Assignment_4

Brianna Viñas

11/4/2021

```
Phar<-read.csv("Pharmaceuticals.csv")

library(cluster)
library(factoextra)
```

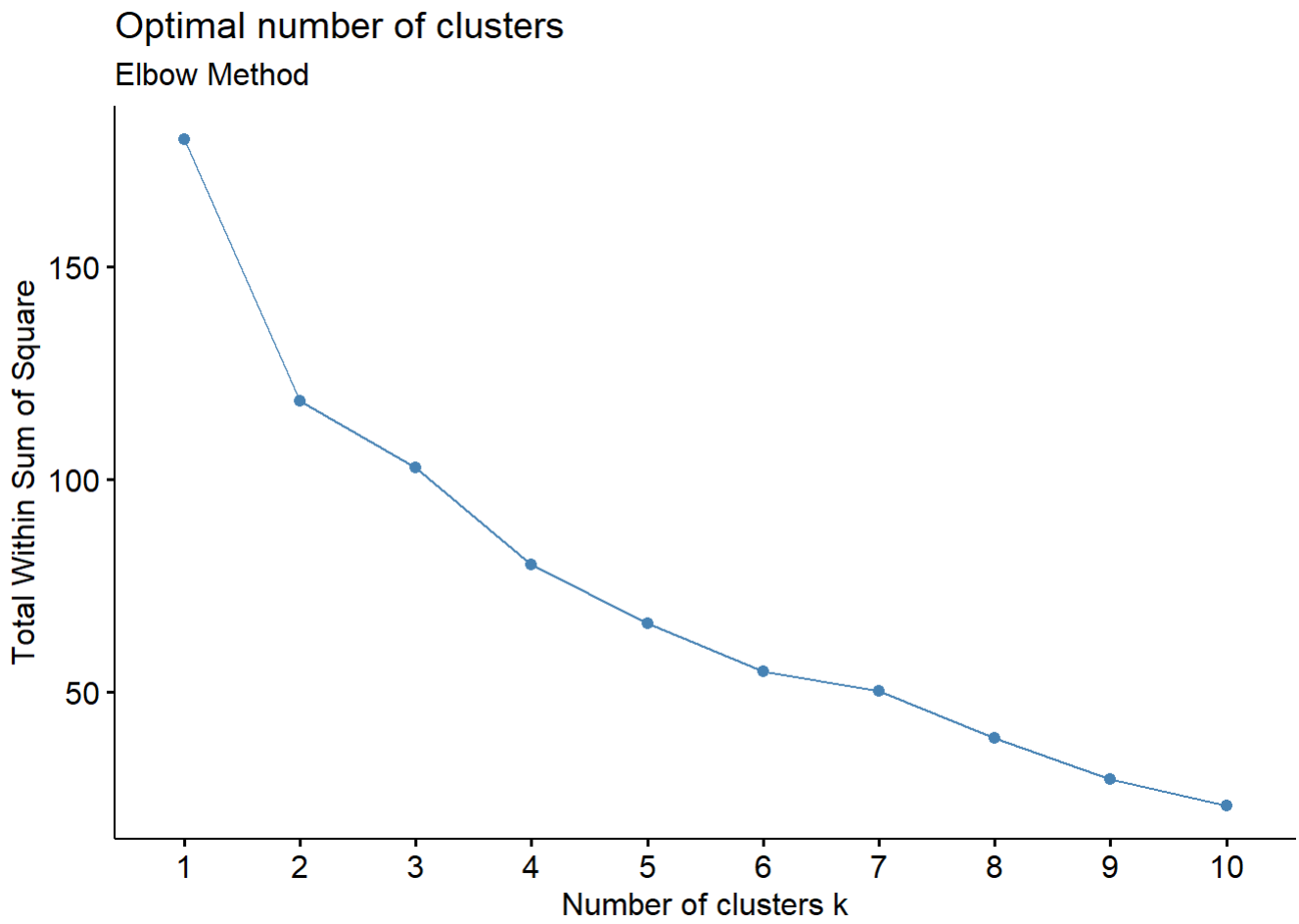
```
## Loading required package: ggplot2
```

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

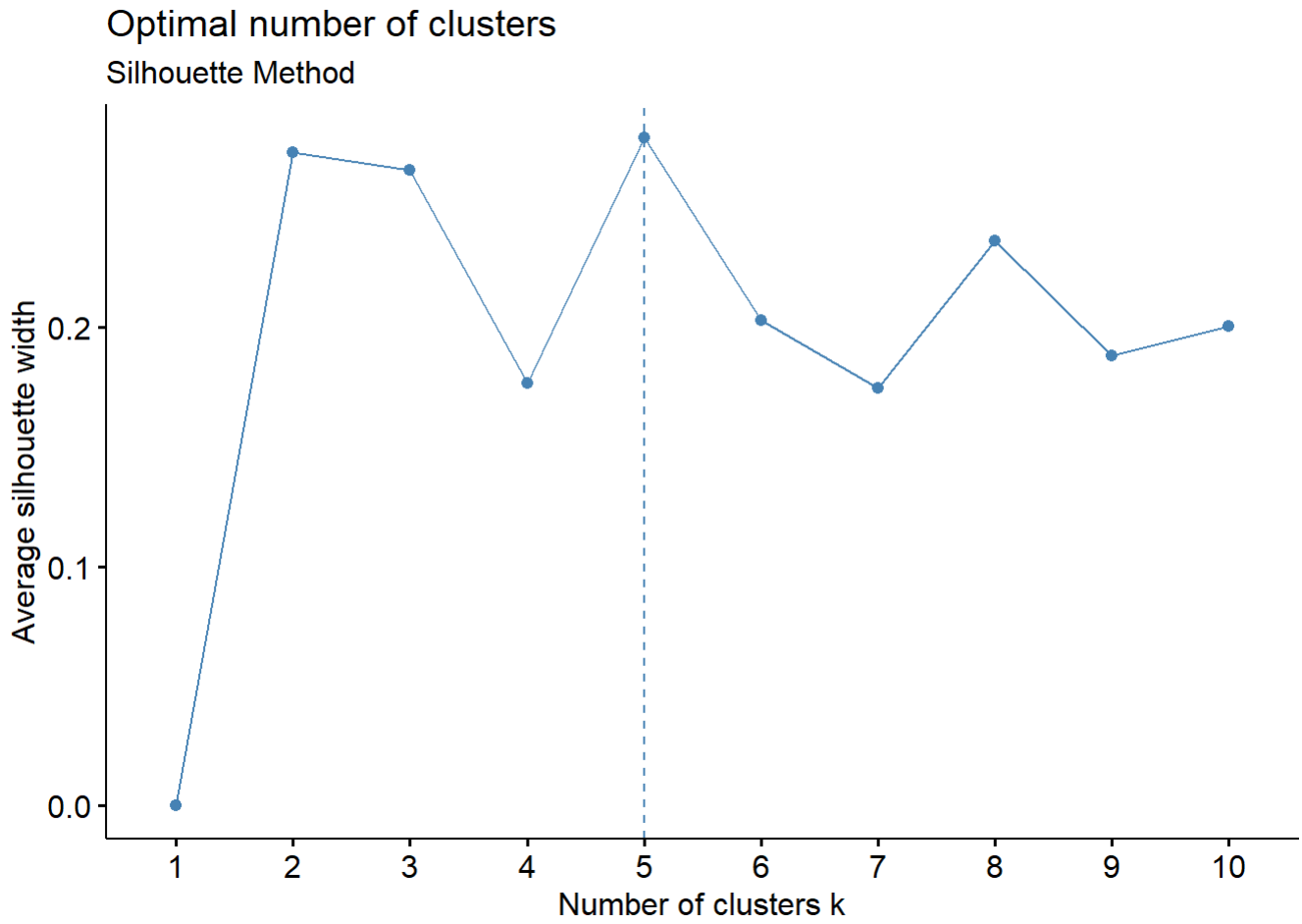
```
# In my first step, I have read the .csv file and installed the necessary packages and libraries
needed in order to complete the assignment.After running the .csv file and labeling it as (Pha
r), I installed the factoextra and cluster packages and libraries.

# a) Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choi
ces made in conducting the cluster analysis, such as weights for different variables, the specif
ic clustering algorithm(s) used, the number of clusters formed, and so on.

Phar1<-Phar[,3:11]
head(Phar1)
```

```
##   Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## 1      68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## 2       7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## 3       6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
## 4      67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## 5      47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 6      16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
##   Net_Profit_Margin
## 1              16.1
## 2               5.5
## 3              11.2
## 4              18.0
## 5              12.9
## 6               2.6
```

```
#Scaling the data of the .csv file.
Phar2<-scale(Phar1)

#Calculating the k value using the Elbow and Silhouette Methods.
fviz_nbclust(Phar2, kmeans, method= "wss") + labs(subtitle="Elbow Method")
```



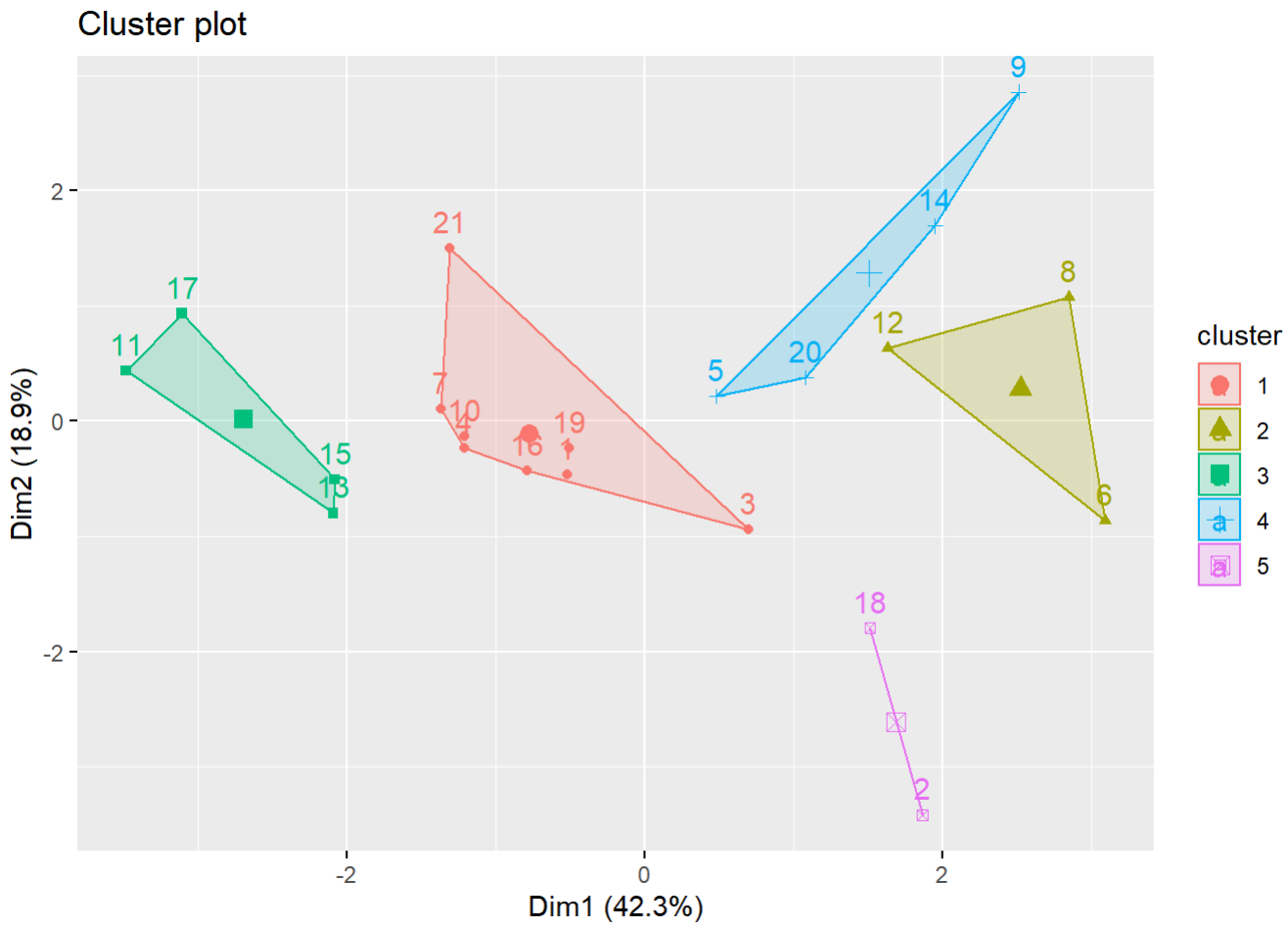

```
fviz_nbclust(Phar2, kmeans, method="silhouette") + labs(subtitle="Silhouette Method")
```




```
set.seed(3)
k5<-kmeans(Phar2, centers = 5, nstart = 25)
k5$centers
```

```
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
## 2 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 3  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 4 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 5 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.27449312 -0.7041516       0.556954446
## 2  1.36644699 -0.6912914      -1.320000179
## 3 -0.46807818  0.4671788       0.591242521
## 4  0.06308085  1.5180158      -0.006893899
## 5 -0.14170336 -0.1168459      -1.416514761
```

```
fviz_cluster (k5, data=Phar2)
```



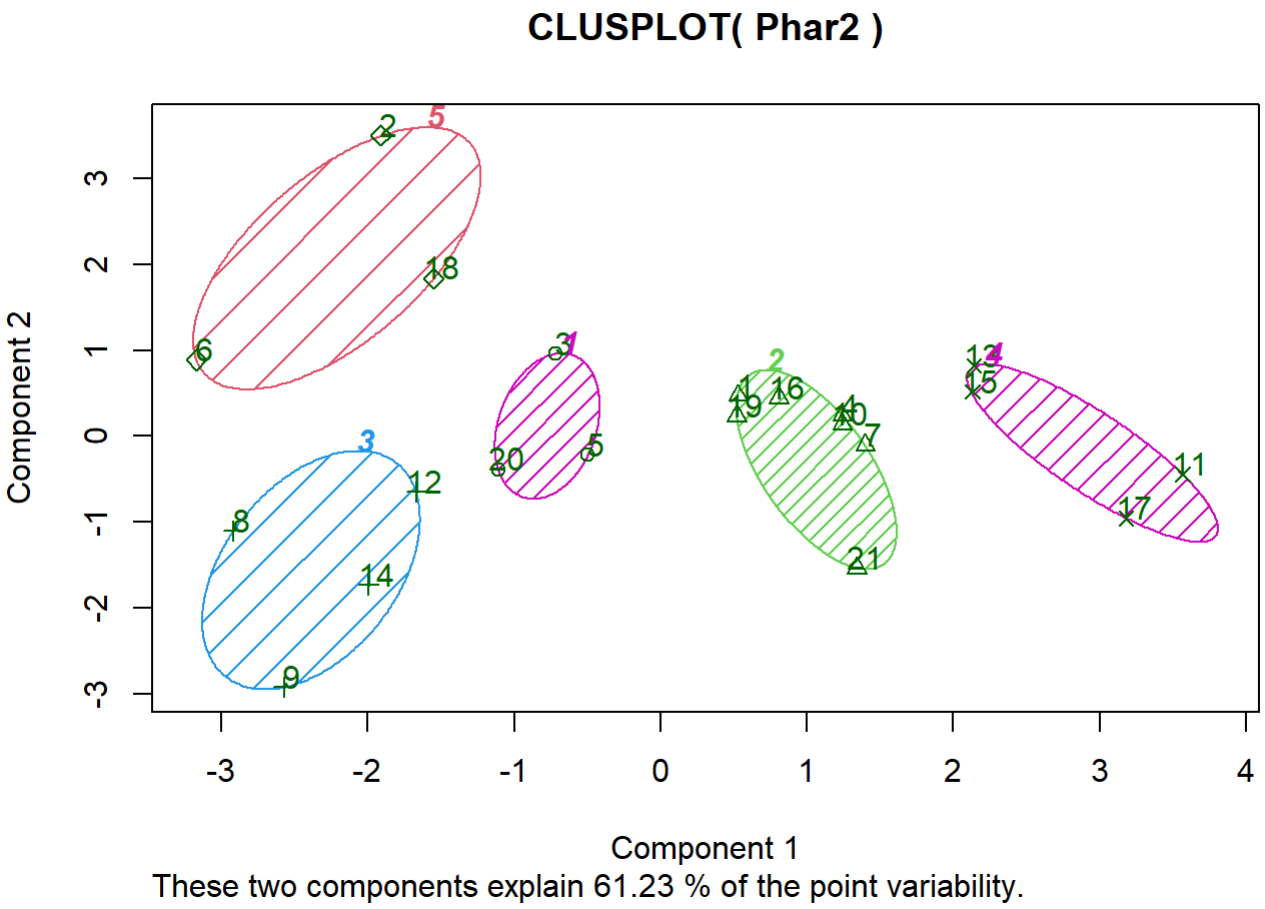

```
#Calculating the KMeans Cluster Analysis and dividing into five different clusters.
fit<-kmeans(Phar2,5)

#In this step, I will calculate the mean value of all of the numerical variables for each of the five clusters.
aggregate(Phar2,by=list(fit$cluster), FUN=mean)
```

```
##   Group.1  Market_Cap       Beta   PE_Ratio        ROE        ROA
## 1       1 -0.66114002 -0.7233539 -0.3512251 -0.6736441 -0.5915022
## 2       2  0.08926902 -0.4618336 -0.3208615  0.3260892  0.5396003
## 3       3 -0.96247577  1.1949250 -0.3639982 -0.5200697 -0.9610792
## 4       4  1.69558112 -0.1780563 -0.1984582  1.2349879  1.3503431
## 5       5 -0.52462814  0.4451409  1.8498439 -1.0404550 -1.1865838
##   Asset_Turnover   Leverage Rev_Growth Net_Profit_Margin
## 1  -1.537552e-01 -0.4040831  0.6917224        -0.4005718
## 2   6.589509e-02 -0.2559803 -0.7230135         0.7343816
## 3  -1.153164e+00  1.4773718  0.7120120        -0.3688236
## 4   1.153164e+00 -0.4680782  0.4671788         0.5912425
## 5   1.480297e-16 -0.3443544 -0.5769454        -1.6095439
```

```
#Adding to the clustering sequence.
Phar3<- data.frame(Phar2, fit$cluster)
Phar3
```

```
##    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
## 7  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## 8  -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## 9  -0.9704532  2.15893320 -1.34037772 -0.70899938 -1.0174553     -1.8450624
## 10  0.2762415 -1.34655112  0.14948233  0.34502953  0.5610770     -0.4612656
## 11  1.0999201 -0.68440408 -0.45749769  2.45971647  1.8389364      1.3837968
## 12 -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.6979905     -0.4612656
## 13  1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544      0.9225312
## 14 -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -1.8450624
## 15  1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577      1.8450624
## 16  0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -0.9225312
## 17  2.4199899  0.48409069 -0.11415545  1.31287998  1.6322239      0.4612656
## 18 -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.9047030     -0.4612656
## 19 -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929      0.4612656
## 20 -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905     -0.9225312
## 21 -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849     -0.4612656
##       Leverage  Rev_Growth Net_Profit_Margin fit.cluster
## 1  -0.21209793 -0.52776752        0.06168225           2
## 2   0.01828430 -0.38113909       -1.55366706           5
## 3  -0.40408312 -0.57211809       -0.68503583           1
## 4  -0.74965647  0.14744734        0.35122600           2
## 5  -0.31449003  1.21638667       -0.42597037           1
## 6  -0.74965647 -1.49714434       -1.99560225           5
## 7  -0.02011273 -0.96584257        0.74744375           2
## 8   3.74279705 -0.63276071       -1.24888417           3
## 9   0.61983791  1.88617085       -0.36501379           3
## 10 -0.07130879 -0.64814764        1.17413980           2
## 11 -0.31449003  0.76926048        0.82363947           4
## 12  1.10620040  0.05603085       -0.71551412           3
## 13 -0.62166634 -0.36213170        0.33598685           4
## 14  0.44065173  1.53860717        0.85411776           3
## 15 -0.39128411  0.36014907       -0.24310064           4
## 16 -0.67286239 -1.45369888        1.02174835           2
## 17 -0.54487226  1.10143723        1.44844440           4
## 18 -0.30169102  0.14744734       -1.27936246           5
## 19 -0.74965647 -0.43544591        0.29026942           2
## 20 -0.49367621  1.43089863       -0.09070919           1
## 21  0.68383297 -1.17763919        1.49416183           2
```

```
#Here I am going to visualize the cluster plot provided.
library(cluster)
clusplot(Phar2,fit$cluster,color = TRUE,shade = TRUE, labels= 2, lines=0)
```


CLUSPLOT(Phar2)
Component 2
Component 1
These two components explain 61.23 % of the point variability.

```
# b.)Interpret the clusters with respect to the numerical variables used in forming the cluster
s.
#Cluster 1: Rows 3,5,20
#Cluster 2: Rows 1,4,7,10,16,19,21
#Cluster 3: Rows 8,9,12,14
#Cluster 4: Rows 11,13,15,17
#Cluster 5: Rows 2,6,18

# When looking at the mean values for all of the numerical values for each cluster it can be con
cluded that:
#Cluster 1 has the lowest Beta, lowest Asset_Turnover
#Cluster 2 has the highest Asset_Turnover, lowest Rev_Growth, highest Net_Profit_Margin
#Cluster 3 has the lowest Market_Cap,highest Beta,lowest PE_Ratio, highest leverage, highest Rev
_Growth
#Cluster 4 has the highest Market_Cap, highest ROE, highest ROA, lowest leverage
#Cluster 5 has the highest PE_Ratio, lowest ROE, lowest ROA, lowest Net_Profit_Margin

# c.) Is there a pattern in the clusters with respect to the numerical variables? #There is a pa
ttern with the Median_Recommendation Variable given that:

#Cluster 1 has the lowest Asset_Turnover which results in mostly Moderate Buy Recommendations.
#Cluster 2 has the highest Net_Profit which results in mostly a Hold Recommendations.
#Cluster 3 has the highest Beta, highest leverage, highest Rev_Growth which results in mostly, a
Moderate Buy Recommendations.
#Cluster 4 has the highest Market_Cap, highest ROE, highest ROA which results in an equal Hold a
nd Moderate Buy Recommendations.
#Cluster 5 has the highest PE_Ratio which results in mostly Hold Recommendations.

#Can we see a certain pattern among the clusters with the variables (10 to 12)?

#Clusters 1,3 have mostly Moderate Buy Recommendations.
#Clusters 2,5 have mostly Hold Recommendations.

#d.) Provide an appropriate name for each cluster using any or all of the variables in the datas
et.

#Cluster 1 has the lowest Beta, lowest Asset_Turnover
#Cluster 2 has the highest Asset_Turnover, lowest Rev_Growth, highest Net_Profit_Margin
#Cluster 3 has the lowest Market_Cap,highest Beta,lowest PE_Ratio, highest leverage, highest Rev
_Growth
#Cluster 4 has the highest Market_Cap, highest ROE, highest ROA, lowest leverage
#Cluster 5 has the highest PE_Ratio, lowest ROE, lowest ROA, lowest Net_Profit_Margin
```