Team Assignment: Pharmaceutical Industry

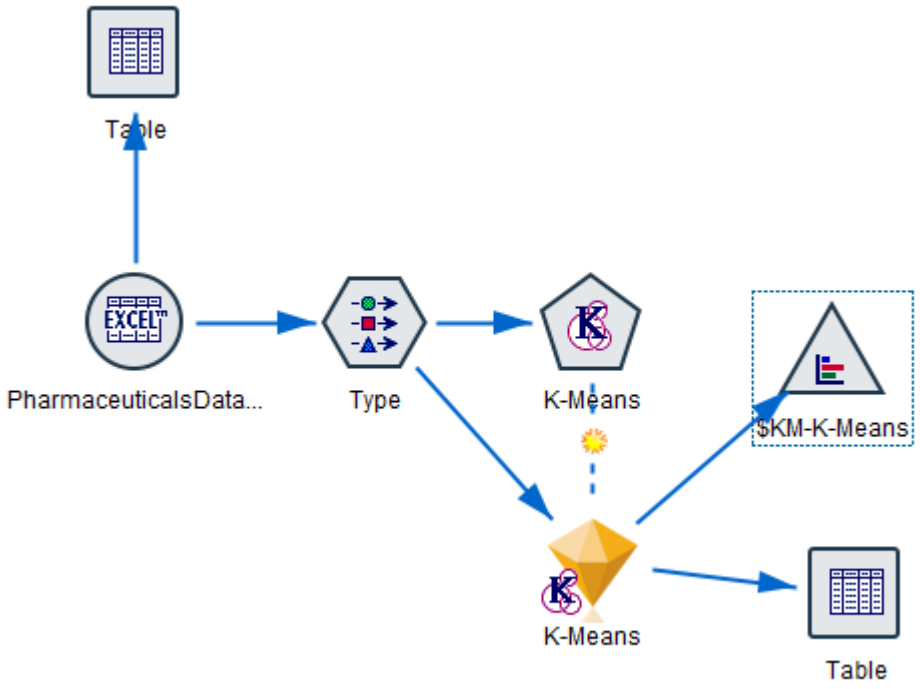
Use cluster analysis to explore and analyze the given dataset as follows:

a) Use the quantitative variables (1 to 9) to cluster the 21 firms. Justify the various choices made when conducting the cluster analysis

Tip: In this case, to understand the structure of the industry in financial terms we might use basic domain knowledge to consider that size and financial performance are fundamental dimensions into which our various metrics fall. To make sense of our analysis, we will need at least two clusters, but more than 3 or 4 will probably defeat the purpose of clustering

NOTE: In order to have a uniform criterion for comparison and although you can run several algorithms and test them with different number of clusters, I would like you to focus on 4 clusters

The following stream is created in SPSS modeler for clustering 21 firms in the given pharmaceuticalsdata.xlsx sheet



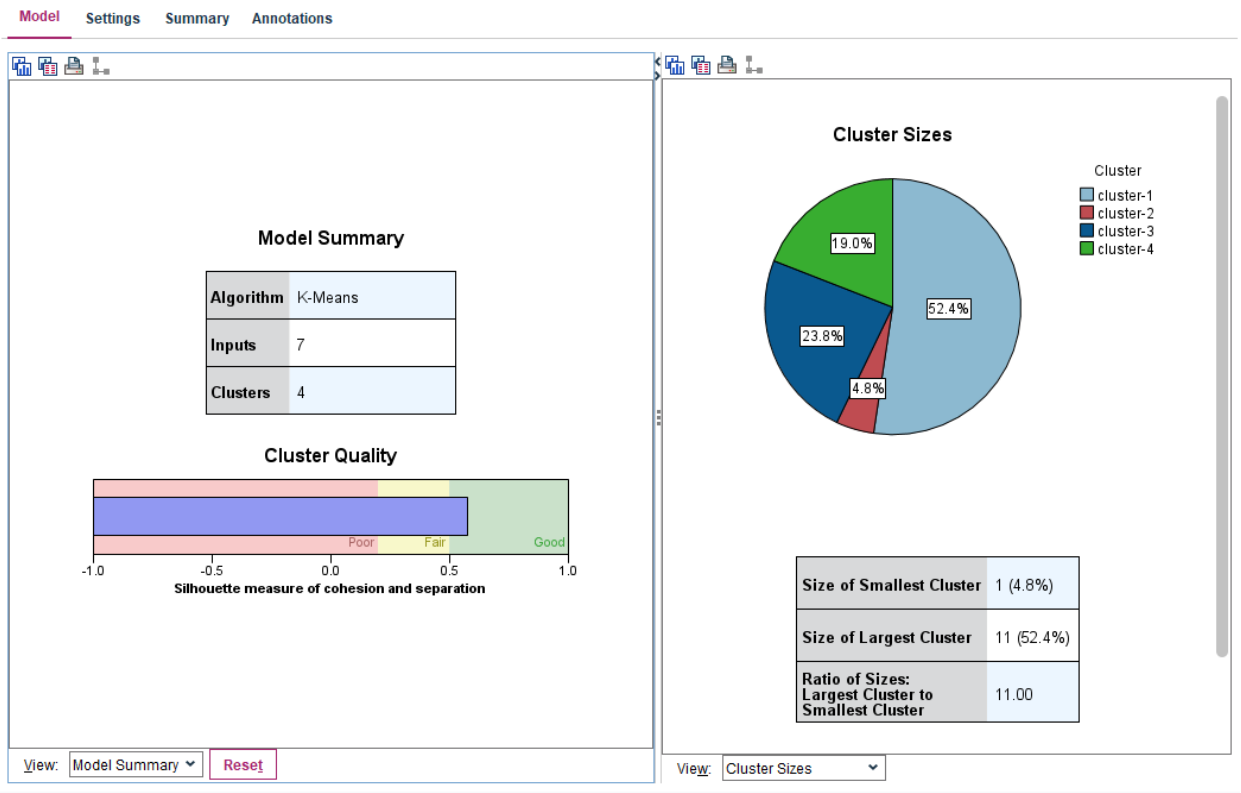
Below picture shows the quantitative variables chosen for clustering. 7 attributes- PE-Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth and Net_Profit_Margin were considered in clustering as they appear to be significant.

Types | Format | Annotations

Read Values | Clear Values | Clear All Values

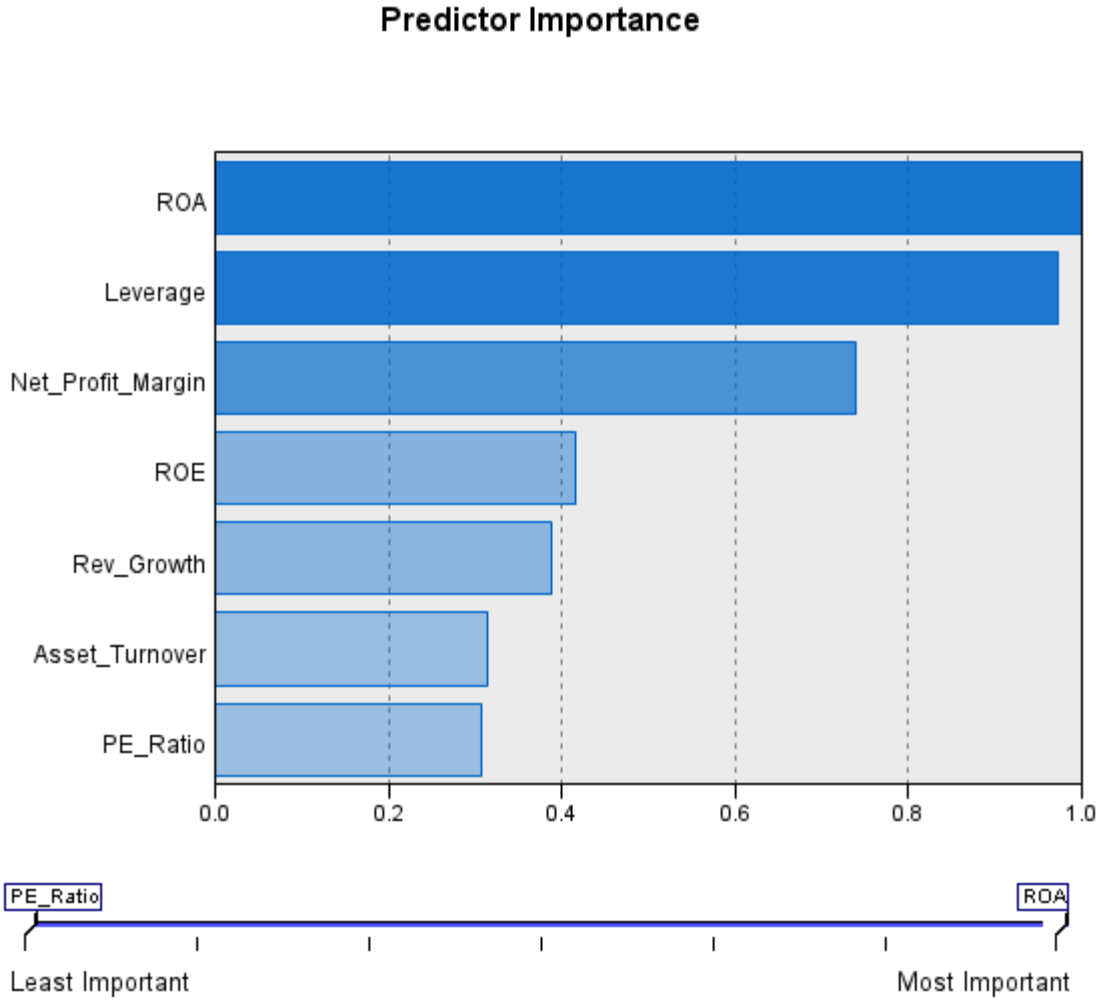
Field	Measurement	Values	Missing	Check	Role
Symbol	Nominal	ABT,AGN,AHM,A...		None	None
Name	Nominal	"Abbott Laborat...		None	None
Market_Cap	Continuous	[0.41,199.47]		None	None
Beta	Continuous	[0.18,1.11]		None	None
PE_Ratio	Continuous	[3.6,82.5]		None	Input
ROE	Continuous	[3.9,62.9]		None	Input
ROA	Continuous	[1.4,20.3]		None	Input
Asset_Turnover	Continuous	[0.3,1.1]		None	Input
Leverage	Continuous	[0.0,3.51]		None	Input
Rev_Growth	Continuous	[-3.17,34.21]		None	Input
Net_Profit_Margin	Continuous	[2.6,25.5]		None	Input
Median_Recommendation	Nominal	Hold,"Moderate ...		None	None
Location	Nominal	CANADA,FRAN...		None	None
Exchange	Nominal	AMEX,NASDAQ,...		None	None

The following is the output when the stream is executed under K-means node.



The 4 clusters induced by 7 inputs gave relatively good cluster quality 0.6, which is silhouette measure of cohesion and separation. The pie chart shows the cluster distribution with 1 record as smallest cluster and 11 records as largest cluster bearing 52.4% of total records.

Below is the predictor importance histogram, shows that ROA, leverage and Net_profit_margin appear to be most relevant attributes.



b) Interpret the clusters with respect to the quantitative variables that were used in forming the clusters

The figure shows 4 generated clusters with ROA, leverage and Net_profit_margin to be most relevant attributes.

There seems to be a significant variation (not too much) of all the attributes except asset_turover which is between 0.8 as highest and 0.46 as lowest value cluster. This significant variation shows that there is no too fitting in between the clusters and they are mostly well defined different from each other.

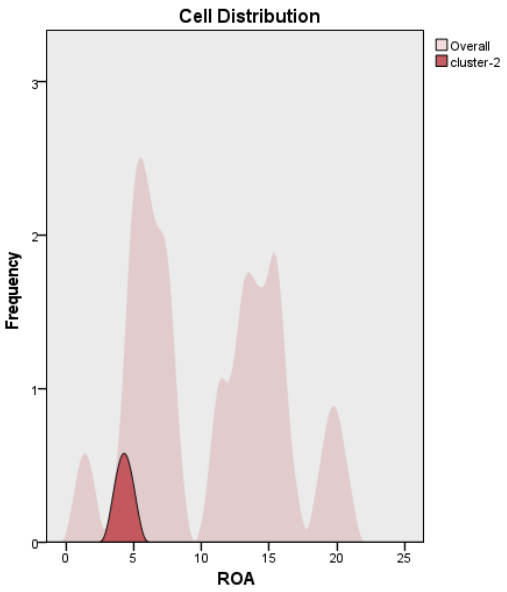
Clusters

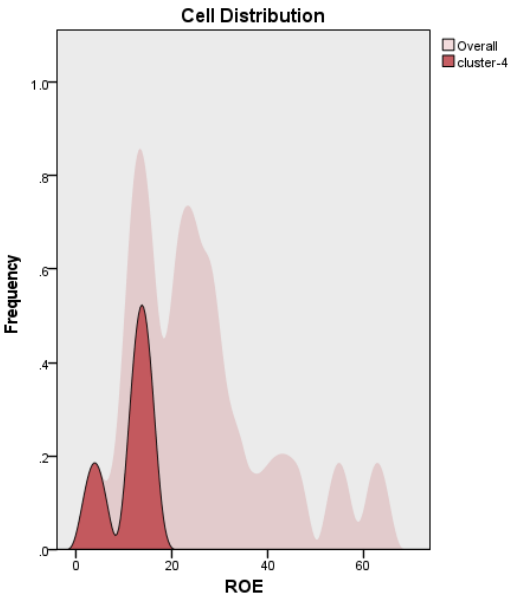
Input (Predictor) Importance

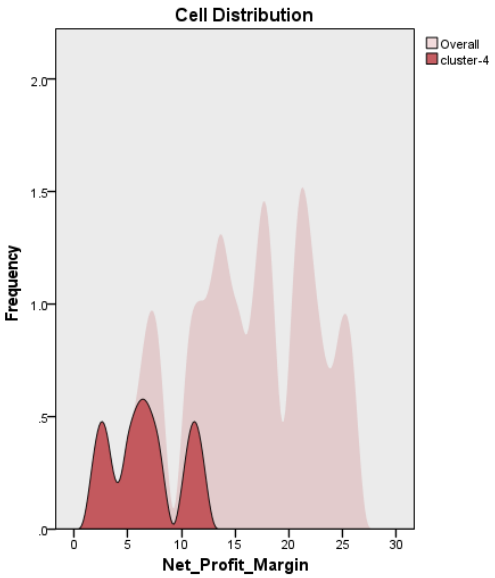
1.0 0.8 0.6 0.4 0.2 0.0

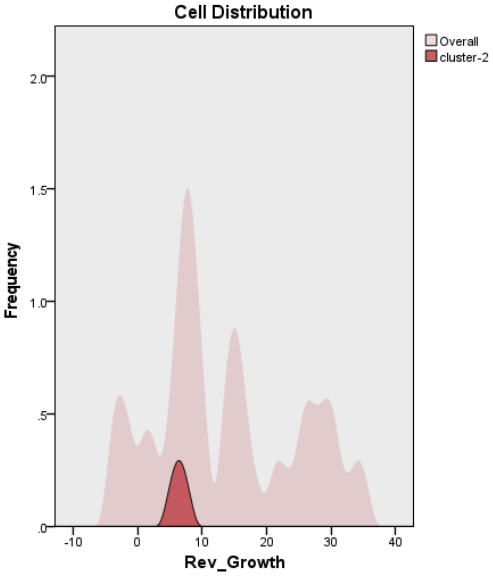
Cluster	cluster-1	cluster-3	cluster-4	cluster-2
Label				
Description				
Size	52.4% (11)	23.8% (5)	19.0% (4)	4.8% (1)
Inputs	ROA 14.95	ROA 6.32	ROA 5.10	ROA 4.30
	Leverage 0.33	Leverage 0.80	Leverage 0.30	Leverage 3.51
	Net_Profit_Margin 20.17	Net_Profit_Margin 14.72	Net_Profit_Margin 6.65	Net_Profit_Margin 7.50
	ROE 35.70	ROE 15.94	ROE 11.30	ROE 24.10
	Rev_Growth 10.16	Rev_Growth 26.91	Rev_Growth 7.01	Rev_Growth 6.38
	Asset_Turnover 0.80	Asset_Turnover 0.46	Asset_Turnover 0.75	Asset_Turnover 0.60
	PE_Ratio 20.95	PE_Ratio 18.12	PE_Ratio 46.90	PE_Ratio 26.00

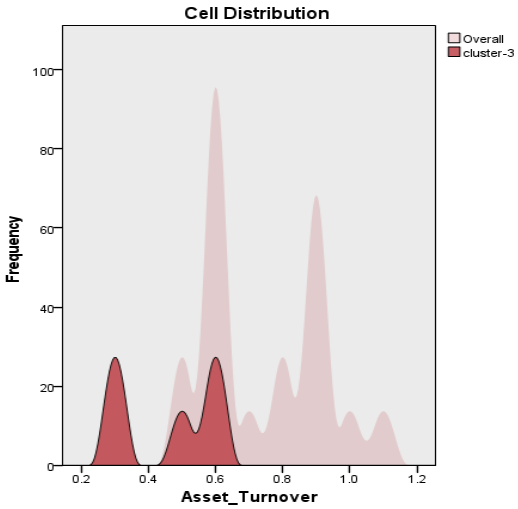
Let's compare the lowest values against cell distribution. The ROA (with importance as 1) shows below graph of **minimum**, cluster 2 against frequency cell distribution of remaining clusters. Similarly, the remaining figures show the frequency distribution of all the attributes against **minimum** cluster size.


Cell Distribution
Overall
cluster-2
Frequency
ROA

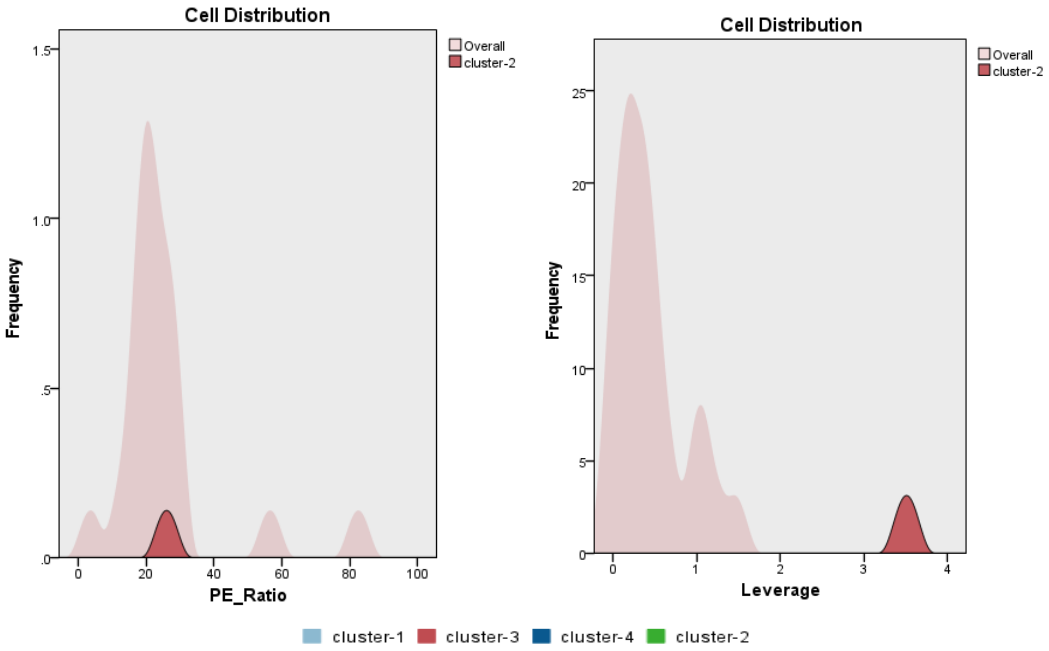

Cell Distribution
Overall
cluster-4
Frequency
ROE

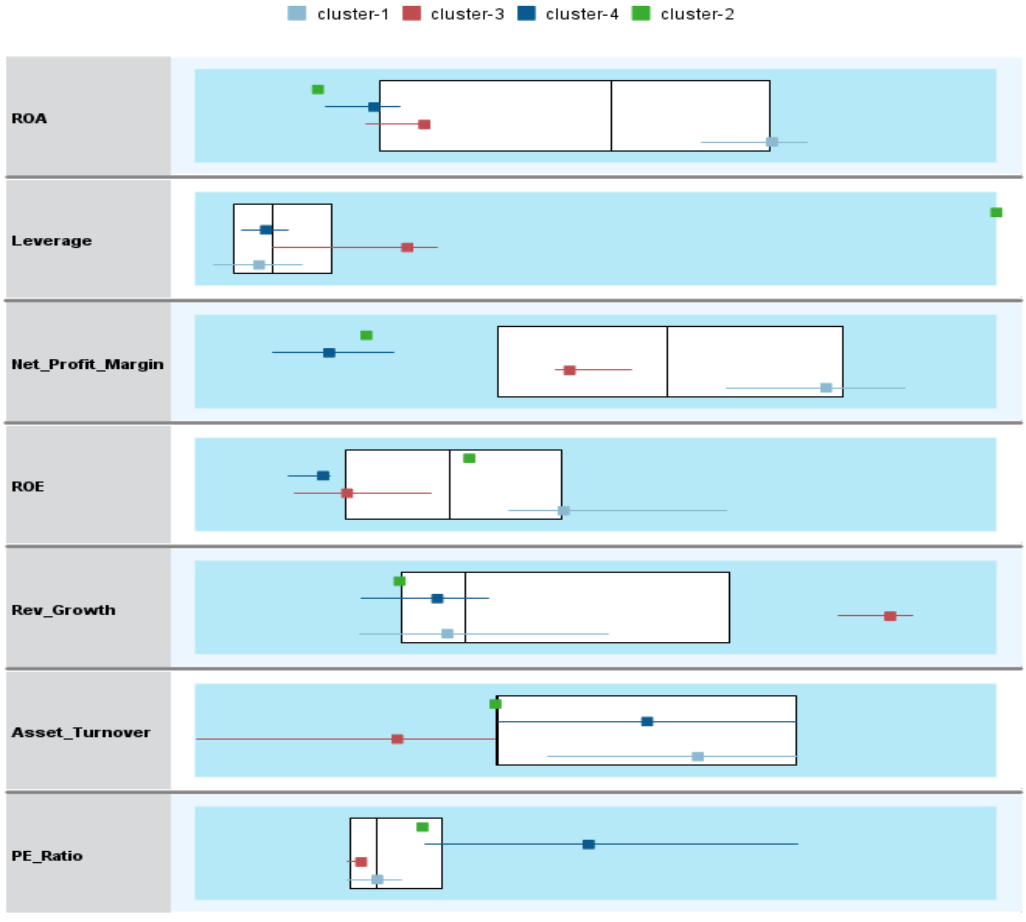

Cell Distribution
Overall
cluster-4
Frequency
Net_Profit_Margin


Cell Distribution
Overall
cluster-2
Frequency
Rev_Growth


Cell Distribution
Overall
cluster-3
Frequency
Asset_Turnover

The below figures show the Leverage frequency distribution of maximum leverage (0.97 importance) of **maximum** of 3.51 of cluster 2 and PE_ratio with **maximum** of 26.00.





The above figure shows the cluster comparison. We could see that there is no similarity among one attribute to another.

c) Is there a pattern in the clusters with respect to the qualitative variables?

There is no defined pattern in the clusters as from the comparison figure we could see that there is no particular tend followed by the clusters with respect to any attributes.

Also, the Cluster table, in answer b shows that the minimums and maximums are varied among the clusters and there is no particular order or trend in the clusters.

d) Provide an appropriate name for each cluster using any or all of the variables in the dataset

Renaming the cluster as – Cluster 1: 1_ROA

Cluster 2: 4_Leverage

Cluster 3: 2_Rev_Growth

Cluster 4: 3_PE_Ratio

The first part in the name denotes the order of size of the cluster and second part denotes the high attribute among other qualitative variabes.