained using the optimal k . The algorithm grouped customers into distinct clusters based on similarities in their income and spending habits.

5. Dimensionality Reduction & Visualization:

To visualize the high-dimensional clustering results, **Principal Component Analysis (PCA)** was applied to reduce features to two dimensions. A scatter plot of clusters made it easy to interpret and distinguish between the customer groups.

Outcome:

The clustering revealed **distinct customer personas**, such as:

- ♦ **High-income low spenders**
- ♦ **Low-income high spenders**
- ♦ **Young high spenders**
- ♦ **Middle-aged moderate spenders**
- ♦ **Older low spenders**

These insights enable marketers and business analysts to:

- ♦ Personalize promotions for different customer groups
- ♦ Allocate budgets efficiently
- ♦ Improve customer retention strategies
- ♦ Identify high-value customers