

**Green University of Bangladesh**

**Department of Computer Science and Engineering (CSE)**

**Faculty of Sciences and Engineering**

**Semester: (Summer, Year:2025), B.Sc. in CSE (Day)**

**Lab Report NO #8**

**Course Title: Data Mining Lab**

**Course Code: CSE-436 Section:213D4**

**Lab Experiment Name:** Clustering Using Hierarchical and Density-Based Algorithms on Real-World Customer Data

**Student Details**

|  |  |  |
| --- | --- | --- |
| **Name** | | **ID** |
| **1.** | Pankaj Mahanto | 213902002 |

**Submission Date : 24/04/2025**

**Course Teacher’s Name : Md. Jahid Tanvir**

|  |
| --- |
| **Lab Report Status**  **Marks: ………………………………… Signature:.....................**  **Comments:.............................................. Date:..............................** |

**1. TITLE OF THE LAB REPORT EXPERIMENT**

* Clustering Using Hierarchical and Density-Based Algorithms on Real-World Customer Data

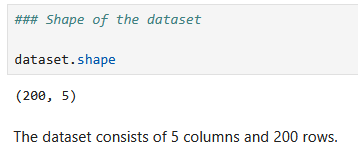
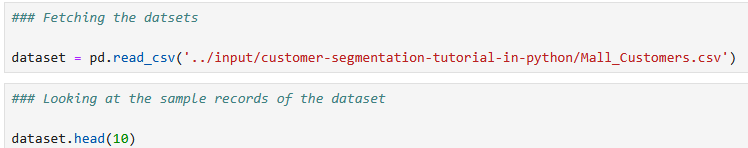
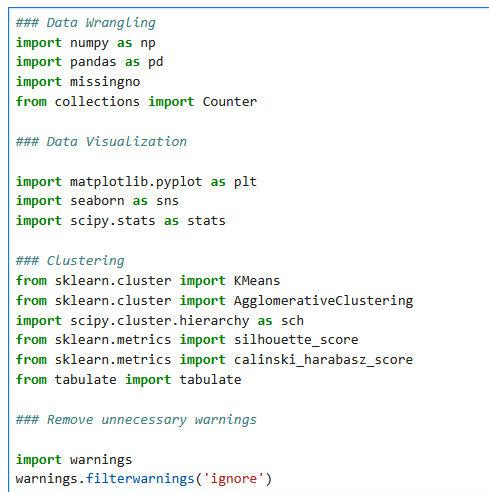
**2. OBJECTIVES/AIM [2 marks]**

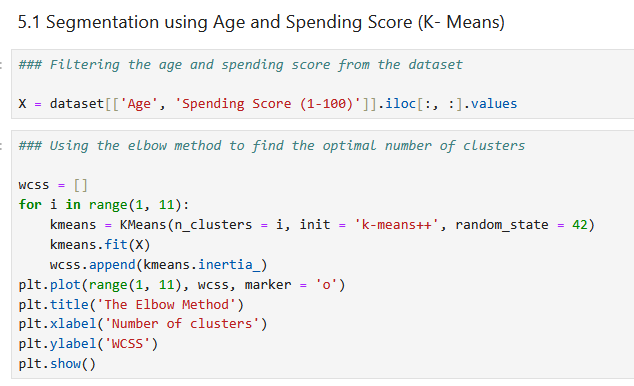
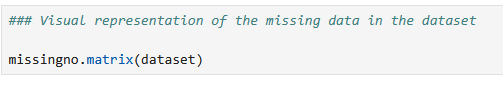
* To implement and understand clustering algorithms that **do not require pre-specifying the number of clusters**.
* To apply **Hierarchical Clustering** and **Density-Based Spatial Clustering (DBSCAN)** on real-world customer data.
* To **analyze the behavior** of clustering results and compare different algorithms on the same dataset.

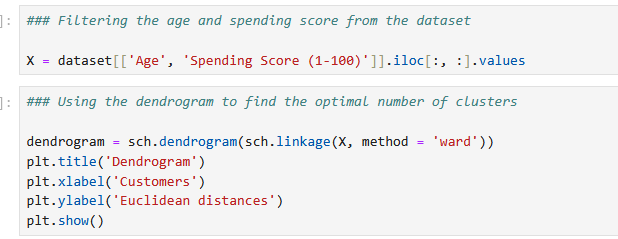
**PROCEDURE / ANALYSIS / DESIGN [3 marks]**

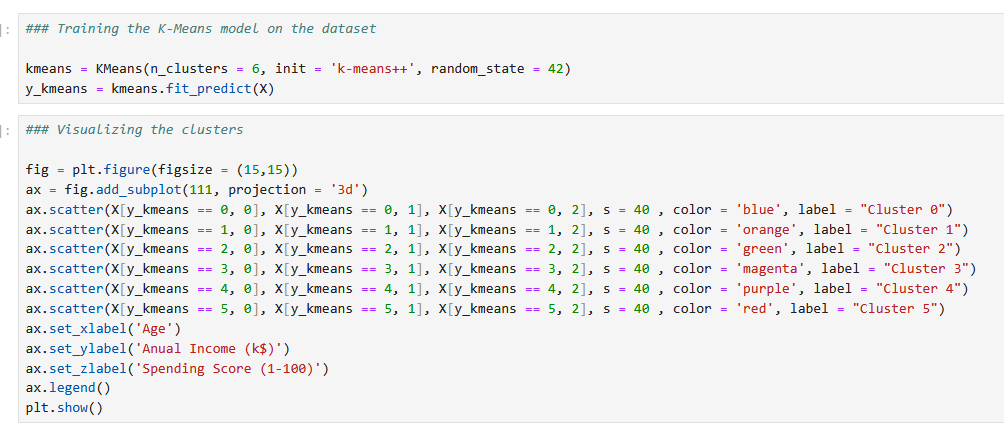
* Step 1: Load the Mall Customer Segmentation dataset which includes features like Age, Annual Income, and Spending Score.
* Step 2: Perform standardization on the dataset using StandardScaler to bring all features onto a similar scale.
* Step 3:
* For Hierarchical Clustering, use ward linkage to compute distances and plot the dendrogram.
* For DBSCAN, use density-based parameters like eps=0.8 and min\_samples=5 to find clusters automatically.
* Step 4: Visualize the clustering results using Principal Component Analysis (PCA) to reduce data into 2D.
* Step 5: Compare and analyze clustering results between the two approaches.

**4. IMPLEMENTATION [3 marks]**

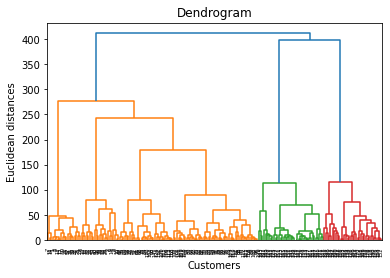
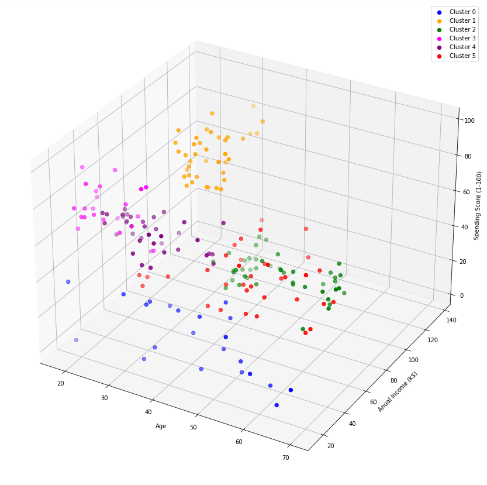


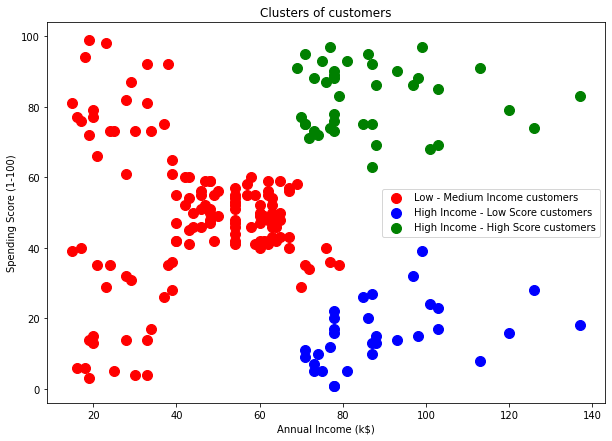
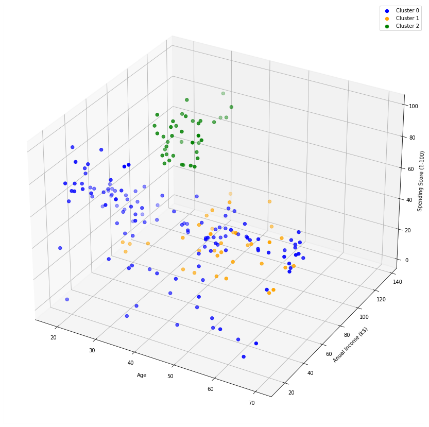






**5. TEST RESULT / OUTPUT [3 marks]**





**6. ANALYSIS AND DISCUSSION [3 marks]**

The code explores how real-world data can be grouped using clustering algorithms that don't require us to know the number of clusters beforehand. Hierarchical clustering builds a tree of clusters and is useful when we want to explore different groupings. Its dendrogram gives a clear picture of how clusters are formed step by step. On the other hand, DBSCAN is powerful for finding irregularly shaped clusters and identifying noise points that don’t belong anywhere. Since both algorithms treat data differently, they reveal unique insights. This is useful in practical scenarios where we don’t always know how many groups we should expect. Overall, the project highlights the flexibility and power of unsupervised learning in extracting patterns from complex datasets.

**7. SUMMARY**

This experiment , I implemented two unsupervised learning algorithms — Hierarchical Clustering and DBSCAN — on the real-life Iris dataset. After preprocessing the data, we used visualizations like dendrograms and 3D plots to understand the structure and outcome of the clustering. Both algorithms worked well, but DBSCAN also detected potential outliers.