

Name: Allan Rodrigues

Class: TE IT A

Roll no: 59









 9.

Q6)

1. What is clustering in data mining?

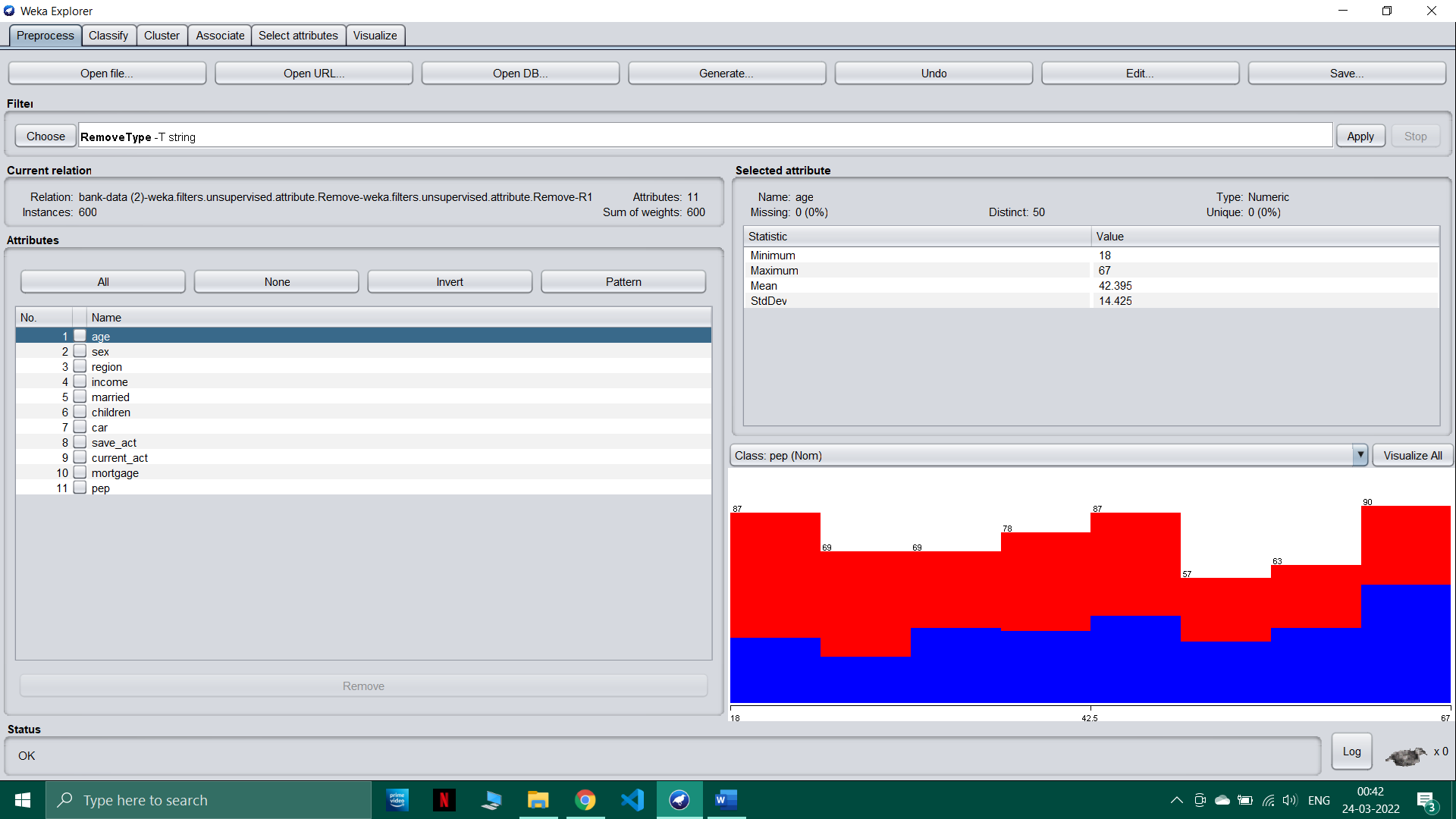
Cluster Analysis is the process to find similar groups of objects in order to form clusters.It is an unsupervised machine learning-based algorithm that acts on unlabelled data. A group of data points would comprise together to form a cluster in which all the objects would belong to the same group.

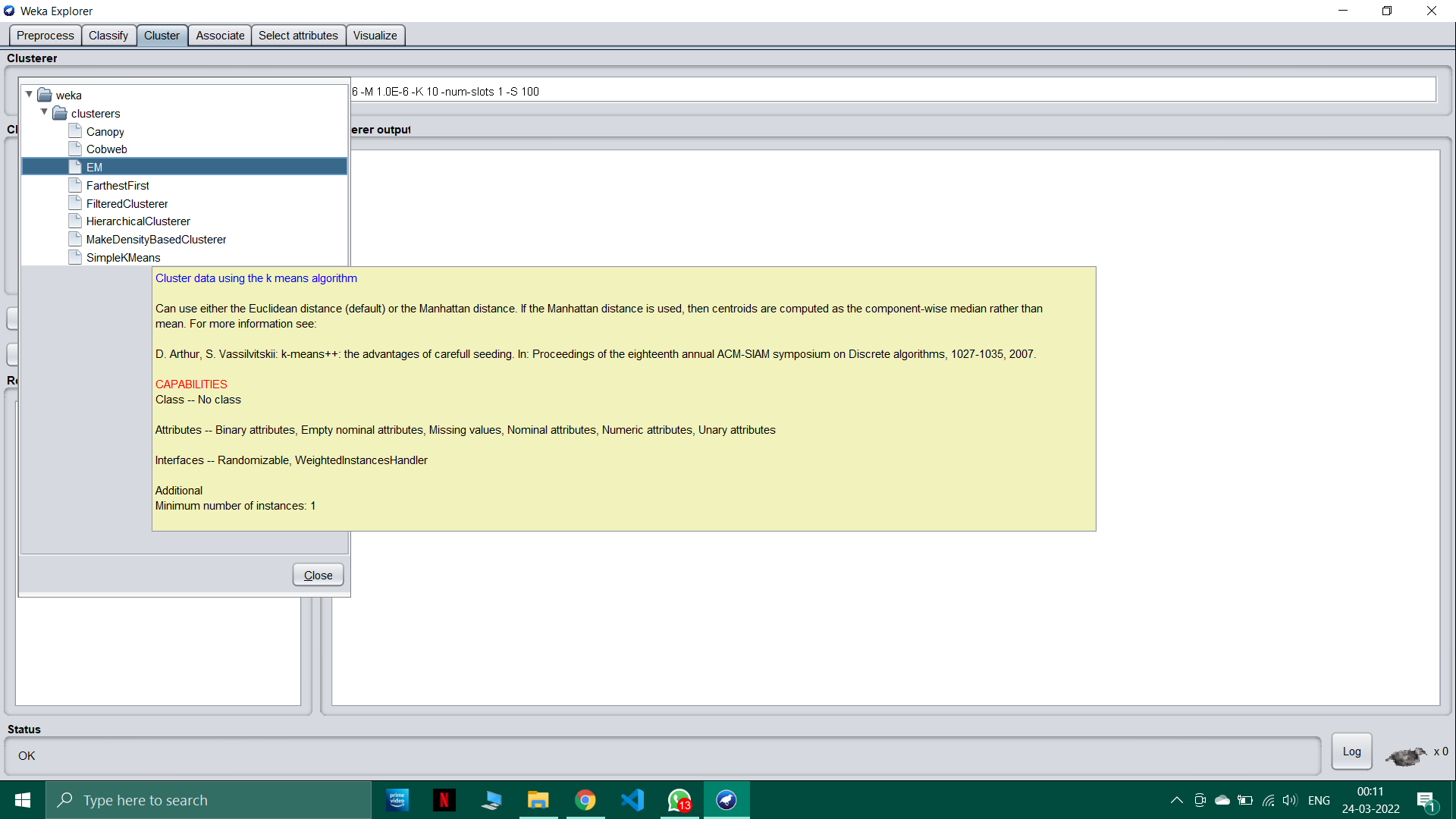
1. Difference between classification and clustering?

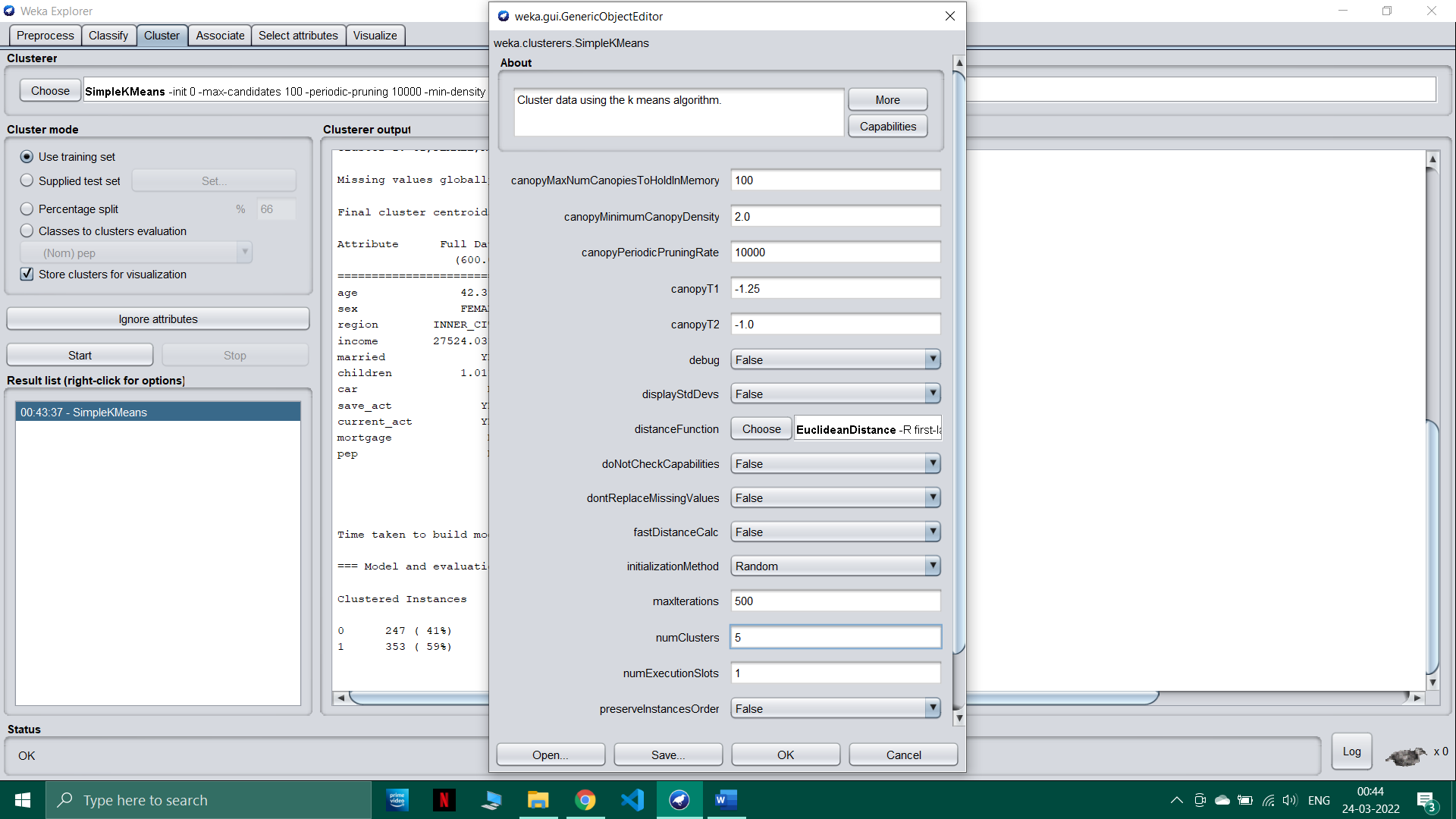
|  |  |
| --- | --- |
| classification | clustering |
| It is used for supervised learning | It is used for unsupervised learning |
| Process of classifying the input instances based on their corresponding class labels | Grouping the instances based on their similarity without the help of class labels |
| It has labels so there is need of training and testing dataset for verifying the model created | There is no need of training and testing dataset |
| Examples : Logistic regression, Naive Bayes classifier, Support vector machines, etc. | Examples : k-means clustering algorithm, Fuzzy c-means clustering algorithm, Gaussian (EM) clustering algorithm, etc. |

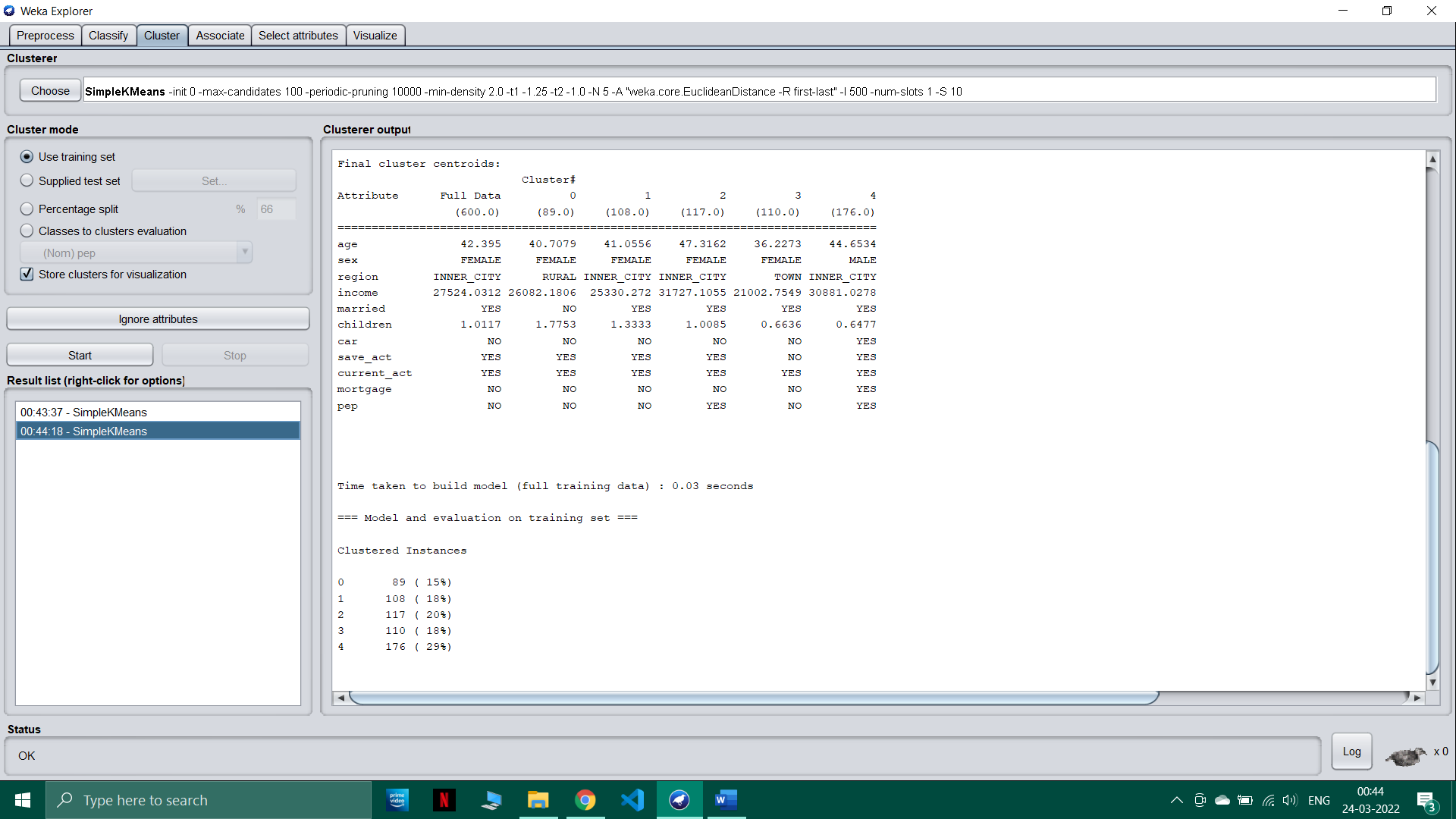
1. Study of algorithms for clustering

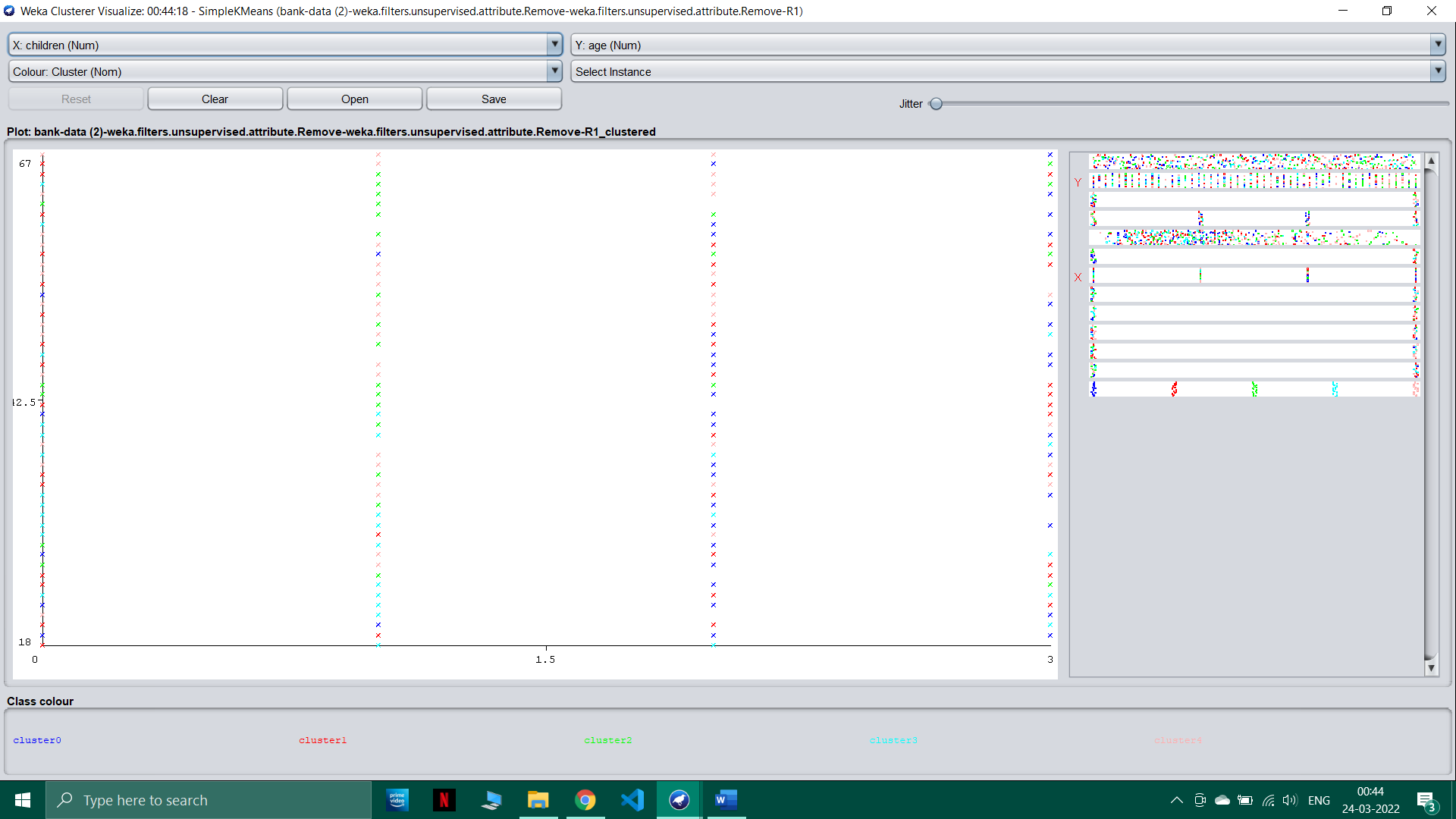
* K-means clustering algorithm
* DBSCAN clustering algorithm
* Gaussian Mixture Model algorithm
* Agglomerative Hierarchy clustering algorithm

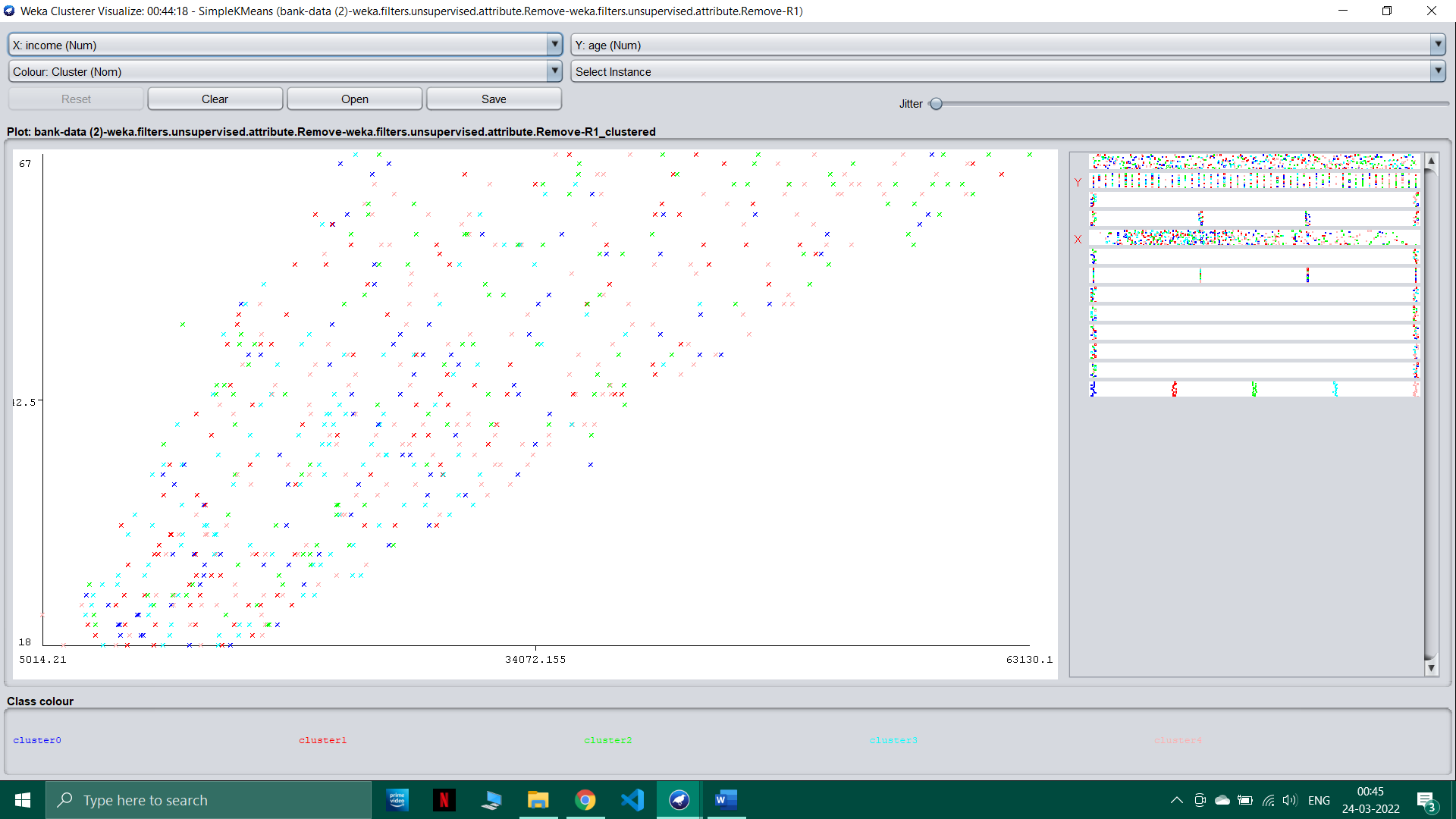












Q8)

