

ACADEMIC YEAR 2020 – 2021

DATA MINING LEARNING COMPONENT

**Implementation Agglomerative Hierarchical Clustering**

*Submitted by:*

***KEERTHAN S 1NT18IS128***

***U R NARASIMHA RANGAN 1NT18IS176***

*Submitted to:*

### Dr. K. Aditya Shastry

Associate Professor

Department of Information Science and Engineering Nitte Meenakshi

Institute of Technology Bangalore-064

# Logo Description automatically generated

# DEPARTMENT OF INFORMATION SCIRNCE AND ENGINEEERING NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

**Autonomous institute accredidited by AICTE, affiliated to VTU-Belgaum Yelahanka Bangalore – 560 064**

# Table of content

1. Introduction
2. Agglomerative hierarchical clustering
3. Dendrogram
4. Implementation

* Dataset
* Code
* outputs

**INTRODUCTION**

Starting from what happens in clustering, a group of different data objects is classified as similar objects. One group means a cluster of data. Data sets are divided into different groups in the cluster analysis, which is based on the similarity of the data. After the classification of data into various groups, a label is assigned to the group. It helps in adapting to the changes by doing the classification.

Cluster Analysis in Data Mining means that to find out the group of objects which are similar to each other in the group but are different from the object in other groups.

There are many uses of Data clustering analysis such as image processing, [data analysis](https://www.upgrad.com/blog/exploratory-data-analysis-and-its-importance-to-your-business/), pattern recognition, market research and many more. Using Data clustering, companies can discover new groups in the database of customers. Classification of data can also be done based on patterns of purchasing.

In this report we are focused on hierarchical clustering methods, more

Specific on agglomerative hierarchical clustering.

### **Hierarchical Clustering Methods:**

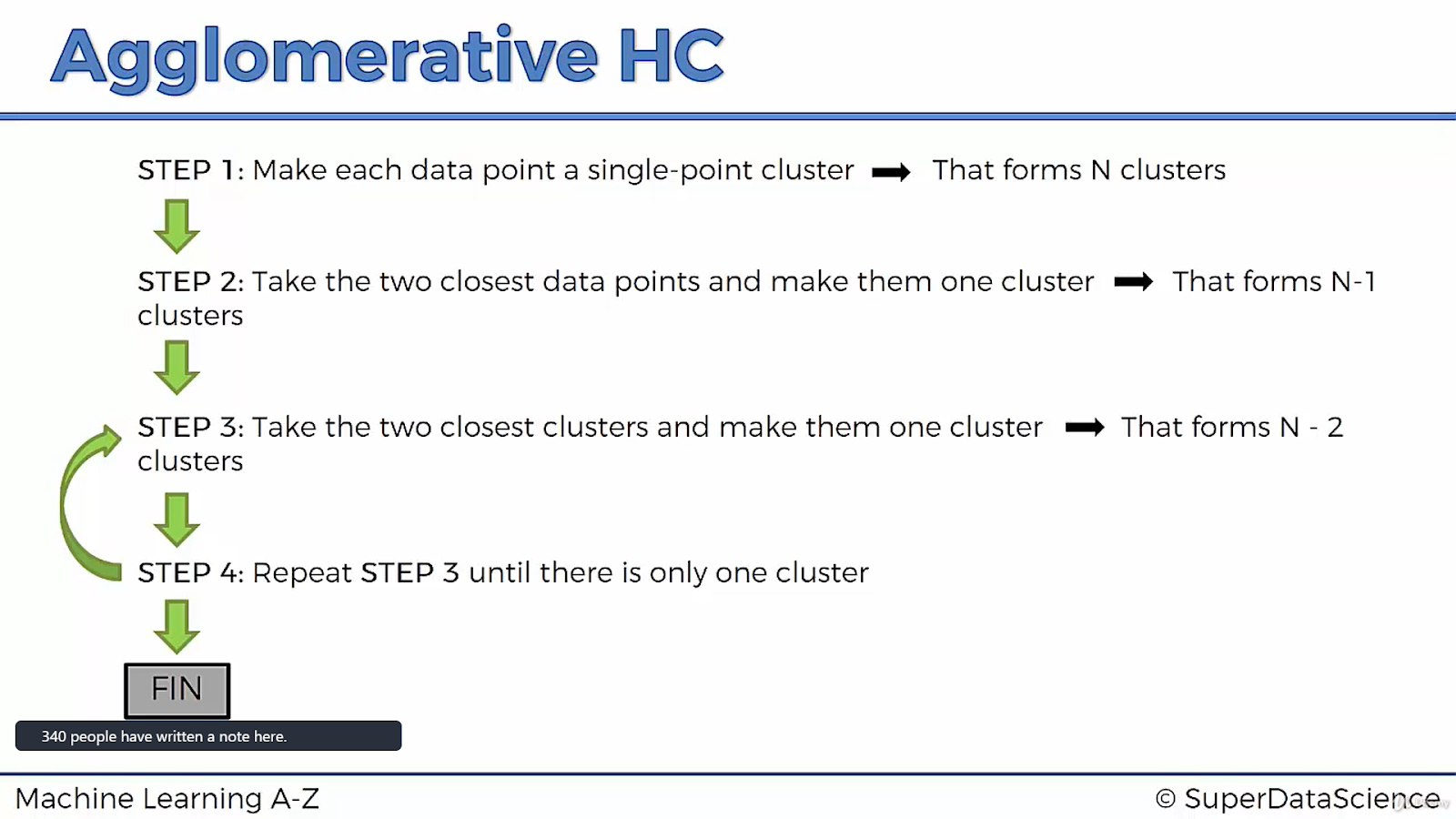
In this hierarchical clustering method, the given set of an object of data is created into a kind of hierarchical decomposition. The formation of hierarchical decomposition will decide the purposes of classification. There are two types of approaches for the creation of hierarchical decomposition,

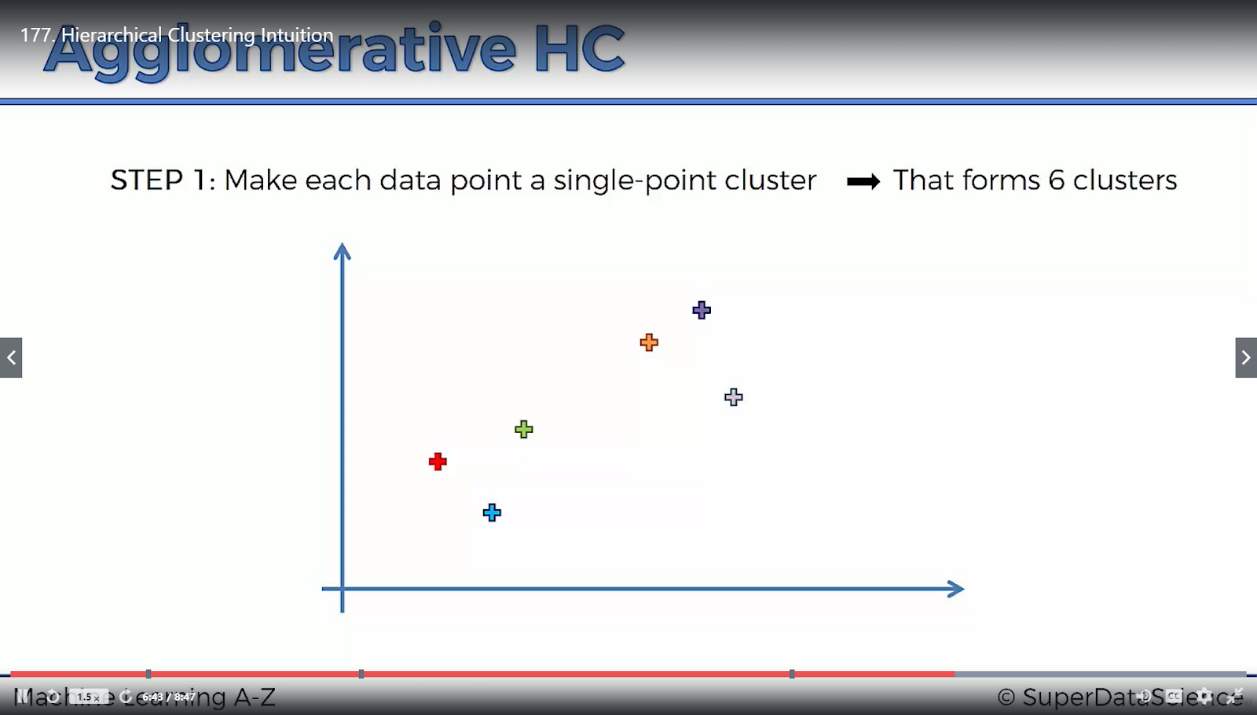
1. **Divisive Approach**
2. **Agglomerative Approach**

Agglomerative Hierarchical clustering(AHC)

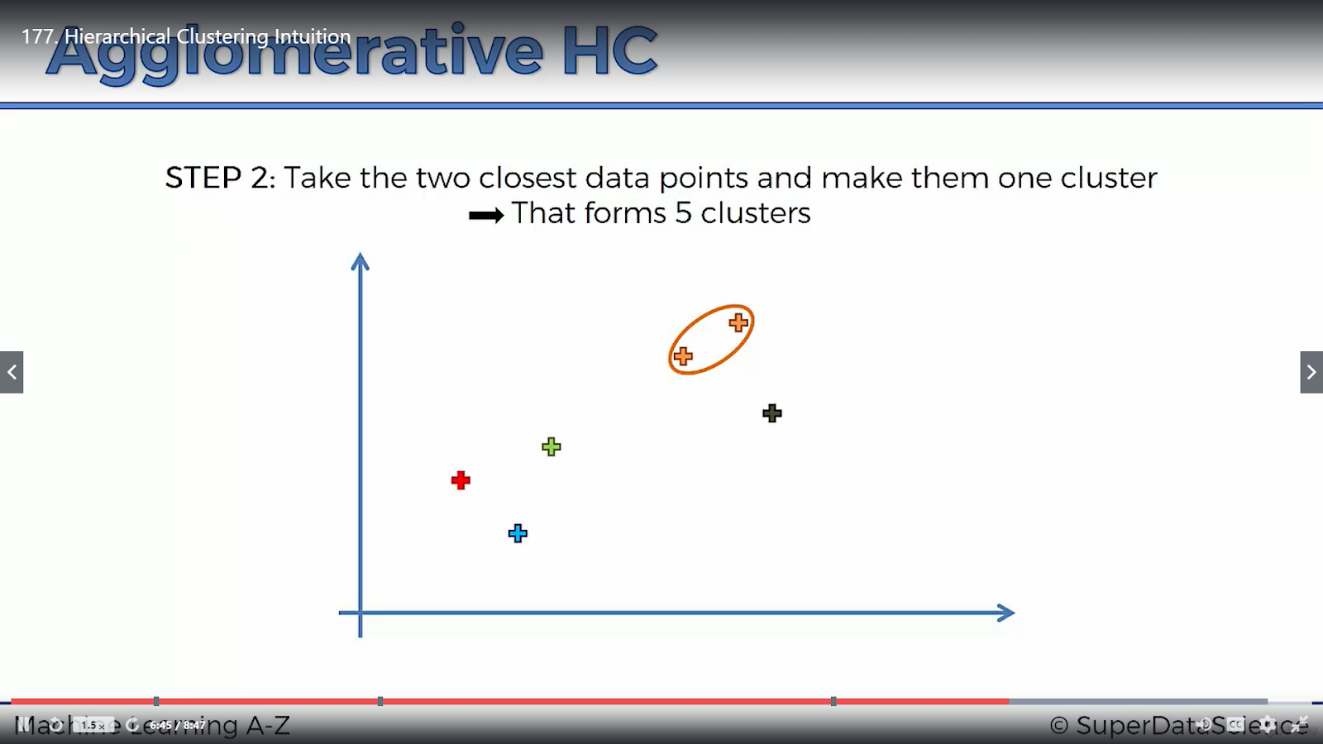
The **agglomerative clustering** is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. It’s also known as AGNES (Agglomerative Nesting). The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named dendrogram.

gglomerative clustering works in a “bottom-up” manner. That is, each object is initially considered as a single-element cluster (leaf). At each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster (nodes). This procedure is iterated until all points are member of just one single big cluster



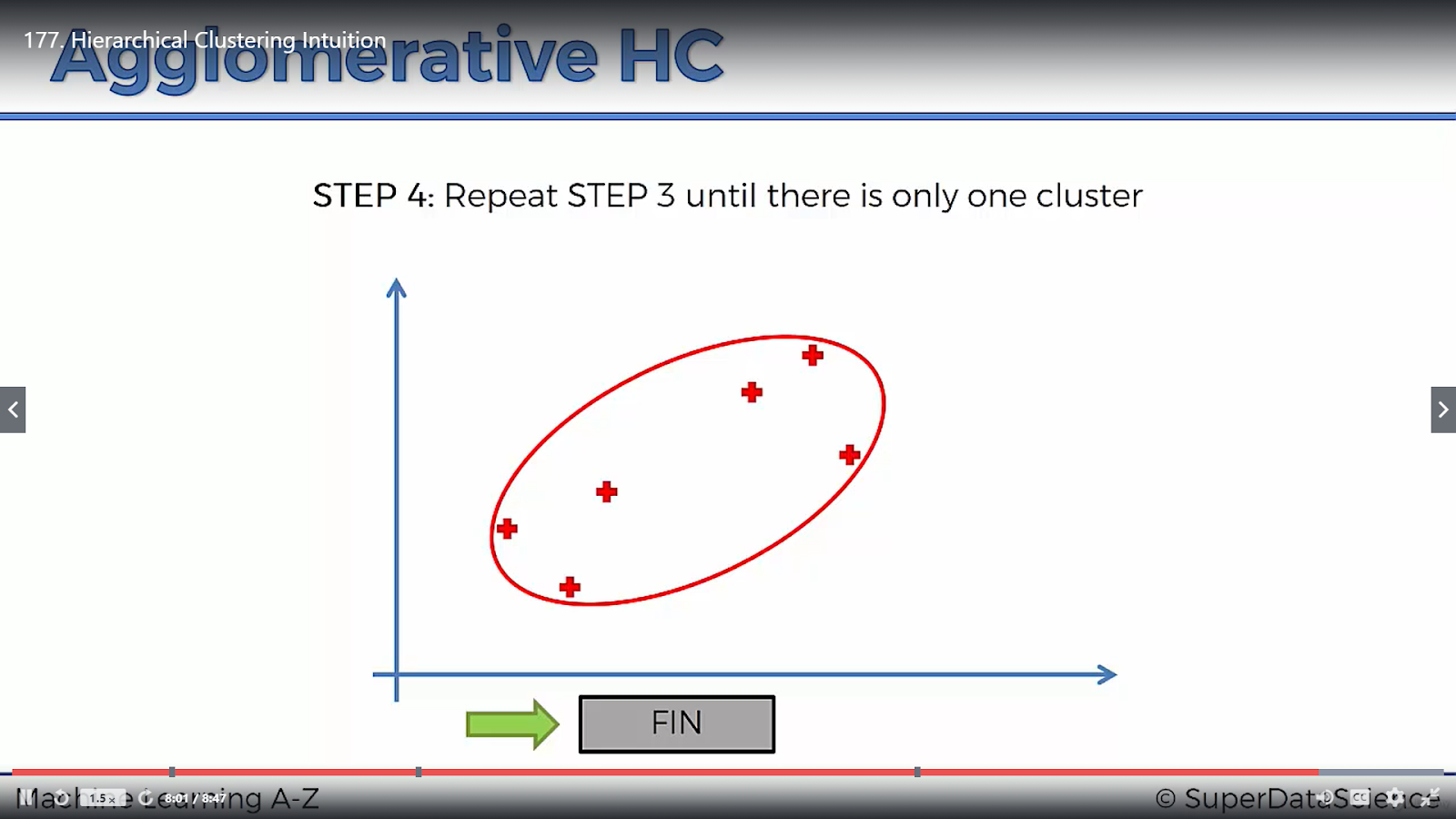


6 clusters initially



5 clusters after one iteration

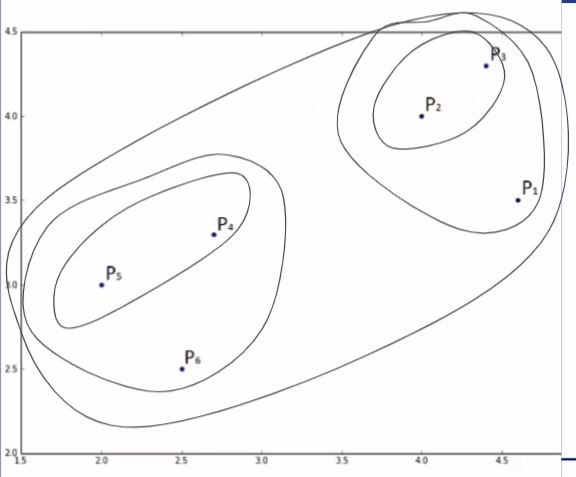
After N-1 iterations



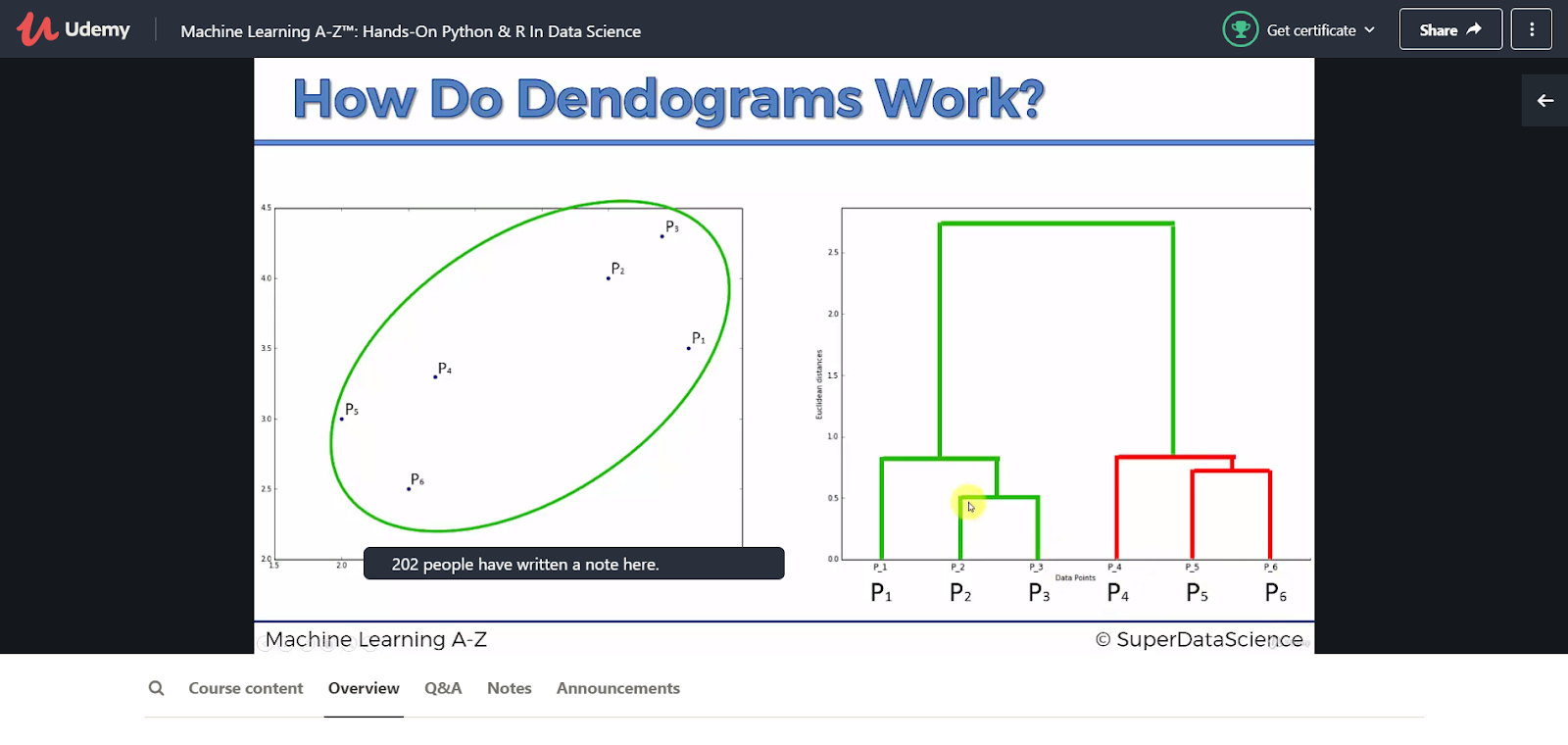
Final cluster

Dendrogram

The dendrogram is a visual representation of the compound correlation data. The individual compounds are arranged along the bottom of the dendrogram and referred to as leaf nodes. Compound clusters are formed by joining individual compounds or existing compound clusters with the join point referred to as a node.

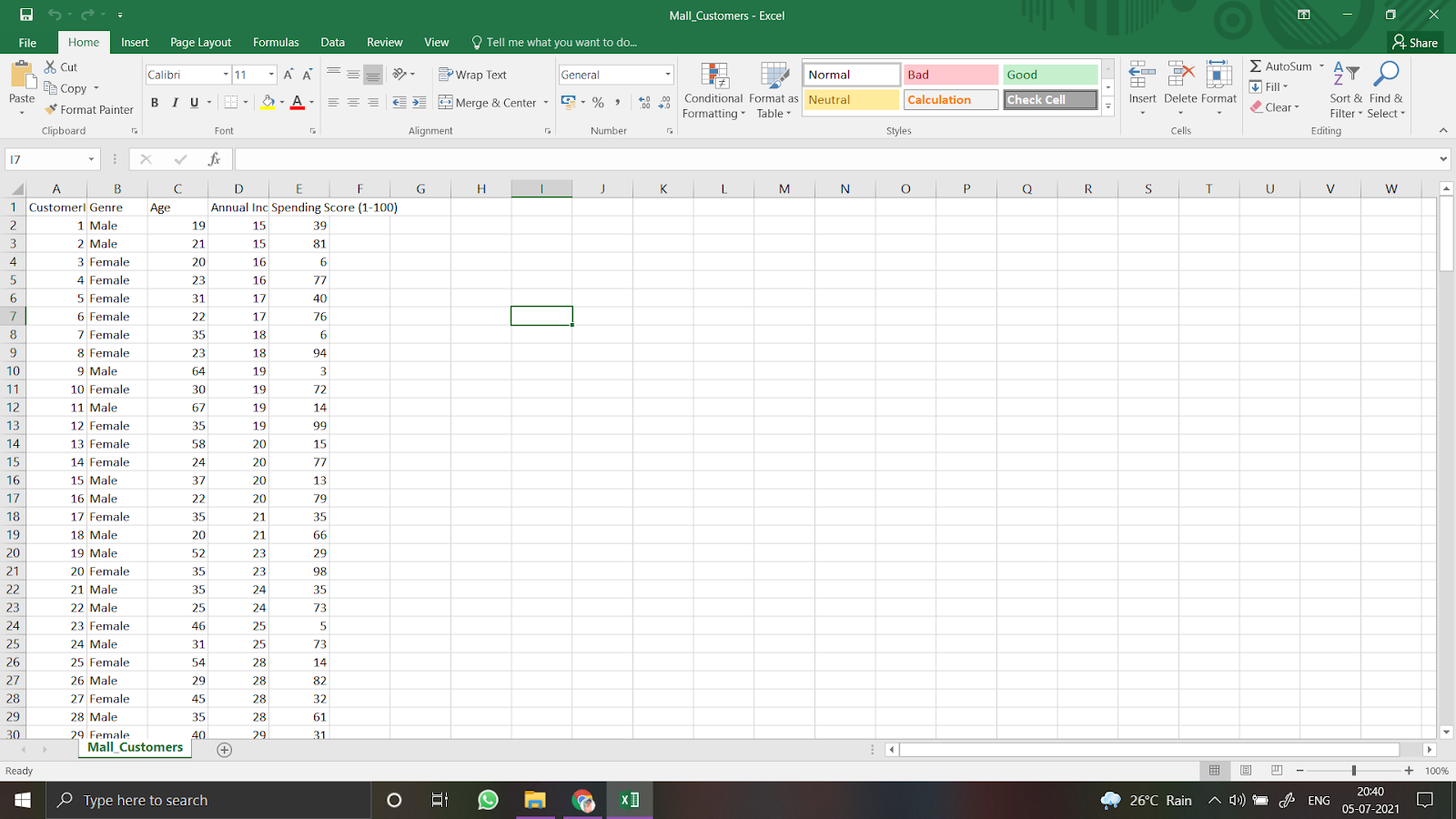


Above figure is the final cluster which is formed, next page shows us the dendrogram of the above figure



**IMPLEMENTATION**

dataset used:



**CODE AND OUTPUT**

