

BA 64060 - Assignment 4

Disney Maxwell

2025-10-25

Install packages

```
# Data manipulation  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## 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  
## v purrr      1.1.0  
## -- 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
```

```
# Clustering algorithms and Visualization  
library(factoextra)
```

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

```
library(dplyr)
```

Question A

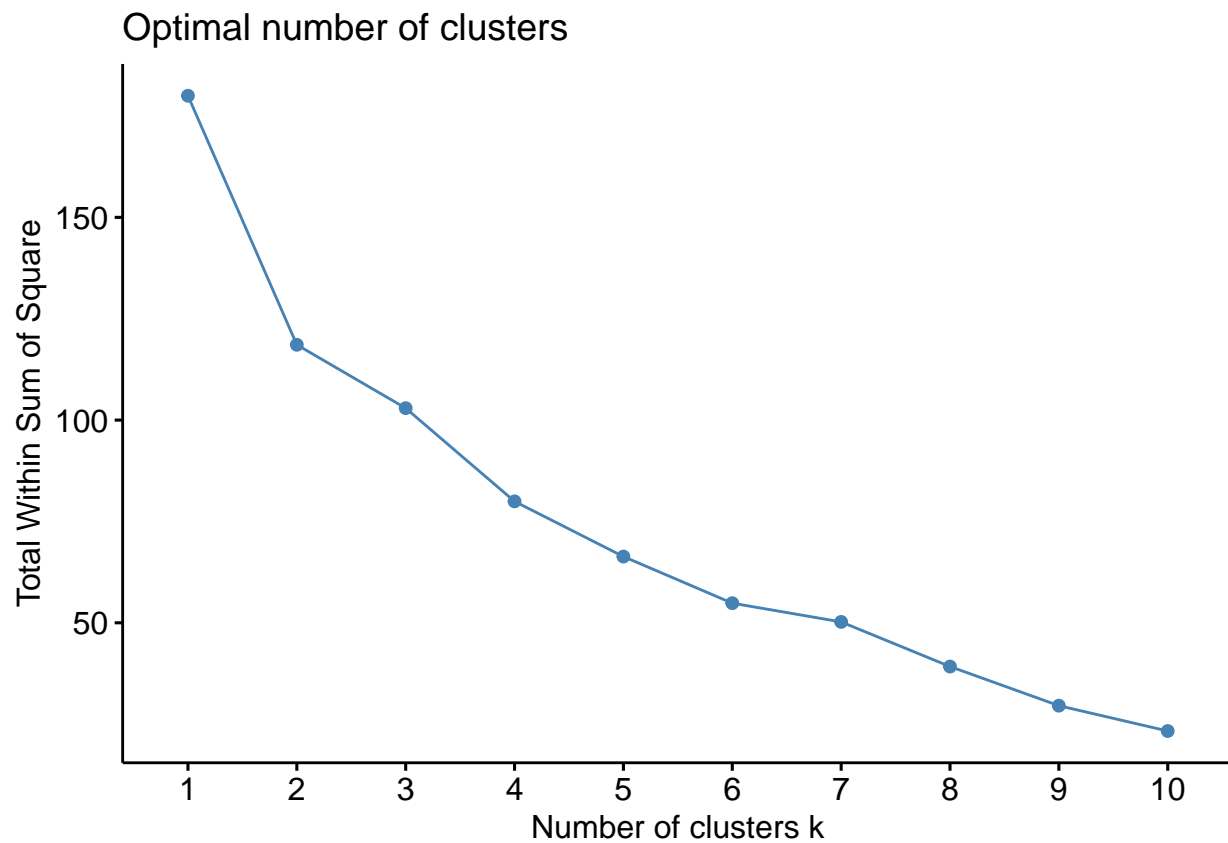
```
pharma_companies_init<-read.csv("Pharmaceuticals.csv")  
  
# Select only numeric variables:  
# Market Cap, Beta, P/E ratio, ROE, ROA, Asset Turnover, Leverage, Estimated revenue growth and Net profit margin  
pharma_companies_data = pharma_companies_init[,c("Market_Cap","Beta","PE_Ratio","ROE","ROA","Asset_Turnover","Leverage","Revenue_Growth","Net_Profit_Margin")]  
  
# The 9 numeric variables have different units and scales. Market cap is in billions of dollars and some are percentages.  
# By normalization, all variables can be equally weighted.  
  
# Normalization  
pharma_companies_norm<-scale(pharma_companies_data)
```

Tuning K-Means Clustering Model- Choice of number of clusters

```
set.seed(123)
print(paste0("Elbow method"))
```

```
## [1] "Elbow method"
```

```
fviz_nbclust(pharma_companies_norm, kmeans, method="wss")
```

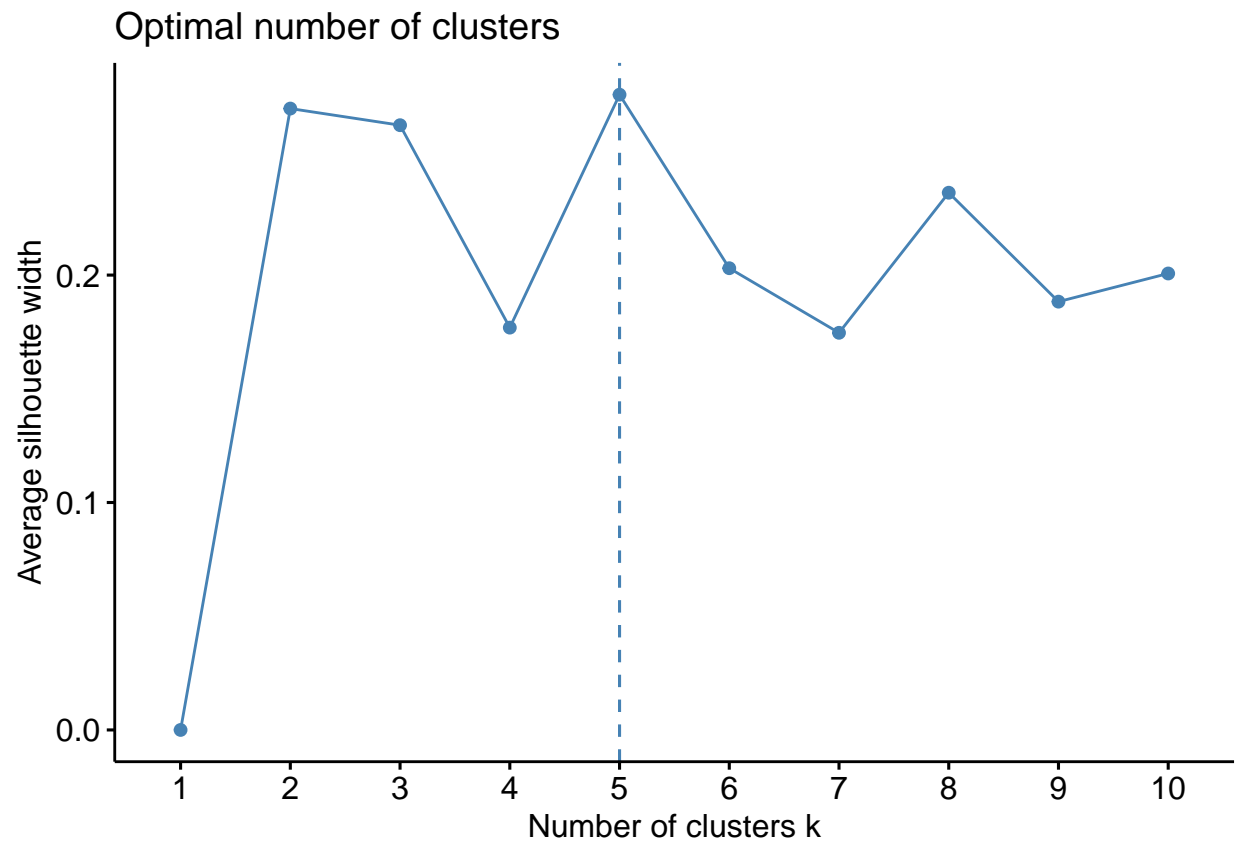


#The above graph from the Elbow method shows that the elbow point 5 is where WSS decreases sharply.

```
print(paste0("Average Silhouette method"))
```

```
## [1] "Average Silhouette method"
```

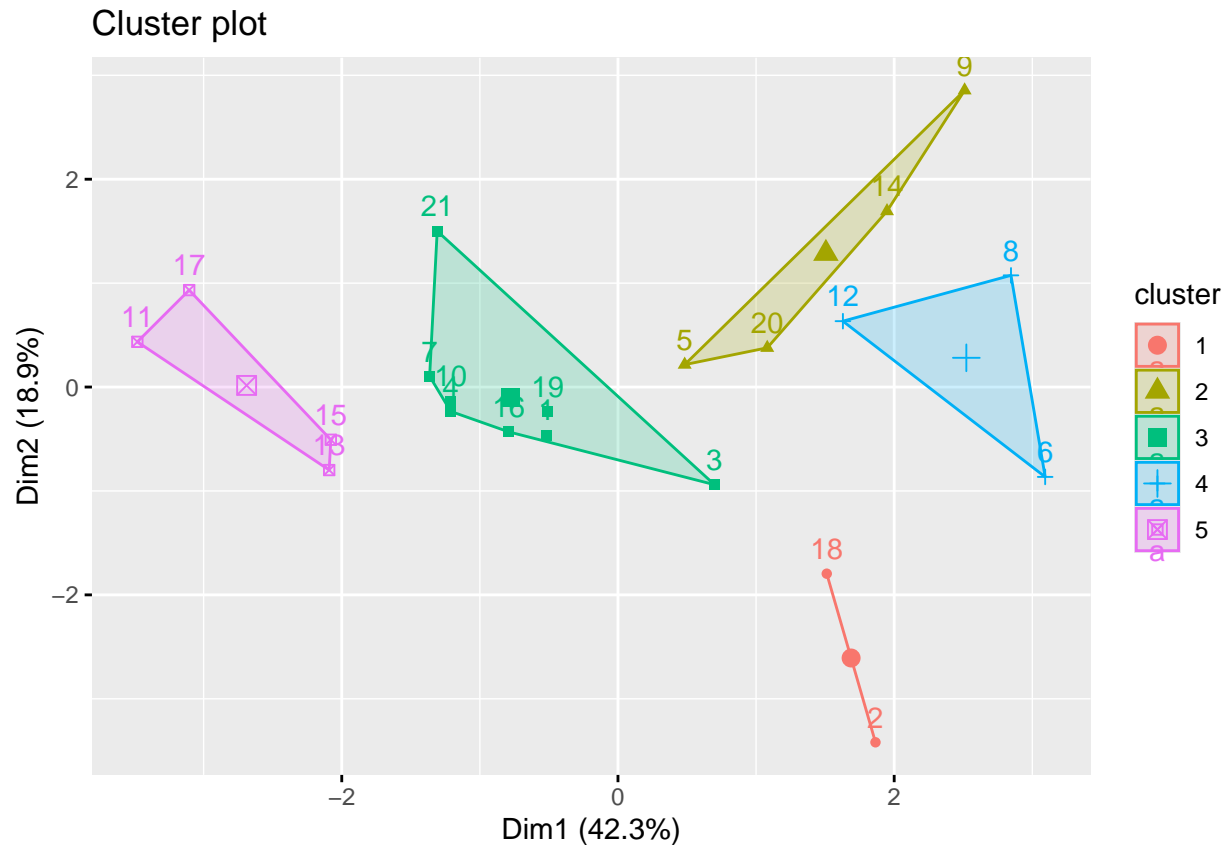
```
fviz_nbclust(pharma_companies_norm, kmeans, method="silhouette")
```



#This is confirmed again in the Silhouette method where k=5 gives the ideal number of clusters.

K-Means Clustering

```
k5 <- kmeans(pharma_companies_norm, centers = 5, nstart = 25)
fviz_cluster(k5, data=pharma_companies_norm)
```



Question B - Characteristics of the Clusters

- a. Companies 2 and 18 are in Cluster 1. These are companies with high Price to Earnings ratios and similar ROE, ROA and moderate revenue growth.
- b. Companies 5, 9, 14 and 20 are in Cluster 2. These are companies with high revenue growths and mostly have small and mid market caps.
- c. Companies 1, 3, 4, 7, 10, 16, 19 and 21 are in Cluster 3. These are companies with good ROE and ROA. They mostly have similar Asset Turnover numbers and low Beta.
- d. Companies 6, 8 and 12 are in Cluster 4. These companies have the same asset turnover but majority have high leverage.
- e. Companies 11, 13, 15 and 17 are in Cluster 5. These are companies with a market cap greater than 100 billion dollars. They have high ROE, ROA and moderate to high revenue growth.

Question C - Characteristics with respect to Median recommendation

When considering the 3 variables: Median recommendation, Location and Exchange, it's reasonable to consider how the Median recommendation relates to the 5 clusters.

- a. Cluster 1 - As they have high P/E ratios, these two companies have “Moderate Buy” and “Hold” recommendations.
- b. Cluster 2 - As these are small and mid market cap companies, these relate to “Moderate Buy” and “Moderate Sell” recommendations.
- c. Cluster 3 - These companies relate to mostly the “Hold” recommendation.
- d. Cluster 4 - These companies relate to mostly the “Hold” recommendation.
- e. Cluster 5 - These companies relate to “Hold” or “Moderate Buy” recommendations.

Question D - Cluster Names

- a. Cluster 1 - Over valued or High Price to Earnings ratio Companies [AGN and PHA]
- b. Cluster 2 - Small and Mid Market Cap Companies with High revenue growth [AVE, ELN, MRX and WPI]
- c. Cluster 3 - Low risk (low beta) Companies [AHM, LLY, NVS, SGP and WYE]
- d. Cluster 4 - High Leverage Companies [BAY, CHTT and IVX]
- e. Cluster 5 - Large Market Capitalization Companies [GSK, JNJ, MRK and PFE]