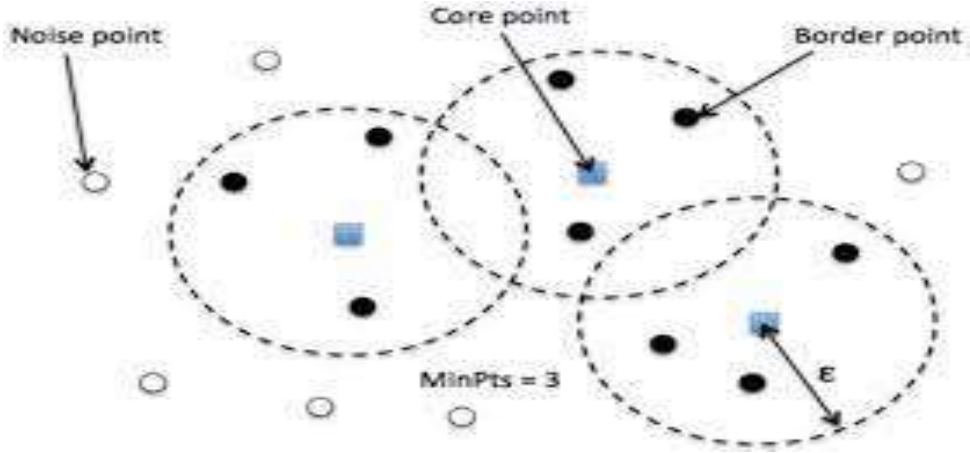
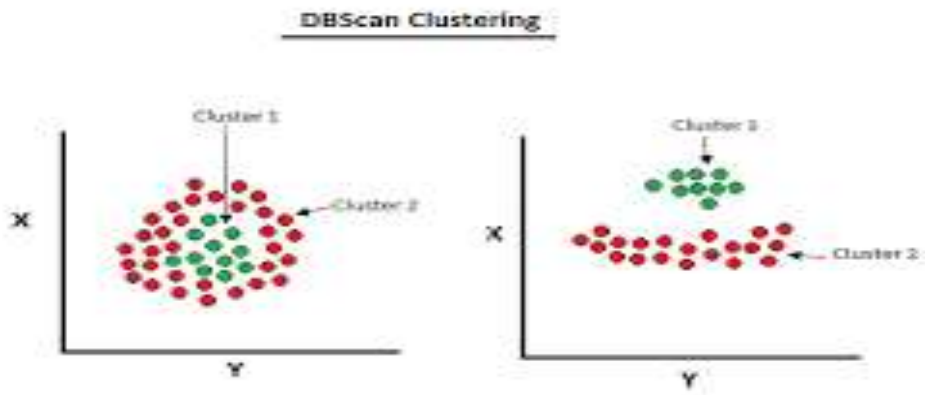
DBSCAN - Density-Based Spatial Clustering

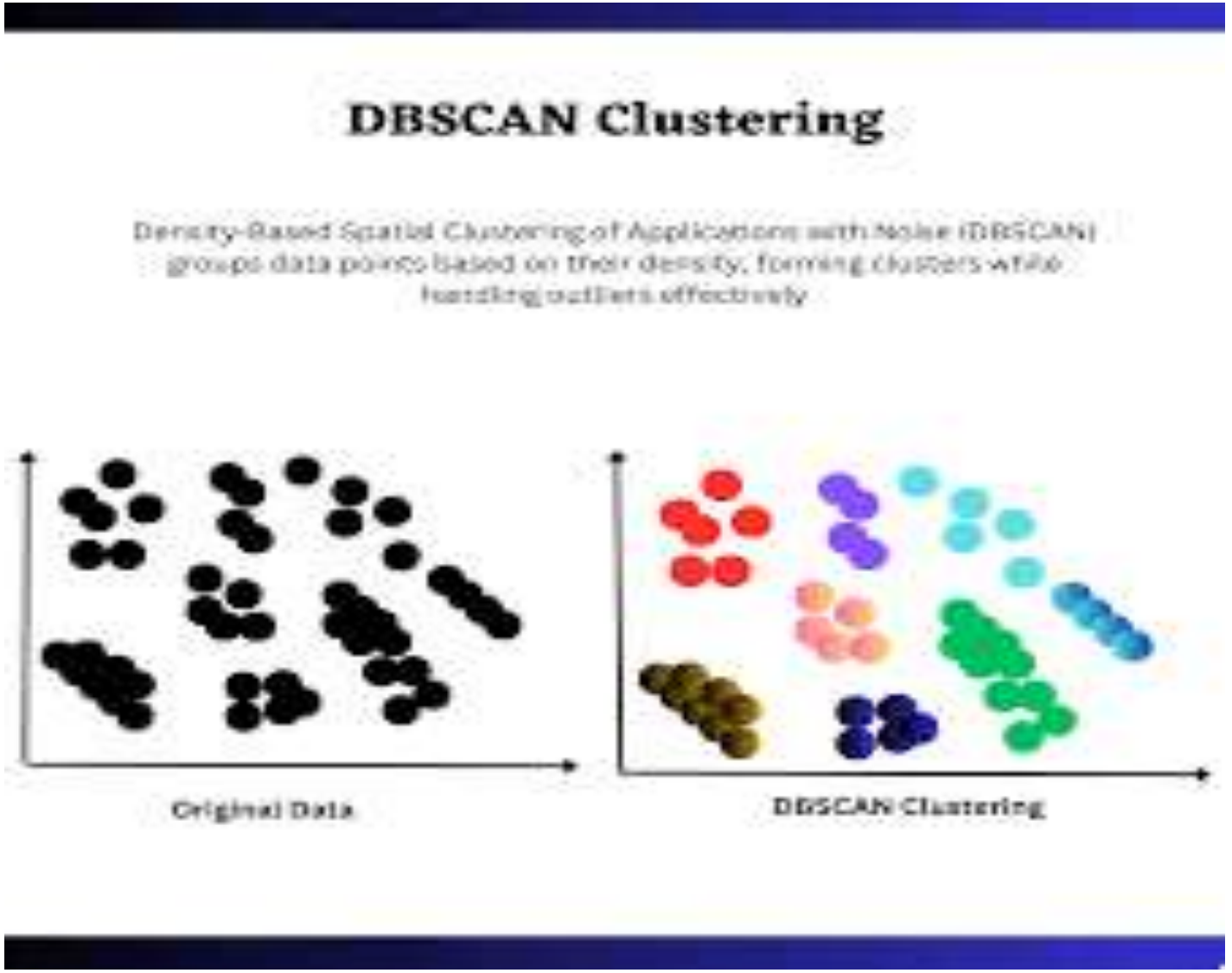
- **DBSCAN Clustering:**
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm that groups together closely packed points while marking points in low-density regions as outliers or noise. It doesn't require specifying the number of clusters beforehand and is capable of identifying clusters of arbitrary shapes and sizes.

- **Importing Libraries and Loading Data:**
- **import numpy as np**
- **import matplotlib.pyplot as plt**
- **import pandas as pd dataset = pd.read_csv('Mall_Customers.csv')**
- **X = dataset.iloc[:, [3, 4]].values**

DBSCAN - Density-Based Spatial Clustering







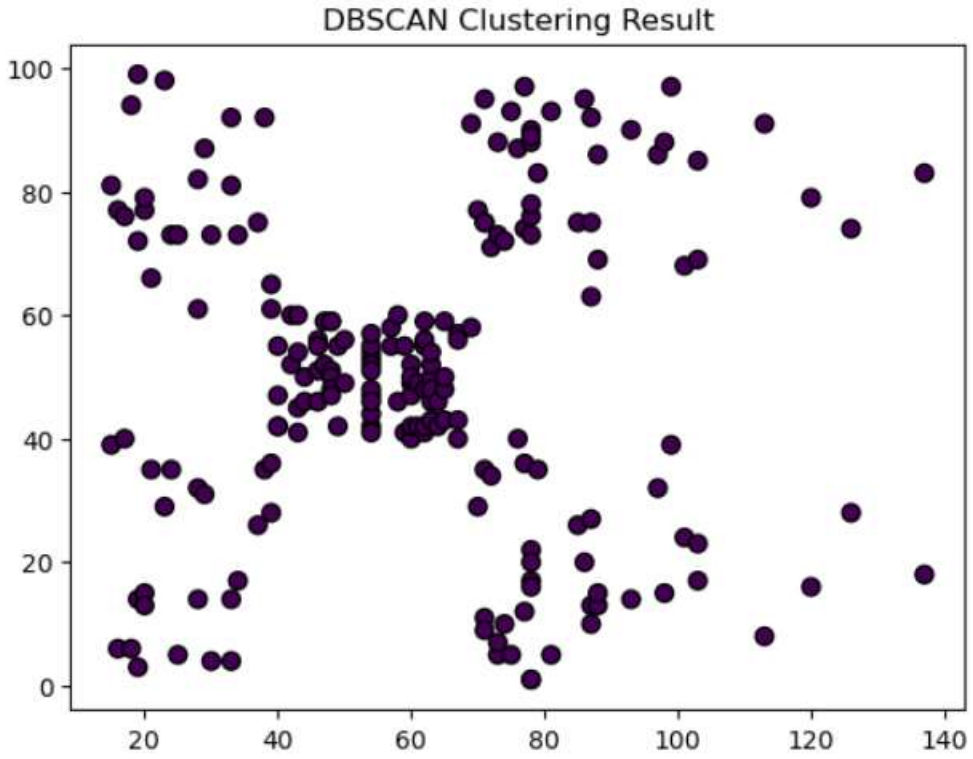
DBSCAN - Density-Based Spatial Clustering

- This part of the code imports necessary libraries and loads the dataset from the CSV file 'Mall_Customers.csv'. It selects columns 3 and 4 (usually representing features) from the dataset for clustering.

- **Applying DBSCAN:**
- **from sklearn.cluster import DBSCAN**
- **dbscan = DBSCAN(eps=0.3, min_samples=5)**
- **labels = dbscan.fit_predict(X)**
- Here, the DBSCAN algorithm is imported from scikit-learn, and an instance of DBSCAN is created with parameters eps (maximum distance between two samples for one to be considered as in the neighborhood of the other) and min_samples (minimum number of samples in a neighborhood for a point to be considered a core point).
- The fit_predict() method is called to fit the model to the data and predict cluster labels**.**

- plt.scatter(X[:, 0], X[:, 1], c=labels, cmap='viridis', marker='o', s=50, edgecolor='k’)
- plt.title('DBSCAN Clustering Result')
- plt.show()

DBSCAN - Density-Based Spatial Clustering



DBSCAN - HDBSCAN

- This part of the code visualizes the clustering result by creating a scatter plot of the data points. Each point is colored according to its assigned cluster label.

- **Difference Between DBSCAN and HDBSCAN:**
- Both DBSCAN and HDBSCAN are density-based clustering algorithms, but HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is an extension of DBSCAN that adds a hierarchical approach to clustering. Here's how they differ:

- **DBSCAN:**
- Requires setting parameters like eps and min_samples.
- Clusters are determined based on density, and points are labeled as noise if they do not belong to any cluster.
- The number of clusters is not pre-defined.
- Sensitive to the choice of parameters.

- **HDBSCAN:**
- Automatically determines the number of clusters without specifying parameters like eps and min_samples.
- Utilizes a hierarchical clustering approach to identify clusters at different levels of density.
- More robust to varying densities and shapes of clusters.
- Produces a hierarchical cluster structure.

DBSCAN - HDBSCAN

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) are similar in that they are both density-based clustering algorithms used for clustering datasets based on the density distribution of data points**.**
- **DBSCAN:** DBSCAN uses a single-level clustering approach. It assigns each point to a cluster, a border point, or noise based on its density and the specified parameters such as epsilon (eps) and minimum samples.
- **HDBSCAN:** HDBSCAN employs a hierarchical clustering approach. It builds a hierarchy of clusters using a minimum spanning tree and then selects clusters based on stability criteria. This hierarchical approach allows HDBSCAN to identify clusters of varying densities and shapes.

- **DBSCAN:** The number of clusters in DBSCAN is not predetermined and depends on the data distribution and parameters specified.
- **HDBSCAN:** HDBSCAN automatically determines the number of clusters in the data without requiring the user to specify any parameters related to the number of clusters. It identifies clusters at different levels of density, resulting in a more flexible clustering solution.
- Handling Noise**:**

DBSCAN - HDBSCAN