ML Assignment 4 2025

Lona Uprety

2025-10-26

```
library(readr)
Pharmaceuticals <- read_csv("C:/Users/lona2/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.
```

```
View(Pharmaceuticals)
```

```
#Load the libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v purrr     1.1.0
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(cluster)
library(factoextra)
```

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

```
#Inspect data
head(Pharmaceuticals)
```

```
## # A tibble: 6 x 14
##   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 5 more variables: Rev_Growth <dbl>, Net_Profit_Margin <dbl>,
## #   Median_Recommendation <chr>, Location <chr>, Exchange <chr>
```

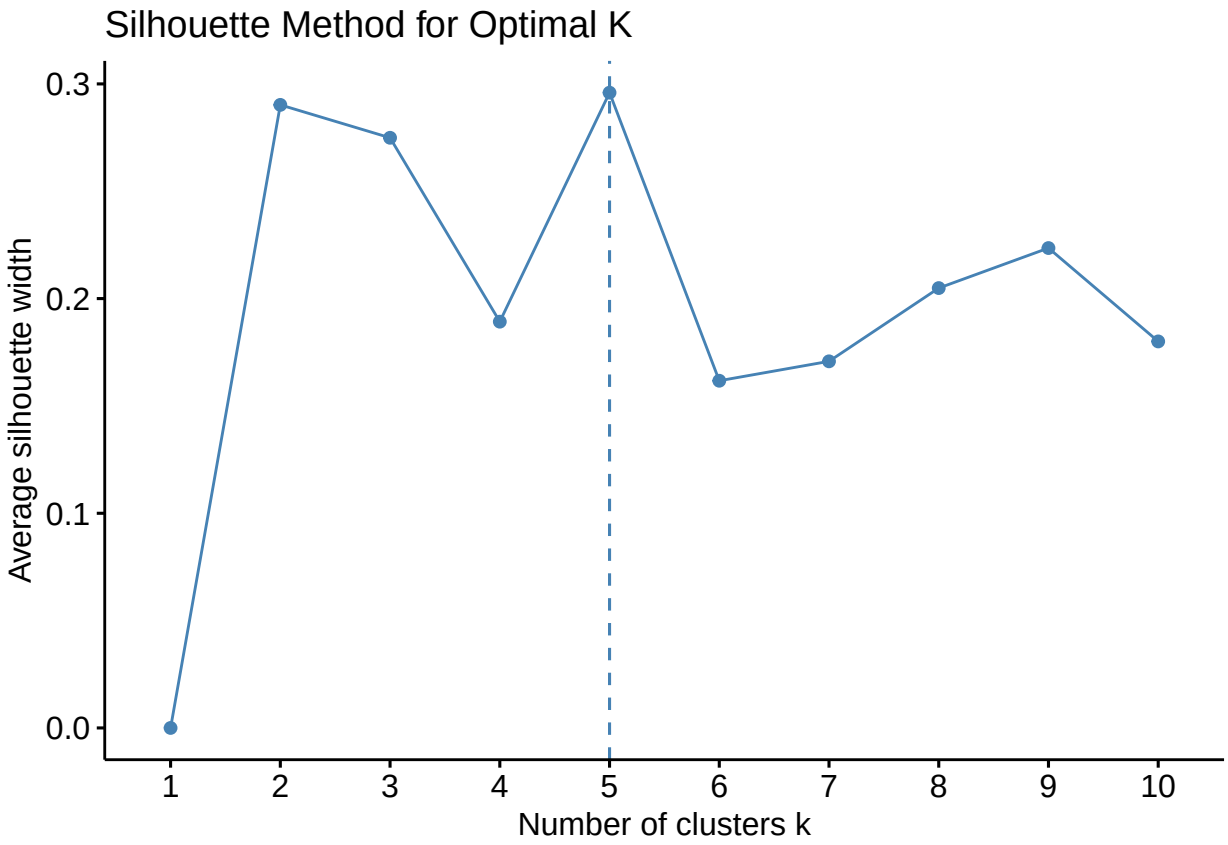
```
summary(Pharmaceuticals)
```

```
##     Symbol              Name             Market_Cap          Beta
##  Length:21          Length:21          Min.   :  0.41   Min.   :0.1800
##  Class :character   Class :character   1st Qu.:  6.30   1st Qu.:0.3500
##  Mode  :character   Mode  :character   Median : 48.19   Median :0.4600
##                                        Mean   : 57.65   Mean   :0.5257
##                                        3rd Qu.: 73.84   3rd Qu.:0.6500
##                                        Max.   :199.47   Max.   :1.1100
##     PE_Ratio           ROE             ROA        Asset_Turnover    Leverage
##  Min.   : 3.60   Min.   : 3.9   Min.   : 1.40   Min.   :0.3    Min.   :0.0000
##  1st Qu.:18.90   1st Qu.:14.9   1st Qu.: 5.70   1st Qu.:0.6    1st Qu.:0.1600
##  Median :21.50   Median :22.6   Median :11.20   Median :0.6    Median :0.3400
##  Mean   :25.46   Mean   :25.8   Mean   :10.51   Mean   :0.7    Mean   :0.5857
##  3rd Qu.:27.90   3rd Qu.:31.0   3rd Qu.:15.00   3rd Qu.:0.9    3rd Qu.:0.6000
##  Max.   :82.50   Max.   :62.9   Max.   :20.30   Max.   :1.1    Max.   :3.5100
##    Rev_Growth    Net_Profit_Margin Median_Recommendation   Location
##  Min.   :-3.17   Min.   : 2.6      Length:21             Length:21
##  1st Qu.: 6.38   1st Qu.:11.2      Class :character      Class :character
##  Median : 9.37   Median :16.1      Mode  :character      Mode  :character
##  Mean   :13.37   Mean   :15.7
##  3rd Qu.:21.87   3rd Qu.:21.1
##  Max.   :34.21   Max.   :25.5
##    Exchange
##  Length:21
##  Class :character
##  Mode  :character
##
##
##
```

```
#Selecet numeric variables (3-11)
numeric_data <- Pharmaceuticals[, 3:11]
```

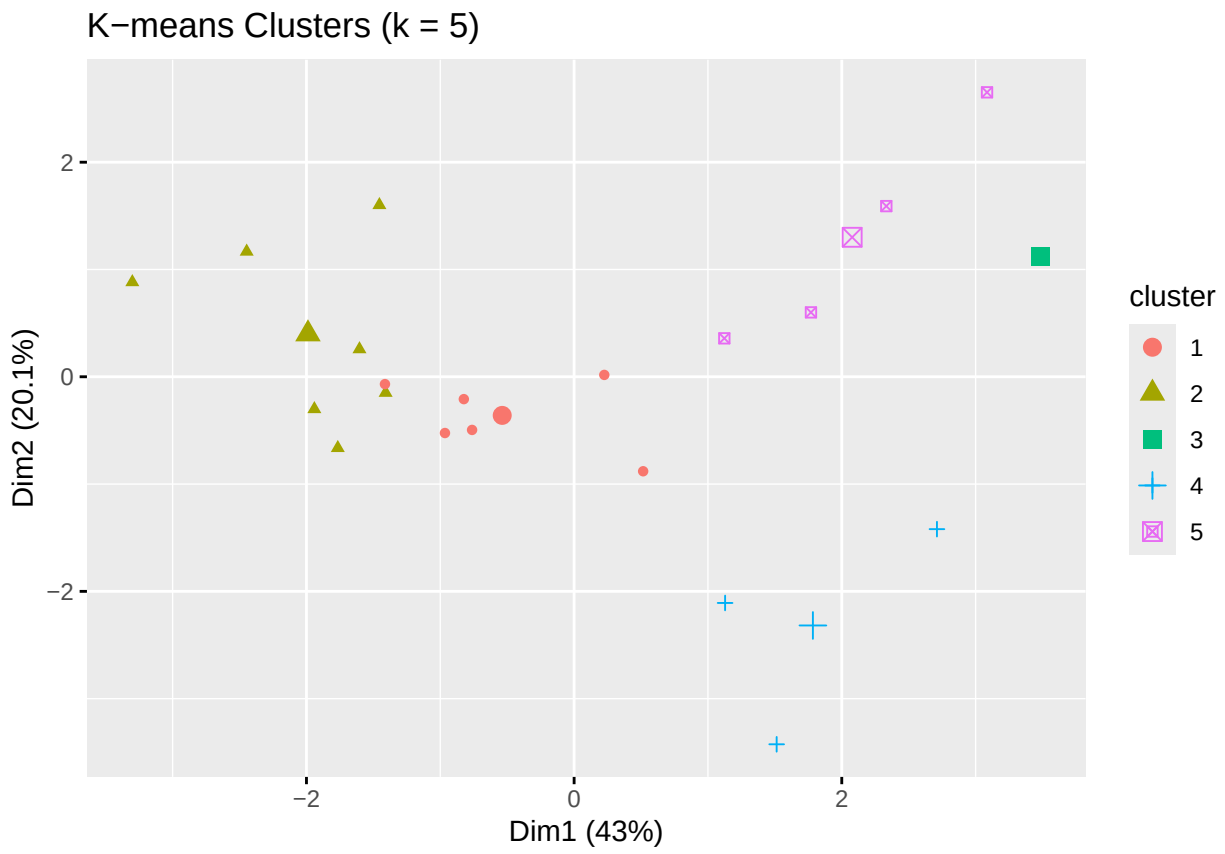
```
#Data preprocessing
if(all(numeric_data[,1] > 0)){
numeric_data[,1] <- log10(numeric_data[,1])
}
scaled_data <- scale(numeric_data)
```

```
#Determine number of clusters
set.seed(123)
fviz_nbclust(scaled_data, kmeans, method = 'silhouette') +
ggtitle('Silhouette Method for Optimal K')
```




```r
#Apply K-means clustering
set.seed(123)
km <- kmeans(scaled_data, centers = 5, nstart = 25)
Pharmaceuticals$Cluster <- as.factor(km$cluster)
# a)
#K-means clustering method was applied to 21 pharmaceutical companies using numeric varibales, standari
```

```r
#Visualize cluster
fviz_cluster(km, data = scaled_data, geom = "point",
             ellipse = FALSE, main = "K-means Clusters (k = 5)")
```




```
#Interpert clusters
aggregate(numeric_data, by = list(Cluster = Pharmaceuticals$Cluster), mean)
```

```
##   Cluster Market_Cap Beta PE_Ratio      ROE      ROA Asset_Turnover  Leverage
## 1       1  1.6157454 0.33 22.31667 22.43333 10.85000      0.6833333 0.2700000
## 2       2  1.9961262 0.51 19.62857 42.11429 16.38571      0.8857143 0.3671429
## 3       3 -0.3872161 0.85 26.00000 24.10000  4.30000      0.6000000 3.5100000
## 4       4  1.2858671 0.64 55.63333 10.10000  4.20000      0.7000000 0.3166667
## 5       5  0.2248667 0.68 17.62500 14.47500  6.02500      0.4250000 0.9125000
##   Rev_Growth Net_Profit_Margin
## 1   8.913333         17.266667
## 2  13.170000         20.342857
## 3   6.380000          7.500000
## 4   6.996667          5.133333
## 5  26.937500         15.175000
```

```
# b)
#1: stable mid-size firms with healthy returns and consistent growth
#2: financially dominant firms with strong performance and global stability
#3: highly leveraged, volatile small firms with weak performance
#4: possibly overvalued or growth-speculative companies with low profitability
#5: firms in expansion phase showing strong top-line growth despite moderate returns
```

```r
#Check patterns for other variables
table(Pharmaceuticals$Cluster, Pharmaceuticals$Median_Recommendation)
```

```
##
##     Hold Moderate Buy Moderate Sell Strong Buy
##   1    3            2             0          1
##   2    3            2             2          0
##   3    0            1             0          0
##   4    2            1             0          0
##   5    1            1             2          0
```

```r
table(Pharmaceuticals$Cluster, Pharmaceuticals$Location)
```

```
##
##     CANADA FRANCE GERMANY IRELAND SWITZERLAND UK US
##   1      0      1       0       0           1  1  3
##   2      0      0       0       0           0  2  5
##   3      0      0       0       0           0  0  1
##   4      1      0       1       0           0  0  1
##   5      0      0       0       1           0  0  3
```

```r
table(Pharmaceuticals$Cluster, Pharmaceuticals$Exchange)
```

```
##
##     AMEX NASDAQ NYSE
##   1    0      0    6
##   2    0      0    7
##   3    0      1    0
##   4    0      0    3
##   5    1      0    3
```

```r
# c)
#1: Mostly Hold/Moderate Buy; mainly U.S./Europe firms on NYSE.
#2: Strong Hold/Moderate Buy sentiment; U.S./UK firms, all NYSE-listed.
#3: Single Moderate Buy; small U.S. firm on NASDAQ.
#4: Mostly Hold; smaller firms in Canada, Germany, and U.S., NYSE-listed.
#5: Mix of Hold/Moderate Sell; U.S./Ireland firms on NYSE/AMEX.

#Summary:
#Larger, stable firms (Clusters 1-2) dominate the NYSE and have positive analyst sentiment, while small
```

```r
# d) Name the clusters
#1: stable performers
#2: global market leaders
#3: high risk niche player
#4: speculative growth firms
#5: emerging growth companies
```