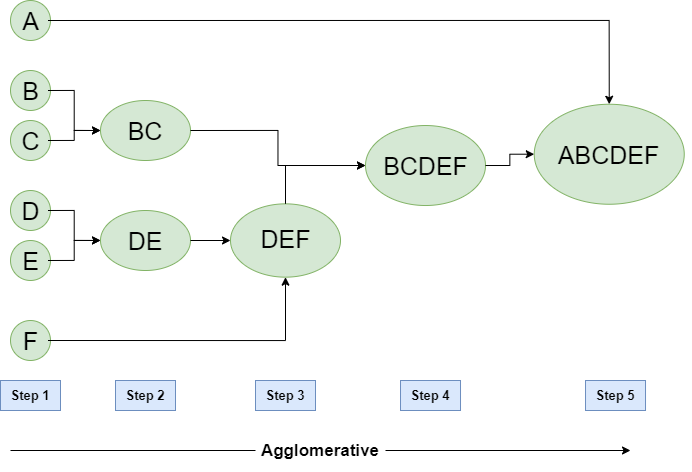
**NAME OF EXPERIMENT**: -Implementation of Agglomerative Hierarchical clustering algorithm using Python.

**INTRODUCTION:** -Hierarchical clustering is a clustering algorithm that group’s similar data points together based on their distance or similarity. It is a type of unsupervised machine learning technique that aims to find patterns or structures in the data without the need for predefined labels or categories.

Agglomerative clustering is a type of hierarchical clustering algorithm which starts with treating each data point as an individual cluster and then merges the clusters that are most similar to each other based on some distance metric. The process continues until all data points are in a single cluster, forming a dendrogram.

Agglomerative clustering is a bottom-up approach to clustering, and it has the advantage of being able to handle a large number of data points. However, it can be computationally expensive for large datasets and sensitive to the choice of distance metric and linkage method.

Algorithm: -

1. Compute the distance matrix between the input data points.
2. Let each data points to be a cluster.
3. Repeat
4. Merge the two clusters.
5. Update the distance matrix.
6. Until only k cluster remains.

MANUAL CALCULATION:

* Perform the agglomerative hierarchy clustering algorithm based on the following data points.

|  |  |  |
| --- | --- | --- |
| Sample No | X | Y |
| A1 | 0.40 | 0.53 |
| A2 | 0.22 | 0.38 |
| A3 | 0.35 | 0.32 |

* Solution,

Step1: Calculate Euclidean distance between each data and every point.

Euclidean distance =

D (A1, A2) = = 0.23

D (A1, A3) = = 0.22

D (A1, A2) = = 0.14

|  |  |  |  |
| --- | --- | --- | --- |
|  | A1 | A2 | A3 |
| A1 | 0 |  |  |
| A2 | 0.23 | 0 |  |
| A3 | 0.22 | 0.14 | 0 |

Here, 0.14 is the smallest in above table. So, merge (A2, A3).

Step2: Calculate Euclidean distance of merged data.

|  |  |  |
| --- | --- | --- |
|  | A1 | A2, A3 |
| A1 | 0 |  |
| A2, A3 | 0.23 | 0 |

Hence, the new cluster group = A1, (A2, A3).

Step3: Construction of dendrogram.

A1 (A2, A3)

A1 A2 A3

ADVANTAGES OF AGGLOMERATIVE CLUSTERING ALGORITHM: -

1. Does not require the number of clusters to be specified in advance, allowing for more flexibility in the analysis
2. Produces a dendrogram that can be used to visually inspect the clustering structure and identify natural groups or outliers
3. Can be used with various distance metrics and linkage methods to fit different types of data.
4. Can handle a large number of data points

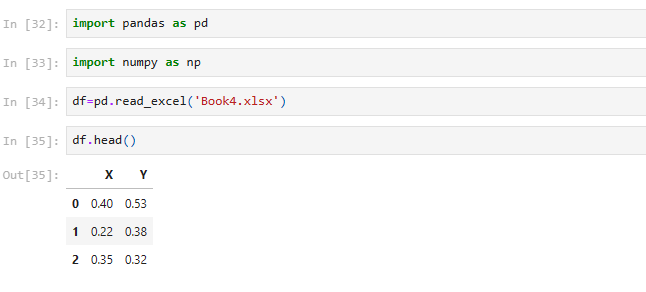
DISADVANTAGES OF AGGLOMERATIVE CLUSTERING ALGORITHM: -

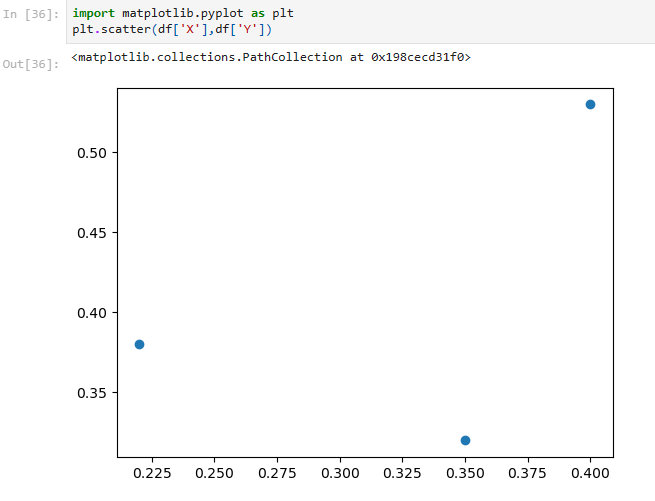
1. It is affected by outliers or noise which can cause suboptimal merging.
2. It has high time and space complexity, making it inefficient for large data sets.
3. Can be computationally expensive, especially for large datasets
4. Does not work well with non-Euclidean distance measures, such as categorical data or binary data.

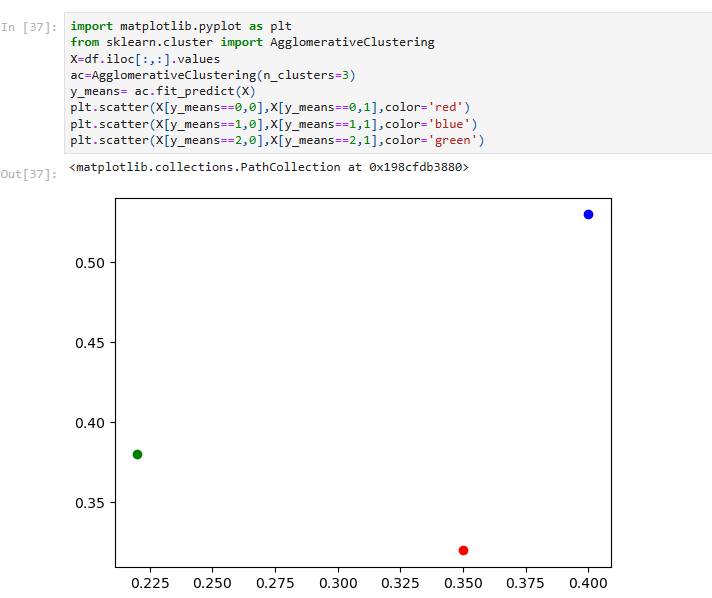
PROGRAM IMPLEMENTATION IN PYTHON: -

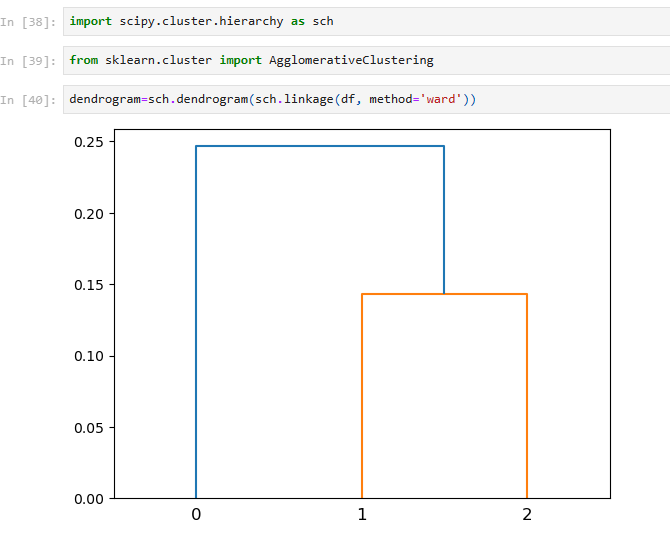
Requirement = Anaconda Navigator

Source Code:









CONCLUSION: -

Hence, Agglomerative clustering algorithm is successful implemented using Python.