Assignment\_4

2025-06-12

# Load necessary libraries  
  
if (!require("cluster")) install.packages("cluster", dependencies = TRUE)

## Loading required package: cluster

## Warning: package 'cluster' was built under R version 4.4.3

if (!require("factoextra")) install.packages("factoextra", dependencies = TRUE)

## Loading required package: factoextra

## Warning: package 'factoextra' was built under R version 4.4.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.4.3

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

if (!require("readr")) install.packages("readr", dependencies = TRUE)

## Loading required package: readr

library(cluster) # For clustering algorithms  
library(factoextra) # For visualizing clustering results  
library(readr) # For reading the excel file  
  
# Load the dataset  
pharma\_data <- read\_csv("C:/Users/arkha/Downloads/Pharmaceuticals.csv")

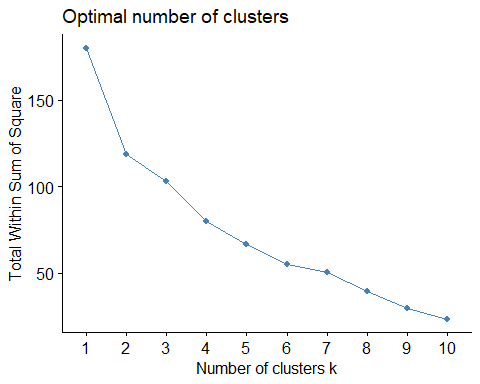
## Rows: 21 Columns: 14

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (5): Symbol, Name, Median\_Recommendation, Location, Exchange  
## dbl (9): Market\_Cap, Beta, PE\_Ratio, ROE, ROA, Asset\_Turnover, Leverage, Rev...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Select only the numerical attributes (columns 3 to 11)  
financial\_data <- pharma\_data[, c(3:11)] # Columns 3 to 11 contain numerical variables  
  
# Standardize the financial data  
standardized\_data <- scale(financial\_data)  
  
# Inspect the scaled data  
head(standardized\_data)

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## [1,] 0.1840960 -0.80125356 -0.04671323 0.04009035 0.2416121 0.0000000  
## [2,] -0.8544181 -0.45070513 3.49706911 -0.85483986 -0.9422871 0.9225312  
## [3,] -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700 0.9225312  
## [4,] 0.1702742 -0.02225704 -0.24290879 0.10638147 0.9181259 0.9225312  
## [5,] -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461 -0.4612656  
## [6,] -0.6953818 2.27578267 0.14948233 -1.45146000 -1.7127612 -0.4612656  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## [1,] -0.2120979 -0.5277675 0.06168225  
## [2,] 0.0182843 -0.3811391 -1.55366706  
## [3,] -0.4040831 -0.5721181 -0.68503583  
## [4,] -0.7496565 0.1474473 0.35122600  
## [5,] -0.3144900 1.2163867 -0.42597037  
## [6,] -0.7496565 -1.4971443 -1.99560225

# Use the elbow method to determine the optimal number of clusters  
set.seed(42) # For reproducibility  
fviz\_nbclust(standardized\_data, kmeans, method = "wss")



# After inspecting the elbow plot, decide on the number of clusters (for example, 3 clusters)  
# Apply K-means clustering with 3 clusters (based on the elbow method result)  
kmeans\_analysis <- kmeans(standardized\_data, centers = 3, nstart = 25)  
  
# Add the cluster assignments to the original dataset  
pharma\_data$Cluster\_Labels <- kmeans\_analysis$cluster  
  
# View the cluster centers to understand the characteristics of each cluster  
print("Cluster Centers (Centroid Values):")

## [1] "Cluster Centers (Centroid Values):"

print(kmeans\_analysis$centers)

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 -0.8261772 0.4775991 -0.3696184 -0.5631589 -0.8514589 -0.9994088  
## 2 0.6733825 -0.3586419 -0.2763512 0.6565978 0.8344159 0.4612656  
## 3 -0.6125361 0.2698666 1.3143935 -0.9609057 -1.0174553 0.2306328  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 0.8502201 0.9158889 -0.3319956  
## 2 -0.3331068 -0.2902163 0.6823310  
## 3 -0.3592866 -0.5757385 -1.3784169

# Investigate patterns in non-numerical attributes (e.g., Recommendation, Headquarters, Exchange)  
print("Cluster Distribution by Recommendation:")

## [1] "Cluster Distribution by Recommendation:"

print(table(pharma\_data$Cluster\_Labels, pharma\_data$Median\_Recommendation))

##   
## Hold Moderate Buy Moderate Sell Strong Buy  
## 1 1 3 2 0  
## 2 6 3 2 0  
## 3 2 1 0 1

print("Cluster Distribution by Location of Firm:")

## [1] "Cluster Distribution by Location of Firm:"

print(table(pharma\_data$Cluster\_Labels, pharma\_data$Location))

##   
## CANADA FRANCE GERMANY IRELAND SWITZERLAND UK US  
## 1 0 1 0 1 0 0 4  
## 2 0 0 0 0 1 2 8  
## 3 1 0 1 0 0 1 1

print("Cluster Distribution by Stock Exchange:")

## [1] "Cluster Distribution by Stock Exchange:"

print(table(pharma\_data$Cluster\_Labels, pharma\_data$Exchange))

##   
## AMEX NASDAQ NYSE  
## 1 1 1 4  
## 2 0 0 11  
## 3 0 0 4

# Assign meaningful names to the clusters based on the analysis of cluster centers  
pharma\_data$Cluster\_Name <- factor(pharma\_data$Cluster\_Labels,   
 levels = c(1, 2, 3),   
 labels = c("Dominant Leaders", "Efficient Innovators", "Moderate Performers"))  
  
# View the dataset with new cluster names  
head(pharma\_data)

## # A tibble: 6 × 16  
## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 ABT Abbott L… 68.4 0.32 24.7 26.4 11.8 0.7 0.42  
## 2 AGN Allergan… 7.58 0.41 82.5 12.9 5.5 0.9 0.6   
## 3 AHM Amersham… 6.3 0.46 20.7 14.9 7.8 0.9 0.27  
## 4 AZN AstraZen… 67.6 0.52 21.5 27.4 15.4 0.9 0   
## 5 AVE Aventis 47.2 0.32 20.1 21.8 7.5 0.6 0.34  
## 6 BAY Bayer AG 16.9 1.11 27.9 3.9 1.4 0.6 0   
## # ℹ 7 more variables: Rev\_Growth <dbl>, Net\_Profit\_Margin <dbl>,  
## # Median\_Recommendation <chr>, Location <chr>, Exchange <chr>,  
## # Cluster\_Labels <int>, Cluster\_Name <fct>

# Summary of clusters  
summary(pharma\_data$Cluster\_Name)

## Dominant Leaders Efficient Innovators Moderate Performers   
## 6 11 4