

Data Mining Project

Data Link : <https://www.kaggle.com/uciml/iris>

Data Description :

The Iris dataset was used in R.A. Fisher's classic 1936 paper, The Use of Multiple Measurements in Taxonomic Problems, and can also be found on the UCI Machine Learning Repository.

It includes three iris species with 50 samples each as well as some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.

The columns in this dataset are:

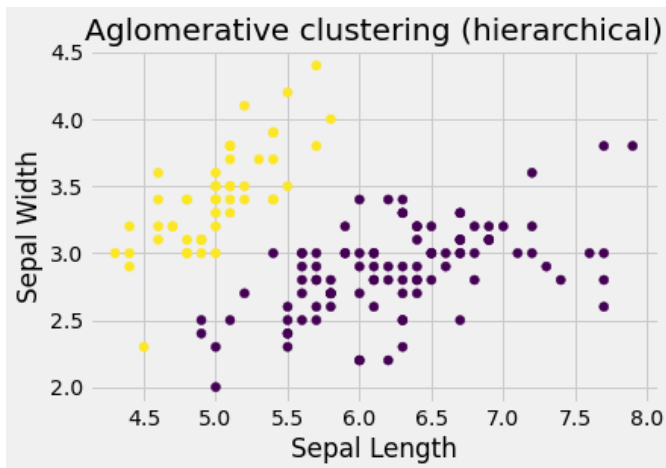
- Id
- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species

Target : clustering similar iris Species that have same characteristics in same group and find out dissimilarities

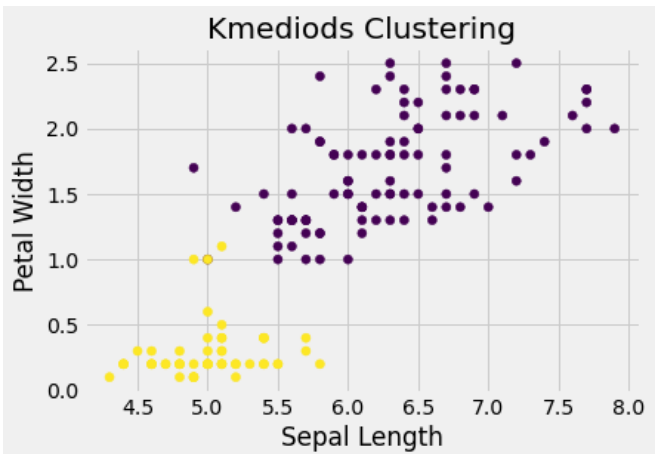
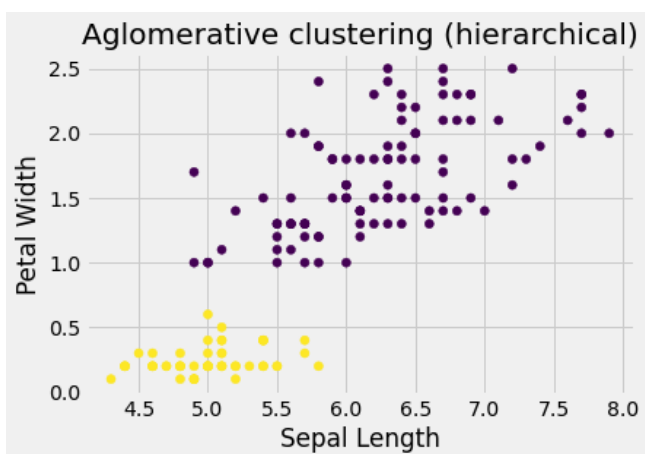
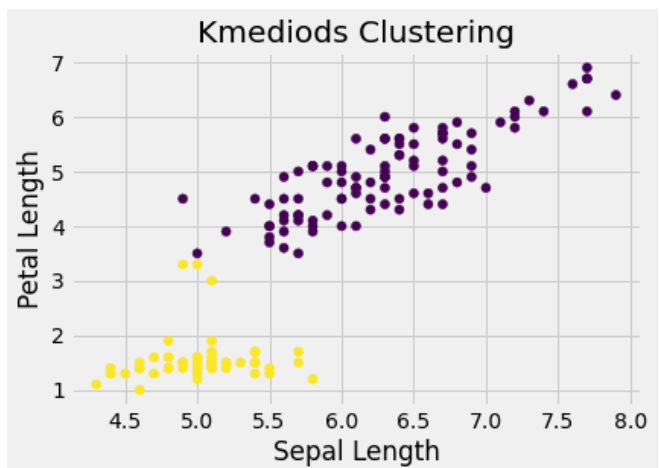
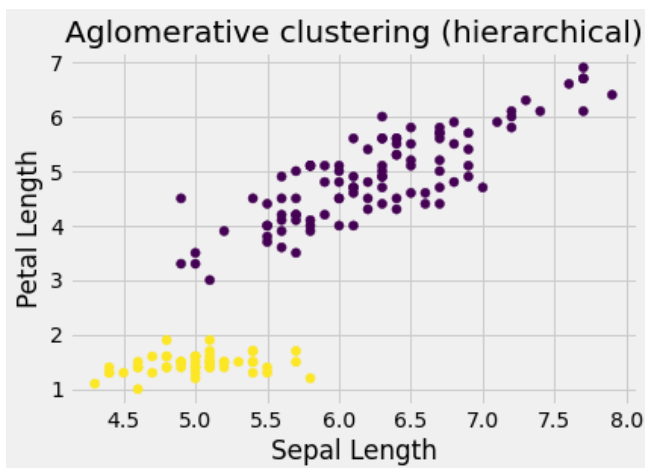
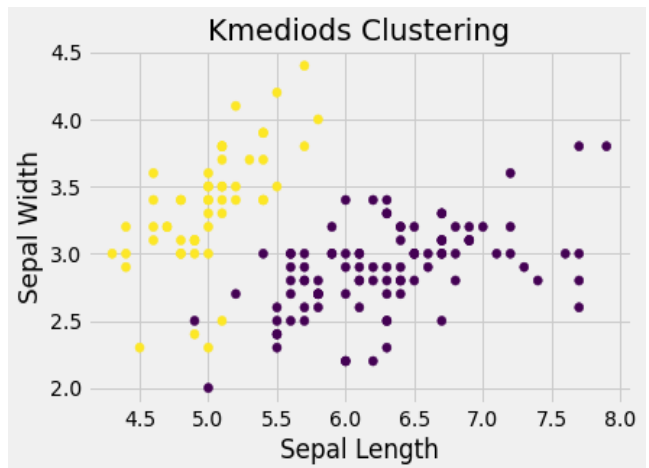
That have been found among groups.

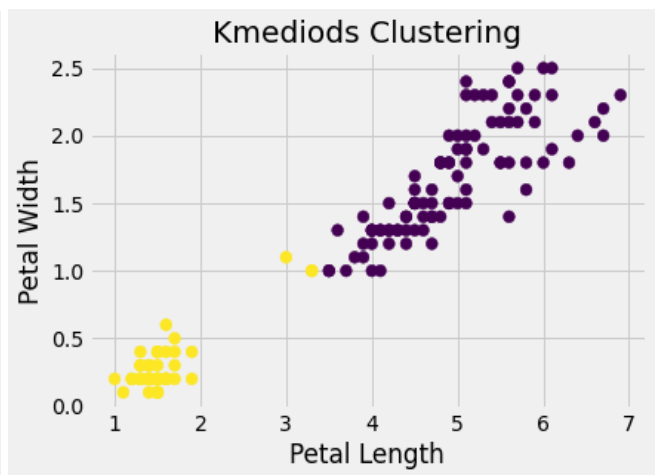
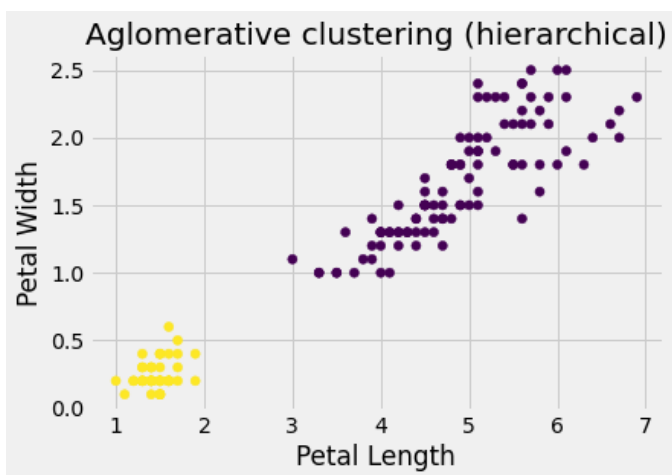
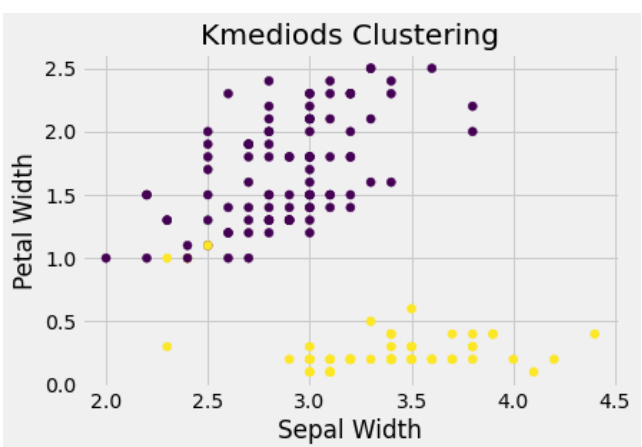
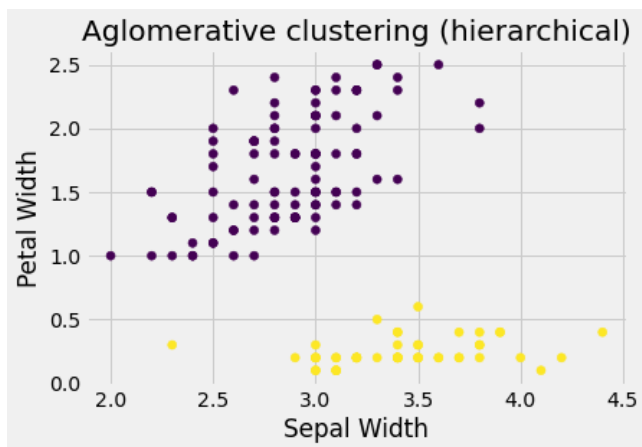
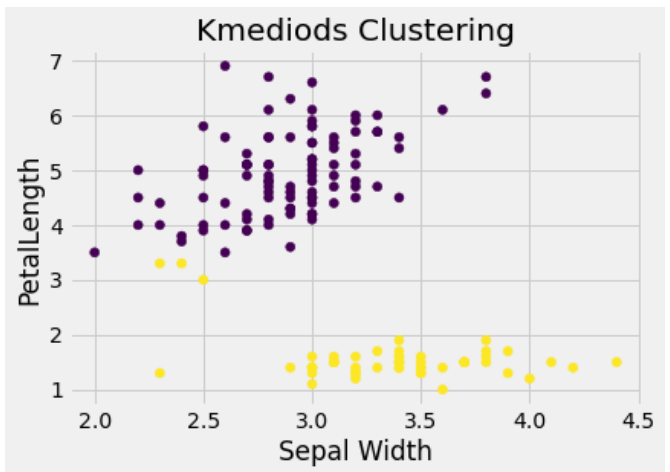
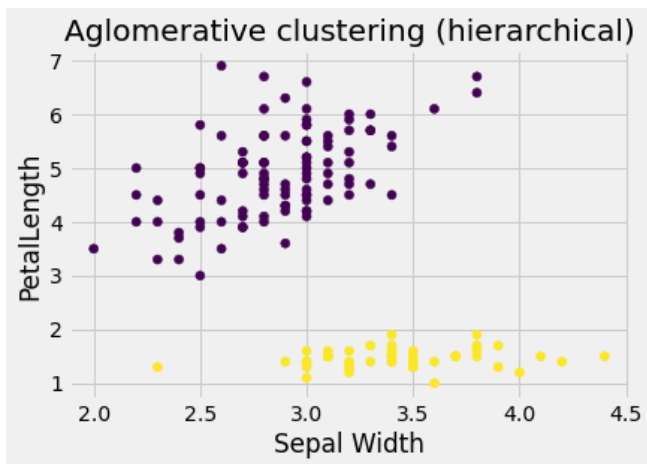
Plotted Graphs

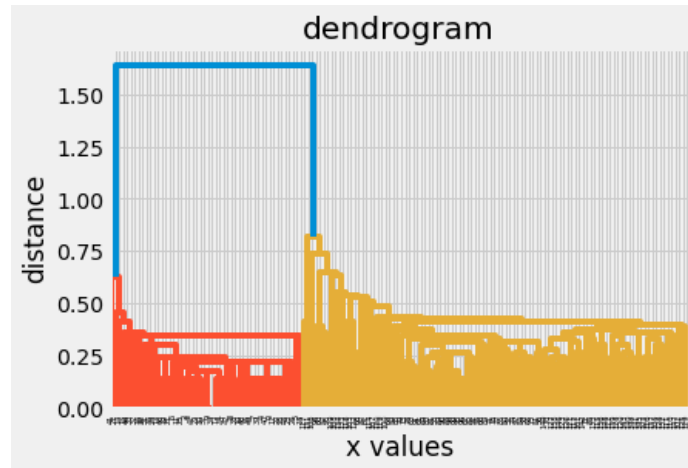
Agglomerative hierarchical clustering



K Medoids Clustering







We can see two colors in Plots, so that means in each algorithm data was divided into two clusters with similarities between each data point and another one in same group and dissimilarities with each data point and another in other group.

K Medoid is Computationally more intensive and we initialized random points.

By watching dendrogram, we can understand that we should use 2clusters in Agglomerative hierarchical clustering.

Comparison

First, let's talk about K medoids, it is robust and more powerful in dealing with outliers as distance between clusters are small and we initialized 2 medoids first and we replace each one of the two medoids with non-medoids if it enhance the total distance of the result clustering.

Second, In Agglomerative hierarchical clustering, We saw dendrogram first so we can decide how many clusters we can made, After that we can conclude that distances between are big so it will may be influenced by outliers but no problem, we can improve that.