Final Exam

Durga Prasad Gandi

2023-12-05

##Load the required Libraries  
library(cluster)   
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(flexclust)

## Warning: package 'flexclust' was built under R version 4.3.2

## Loading required package: grid

## Loading required package: lattice

## Loading required package: modeltools

## Loading required package: stats4

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.3.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.3.2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(FactoMineR)

## Warning: package 'FactoMineR' was built under R version 4.3.2

library(ggcorrplot)

## Warning: package 'ggcorrplot' was built under R version 4.3.2

library(dbscan)

## Warning: package 'dbscan' was built under R version 4.3.2

##   
## Attaching package: 'dbscan'

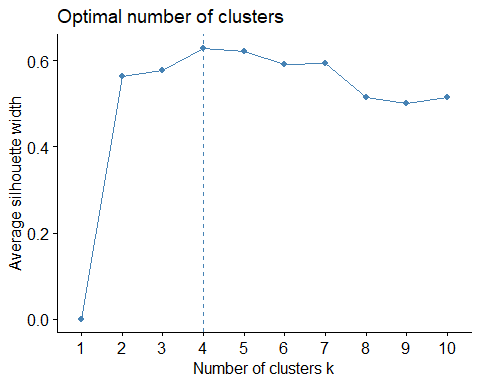
## The following object is masked from 'package:stats':  
##   
## as.dendrogram

#Read the datbase  
fuel\_met= read.csv("C:/Users/gdurg/Downloads/fuel\_receipts1.csv")   
filter\_data = fuel\_met[10:12]   
head(filter\_data)

## sulfur\_content\_pct ash\_content\_pct fuel\_cost\_per\_mmbtu  
## 1 0.49 5.4 2.135  
## 2 0.48 5.7 2.115  
## 3 0.00 0.0 8.631  
## 4 1.69 14.7 2.776  
## 5 0.84 15.5 3.381  
## 6 1.54 14.6 2.199

norm\_data = scale(filter\_data)   
row.names(norm\_data) = fuel\_met[,1]   
distance = get\_dist(norm\_data)   
corr = cor(norm\_data)

fviz\_nbclust(norm\_data, kmeans, method = "silhouette")



**The Sharp rise in the above graph indicates k=4 is the best value**

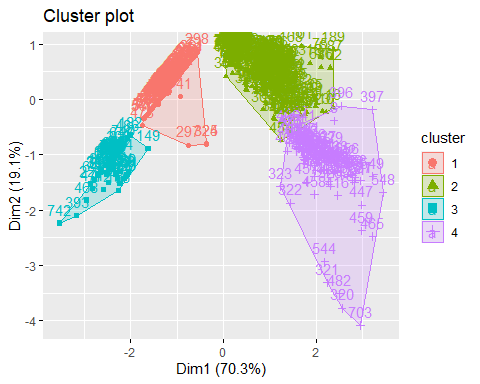
set.seed(1)   
k5 = kmeans(norm\_data, centers = 4, nstart = 25)   
k5$centers

## sulfur\_content\_pct ash\_content\_pct fuel\_cost\_per\_mmbtu  
## 1 -0.68285701 -0.9756003 0.4267807  
## 2 0.03743433 0.7622339 -0.6717529  
## 3 -0.57343169 -0.9756003 2.6109757  
## 4 2.31460610 0.8087649 -0.7318205

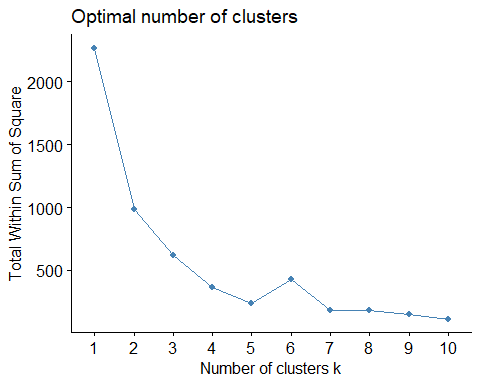
k5$size

## [1] 267 332 67 90

fviz\_cluster(k5, data = norm\_data)



fviz\_nbclust(norm\_data, kmeans, method = "wss")



The Elbow graph here shows the sharp decrease at k=5

***The elbow method’s main concept is to plot the explained variation, also known as the within-cluster sum of squares, as a function of the number of clusters after running the clustering algorithm for a range of cluster numbers. The “elbow” of the plot denotes the point k=5 which the variation begins to decline more slowly, suggesting that the performance of the model remains mostly unchanged even with the addition of more clusters.***

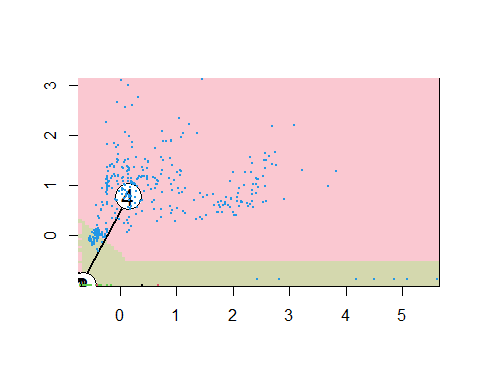
set.seed(1)   
k51 = kcca(norm\_data, k=4, kccaFamily("kmedians"))   
k51

## kcca object of family 'kmedians'   
##   
## call:  
## kcca(x = norm\_data, k = 4, family = kccaFamily("kmedians"))  
##   
## cluster sizes:  
##   
## 1 2 3 4   
## 160 107 67 422

clusters\_index = predict(k51)   
dist(k51@centers)

## 1 2 3  
## 2 0.3849698   
## 3 1.9342521 2.3191240   
## 4 2.3286064 2.1421611 3.7436155

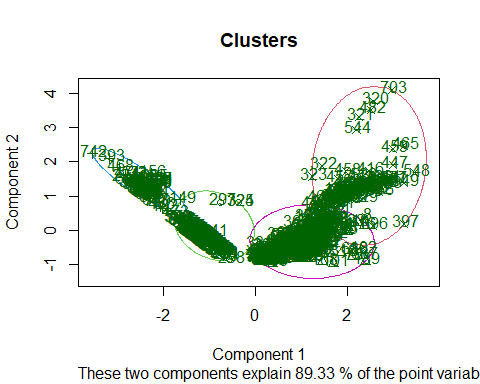
image(k51)   
points(norm\_data, col=clusters\_index, pch=19, cex=0.3)



filter\_data %>% mutate(Cluster = k5$cluster) %>% group\_by(Cluster) %>% summarise\_all("mean")

## # A tibble: 4 × 4  
## Cluster sulfur\_content\_pct ash\_content\_pct fuel\_cost\_per\_mmbtu  
## <int> <dbl> <dbl> <dbl>  
## 1 1 0.0167 0 8.06  
## 2 2 0.771 8.90 2.20  
## 3 3 0.131 0 19.7   
## 4 4 3.15 9.14 1.88

clusplot(norm\_data,k5$cluster, main="Clusters",color = TRUE, labels = 3,lines = 0)

 ##Summary:

Based on their lower-than-average sulfur and low-ash contents, Cluster 1 looks to be a set of fuel samples with favorable combustion properties. Reduced emissions of sulfur dioxide during burning are suggested by the lower sulfur content, which improves the environment. Cluster 1 is not the most affordable option, but it strikes a compromise between reasonable pricing and cleaner combustion qualities, making it a potentially appealing option for enterprises looking for an ecologically friendly and cost-effective fuel source.

With an average sulfur content, a higher-than-average ash concentration, and a noticeably reduced fuel cost per MMBtu, Cluster 2 has a unique composition. The higher proportion of ash points to a higher degree of impurity in the form of non-combustible residue, which could provide problems for energy production in general because of increased emissions and the possibility of equipment corrosion.

In cluster 3 the fuels are particularly noteworthy because of their clean combustion characteristics and noticeably greater cost per MMBtu. These fuels include low ash content, which suggests a more effective and ecologically friendly energy production process, and low sulfur content, which helps to reduce emissions during combustion. These fuels appear to be more expensive than others based on the significant rise in fuel cost per MMBtu seen in this cluster.

Cluster 4, which includes fuels with a unique composition marked by greater levels of ash and sulfur, offers an intriguing contrast. Fuels in this cluster have a cost benefit over average fuel costs per MMBtu, despite these elevated levels of impurity. The intriguing combination of lower production costs and higher pollutant content could mean that the fuels in Cluster 4 are less processed or refined, which would make them a more affordable choice for producing energy. Higher amounts of ash and sulfur should be carefully addressed in terms of their possible environmental effects, since greater emissions during combustion may outweigh the benefits to the economy.

Realative composition:

For Cluster1:

Content of Sulfur: Low (0.0167) Ash Content: Very Minimal (0) Moderate fuel cost per MMBtu (8.06). Cluster 1 fuels have reasonable combustion qualities and a very low sulfur and ash level, making them environmentally beneficial. The feature that strikes a compromise between cost-effectiveness and cleaner combustion is indicated by the fuel cost per MMBtu.

Cluster2:

Moderate Sulfur Content (0.771) Average Ash Content (8.90) Relatively cheap fuel cost per MMBtu (2.20) The fuels in Cluster 2 are less expensive per MMBtu, but they have a moderate sulfur level and a somewhat greater ash percentage. A trade-off between cost-effectiveness and impurity levels is represented by this cluster.

cluster3:

Minimal Sulfur Content (0.131) Ash Content: Very less (0) Expensive per MMBtu of fuel (19.7) The low sulfur and ash level of the fuels in Cluster 3 points to cleaner combustion. The demand for cleaner fuels at a higher production cost is indicated by the higher fuel cost per MMBtu.

Cluster4:

High Sulfur Content (3.15). Ash Content: 9.14, above normal comparatively cheap fuel cost per MMBtu (1.88). The fuels in Cluster 4 are less expensive per MMBtu, but they include more ash and sulfur. This cluster may indicate more affordable, less refined.

In conclusion, the proportional composition of various fuel types with respect to clusters indicates that every cluster is a distinct mixture of economic factors and combustion characteristics. When choosing fuels based on particular standards, the clustering analysis offers insightful information.