**Comparison between the K-Means and Hieratical algorithms:**

* K-Means is that it needs us to pre-enter the number of clusters (K) but Hierarchical clustering has no such requirements. The algorithm on itself deduces the optimum number of cluster and displays it form of dendrogram.
* Performance of K-Means on spherical data is better than that of Hierarchical Clustering Algorithm
* Hierarchical clustering is a purely agglomerative approach and goes on to build one giant cluster. K-Means algorithm in all its iterations has same number of clusters.
* K-Means need circular data, while Hierarchical clustering has no such requirement.
* K-Means uses median or mean to compute centroid for representing cluster while Hierarchical Clustering Algorithm has various linkage methods that may or may not employ the centroid.
* With the introduction of mini-batches K-Means can work with very large datasets but the Hierarchical Clustering Algorithm lacks in this regard.
* Hierarchical methods are suited for cases that require an arrangement of the clusters into a natural hierarchy. In K-means all clusters are on the same level.
* Hierarchical Clustering Algorithm can produce reproducible results while older versions of K-Means can’t
* K-Means simply divides data into mutually exclusive subsets while Hierarchical Clustering Algorithm arranges it into a tree format.
* Following are the difference between K-Means and Hierarchical Clustering Algorithm (HCA) K-Means is that it needs us to pre-enter the number of clusters (K) but Hierarchical clustering has no such requirements. The algorithm on itself deduces the optimum number of clusters and displays it form of the dendrogram. Performance of K-Means on spherical data is better.

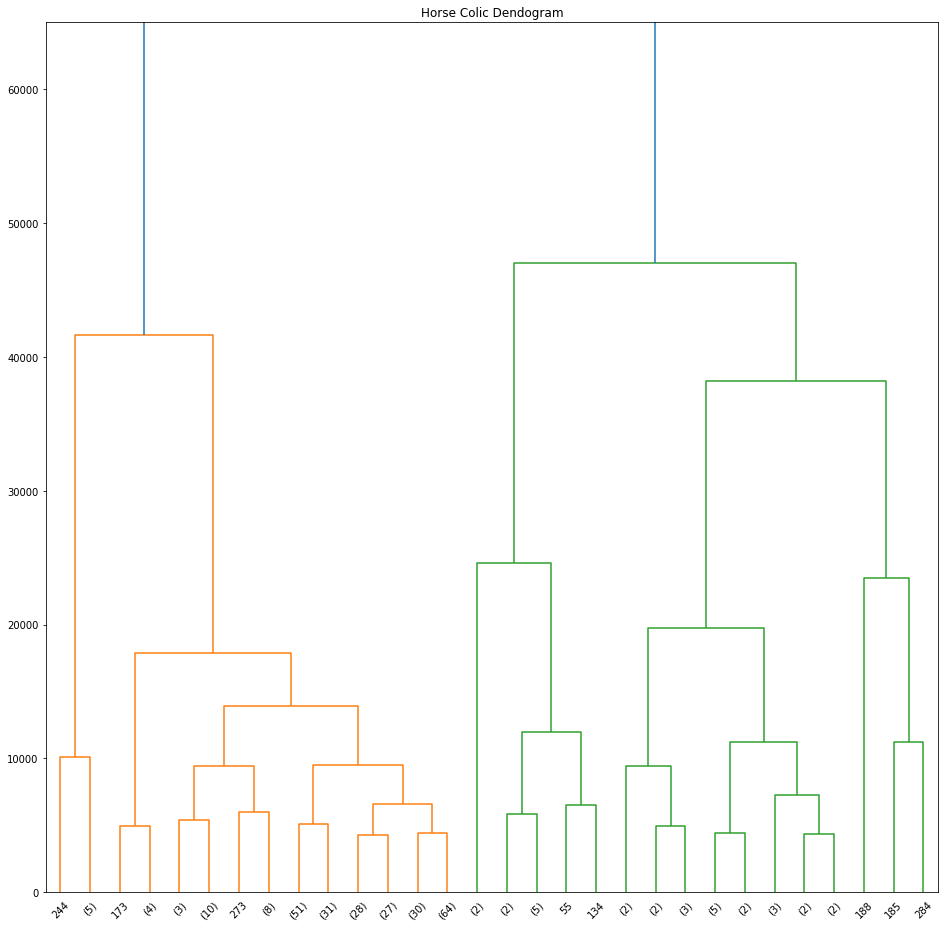
**Visualization:**

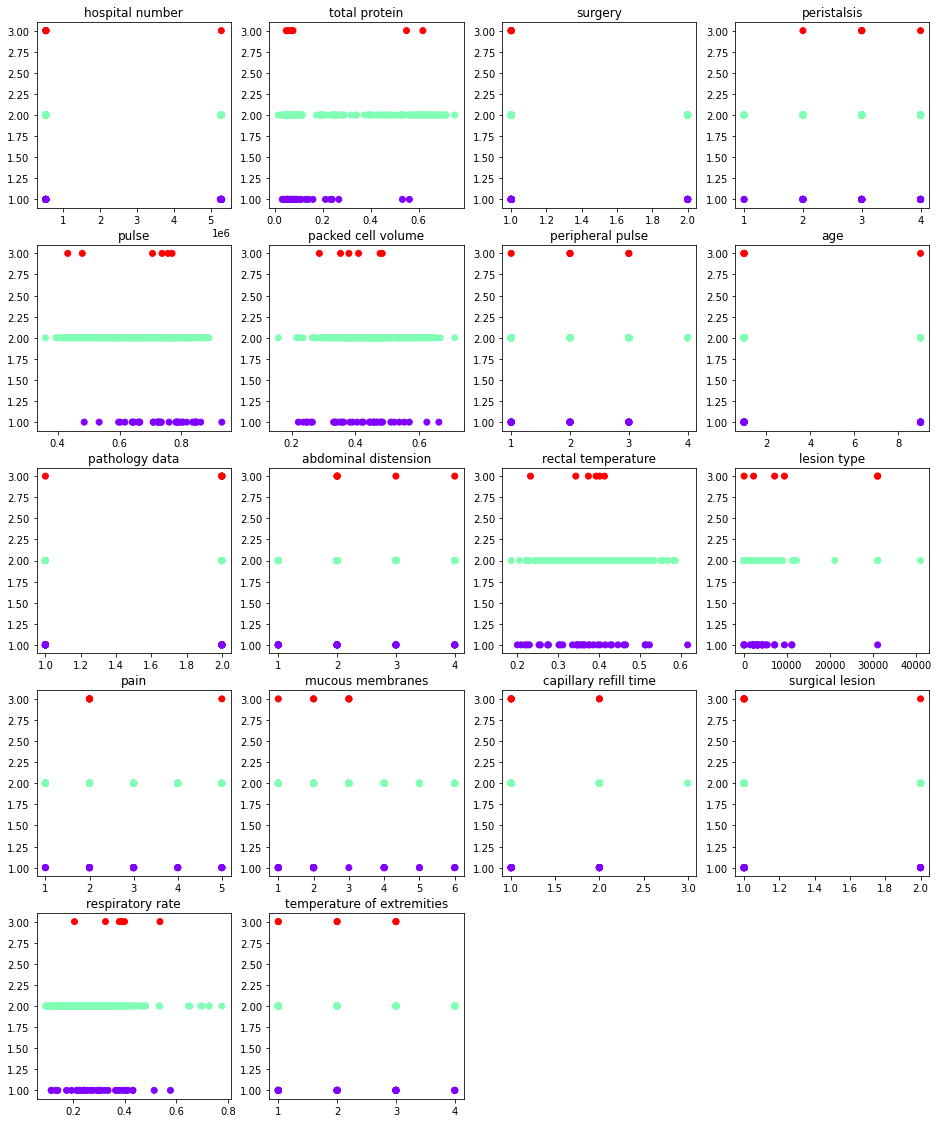
The goal of visualization is the visual representation of data with the help of comprehensive charts, images, lists, charts, and other visual objects. And the following graphs represent the data after pre-processing step that contains updating missing values, Categorical attributes, and normalization of the data set by using the two algorithms k-means and Hieratical algorithm.

**K-Means:A picture containing window, overhead, open, platform

Description automatically generated**

**Hieratical:  
  
 Dendrogram:**

****



**Accuracy:**

Table

Description automatically generated

The accuracy of a classifier is given as the percentage of total correct predictions divided by the total number of instances. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known.

In our data set, the accuracy of the k-means algorithm that resulted from the classification report is 57%, and the accuracy of the hierarchical algorithm is 26%

**Conclusion:**

To sum up the report, it was concluded that Undirected data mining is a technique that involves finding hidden patterns in data without coming up with a hypothesis and it is called clustering. It is commonly used to identify similarities between different objects. There are two types of clustering: hierarchical and non-hierarchical. Non-hierarchical clustering is a technique that involves splitting a dataset into various M clusters. This method is commonly used in business intelligence and the most widely used non-hierarchical clustering technique is K-means. While in hierarchical clustering it involves creating multiple nested sets of objects. Each pair of objects is then added to the set to form a larger cluster. In our data set the accuracy of the k-means algorithm is 57% while the accuracy of the hierarchical algorithm is 26% so k-means algorithm is better and more accurate hierarchical algorithm.