

Assignment 4: Cluster Analysis

Overview: An equities analyst is studying the pharmaceutical industry and would like your help in exploring and understanding the financial data collected by her firm. Her main objective is to understand the structure of the pharmaceutical industry using some basic financial measures.

Financial data were gathered on 21 firms in the pharmaceutical industry, and is available in the file Pharmaceuticals.csv. For each firm, the following variables are recorded:

- (a) market capitalization (in \$billion)
- (b) beta
- (c) price/earnings ratio
- (d) return on equity
- (e) return on assets
- (f) asset turnover
- (g) leverage
- (h) estimated revenue growth
- (i) net profit margin
- (j) median recommendation (across major brokerages)
- (k) location of firm's headquarters
- (l) the stock exchange on which the firm is listed

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

1. (30) Use only the quantitative variables (a)-(i) to cluster the 21 firms. Provide final result about number of clusters chosen and cluster memberships.
2. (25) Justify the choices made in conducting the cluster analysis (in the previous question) such as weights accorded different variables, the specific clustering algorithm/s used, the number of clusters formed, etc.
3. (25) Interpret the clusters with respect to the quantitative variables that were used in forming the clusters.
4. (20) Provide an appropriate name for each cluster using key variables in the dataset.

Deliverable: Submit a R Markdown file that contains R codes, output and responses as required. The file submitted should be Word or PDF.