



Assignment 9

Clustering Stocks

Applying PCA and K-Means Clustering to Identify Similar Stocks

Use the data on the G-Drive sourced from [Kaggle](#), to group together stocks. The student should choose the most recent period e.g. ~December 2016 period of reporting and choose what data fields they want to include. The student should apply PCA and justify the number of dimensions they choose, then use K-Means, or another clustering technique to group stocks together. The student should also justify the number of clusters chosen.

In addition to the analysis, please identify and then explain the differences in clusters.