Hierarchical Clustering (HC) is an unsupervised clustering method that creates clusters by arranging or classify things according to inclusiveness of HC. It is an organic way to summarise and abstract the dataset for better understanding.

In this case, we use Agglomerative Clustering (also known as AGNES method) to build a multilevel hierarchy of clusters from point to clusters. We chose k=2 as the number of clusters to cut the dendogram.

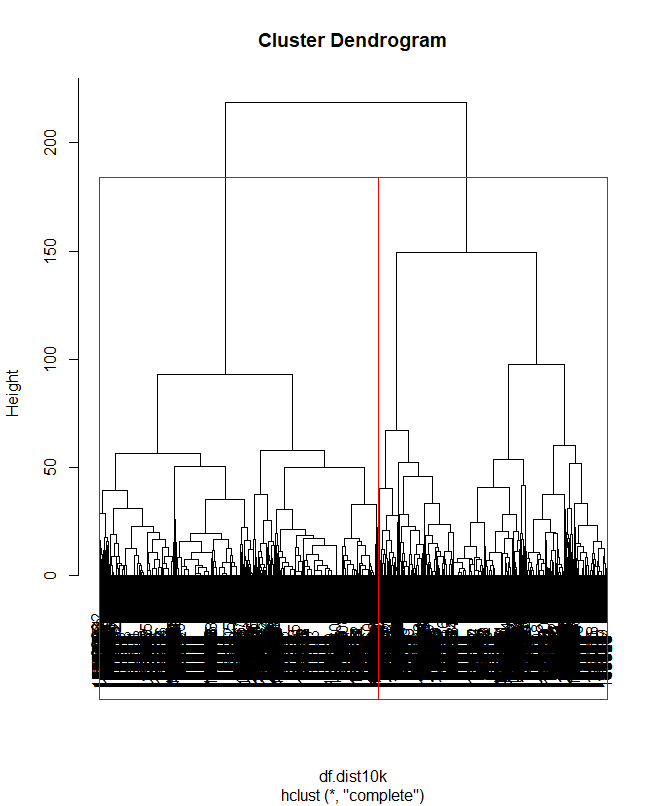


Figure 2.1

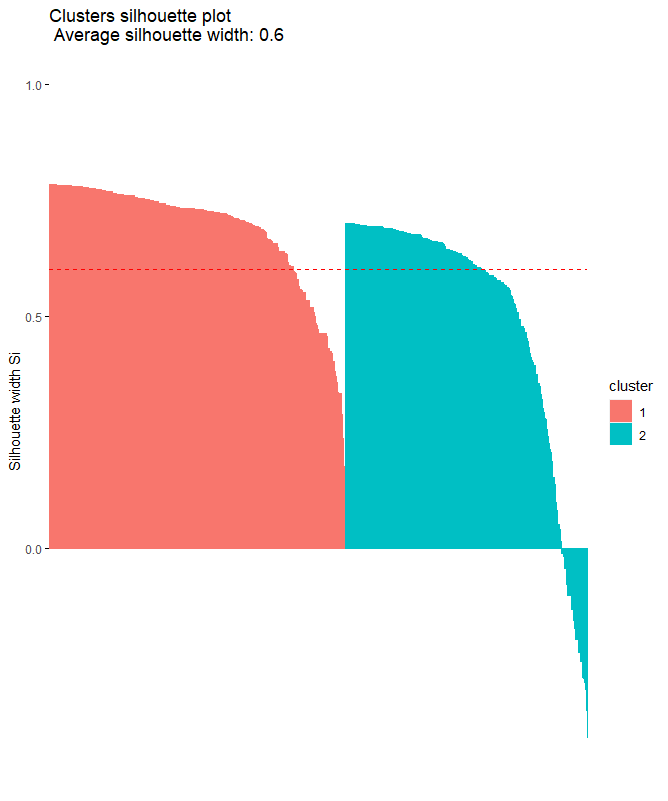


Figure 2.2

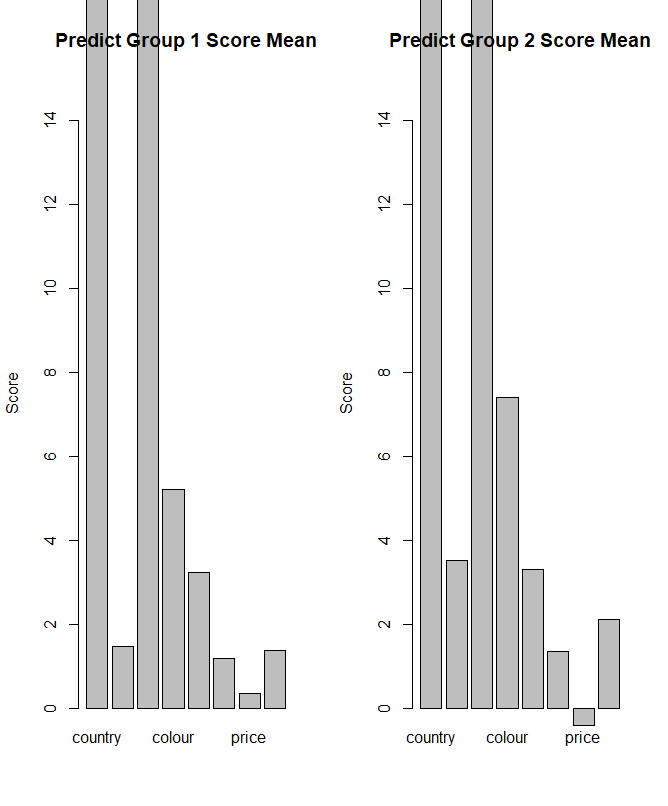
From Figure 2.2, the average silhouette width for Cluster 1 and Cluster 2 are 0.685 and 0.4999 respectively. The average silhouette width is 0.6015 for the chart. 

Figure 2.3

Afterwards, we compare group 1 and group 2 with a randomly sampled training datasets of 10,000 from training data. Then, we test the accuracy of the classification of these 2 groups to then sampled training data.

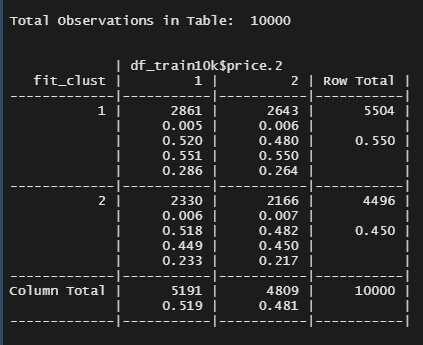


Figure 2.4

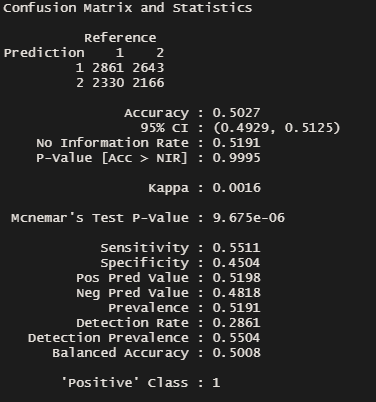


Figure 2.5

The accuracy was 50.267% through applying the hierarchical clustering method in this method.