# Using the K-means Clustering Analysis for A Unleveraged Portfolio

Fangchi Wu, Lantian Jia, Wenbo Yu Instructor: Dragos Bozdog



Business Intelligence & Analytics

#### Introduction

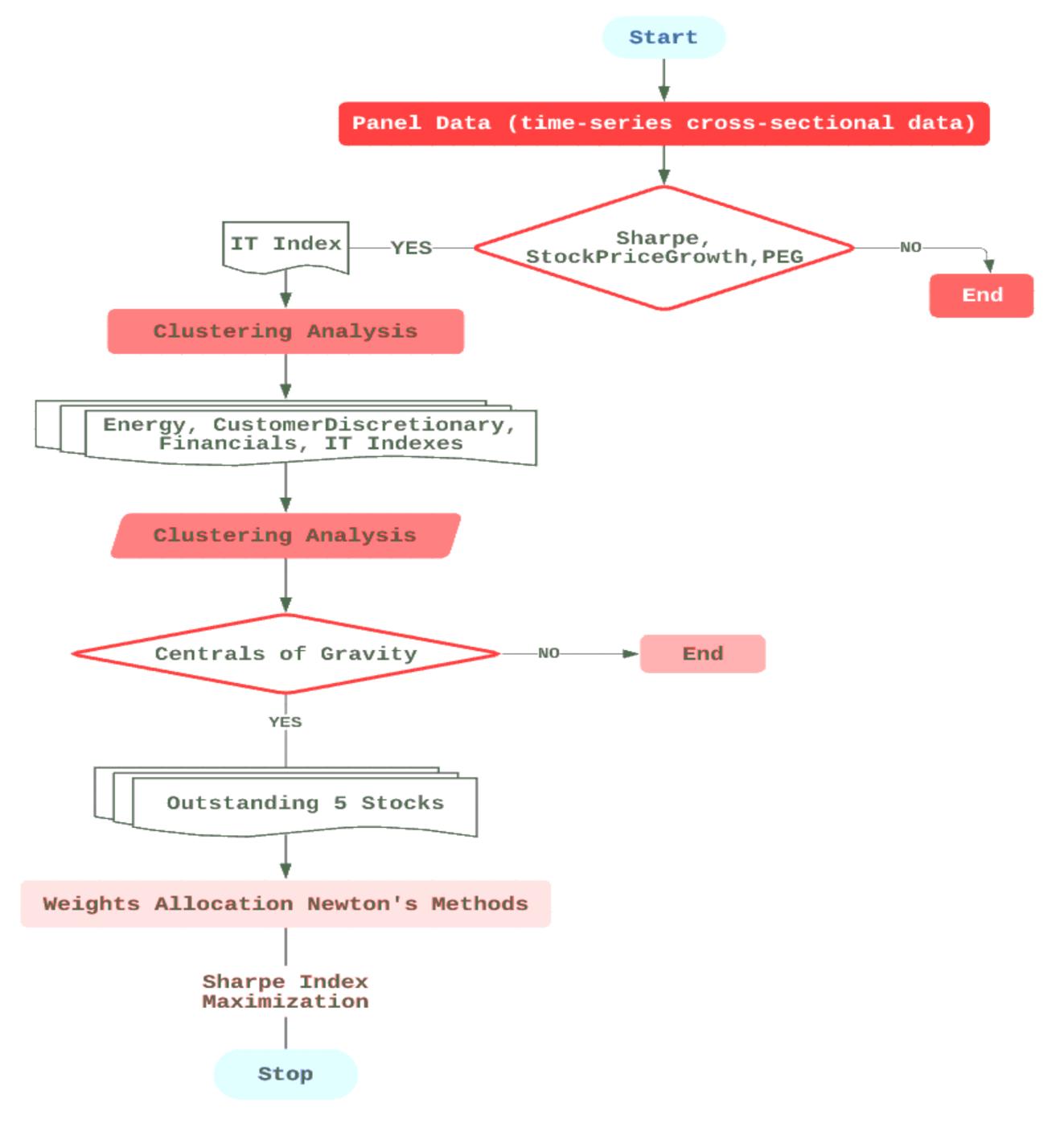
People of different ages maybe tend to have different investment choices in the stock market. In general, older people are more likely to choose conservative stocks to invest in, while younger prefer to invest in high-yield stocks. The risk and benefit are two important factors in stock investment. Young people in the 20-30 age range are becoming more conservative because of the limited amount of funds, and they focus on risks as well as benefits. For this practical problem, we will research the portfolio of stock in the project, which can provide some advice for those young people who want to maximize profits and minimize risks.

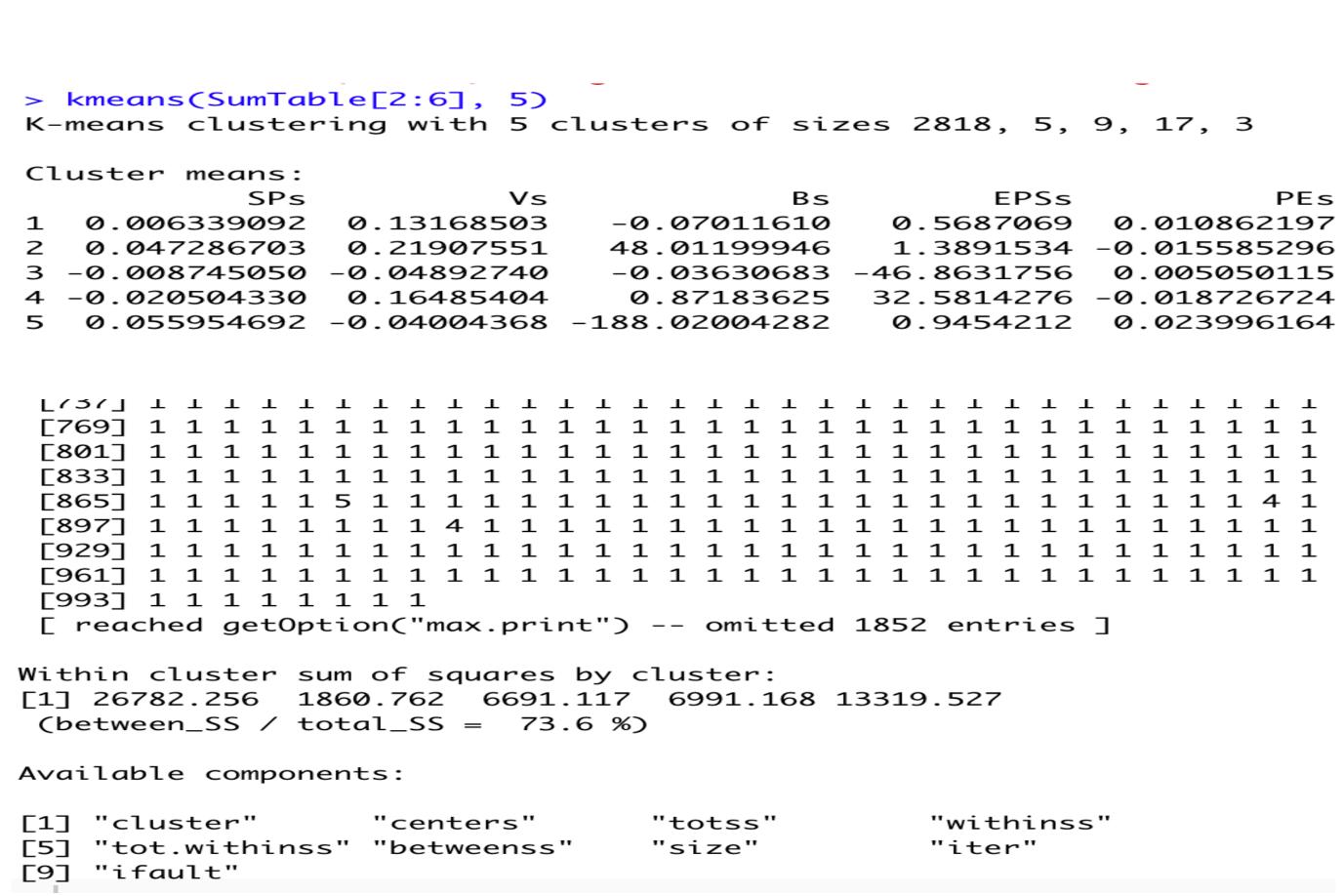
#### Data Process

Data Integration. After integrating the original data with the calculated data, we obtain a data set that totals 11 dimensional variables.

Handling outliers and missing values.

Standardizing data.





## Method and Technology

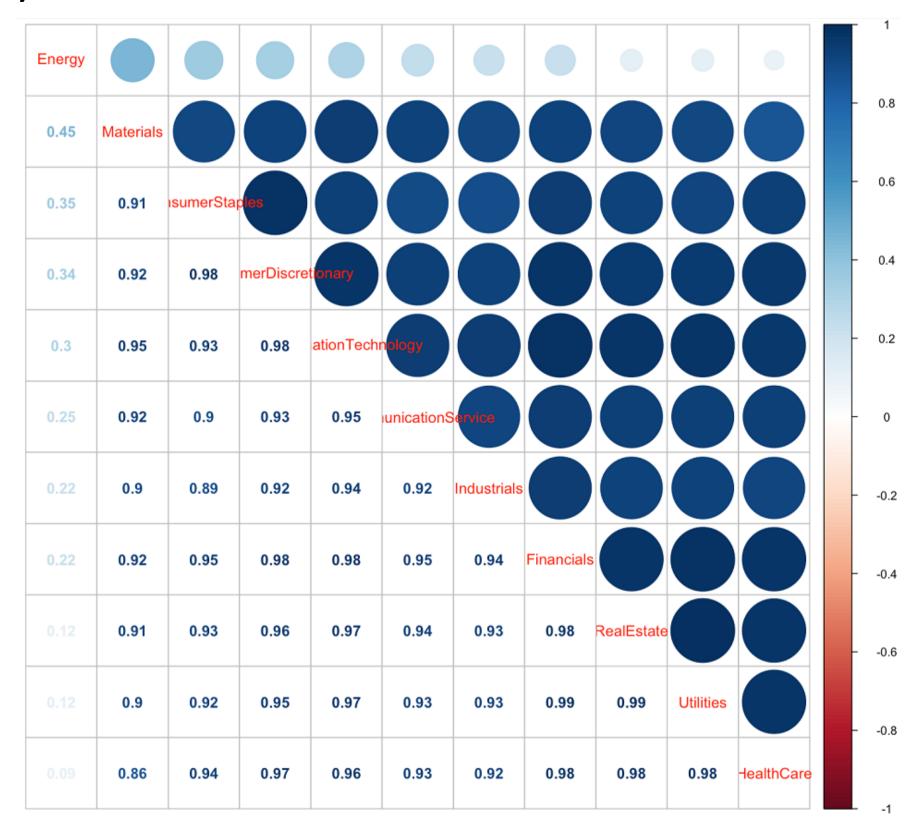
Pearson correlation coefficient analysis

K-means cluster analysis

Newton method to find the optimal solution of weight combination

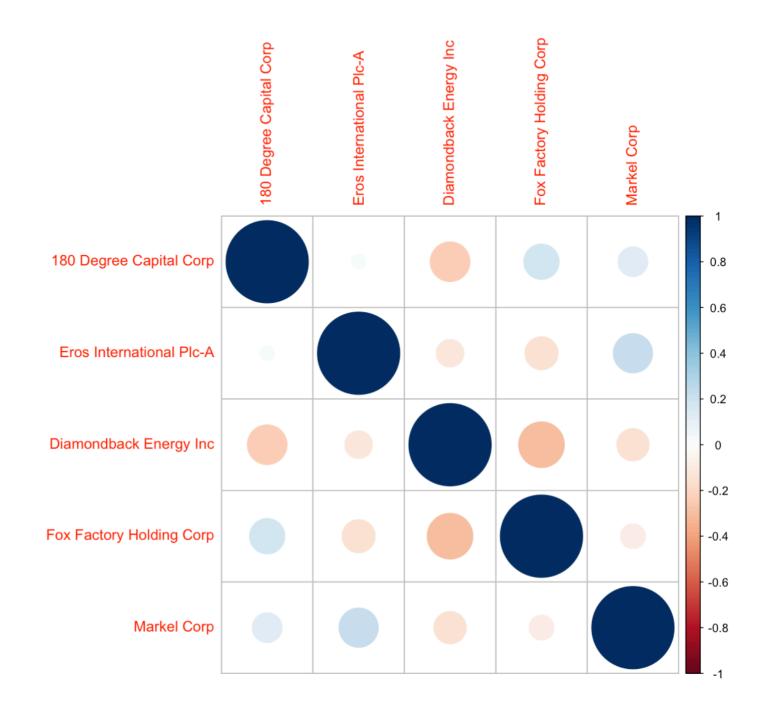
### Results and Conclusion

The correlation coefficient between the rise and fall of the total stock price of each industry is obtained.



The k-means clustering of all stocks in 4 specific industries (The overall stock price rise negative correlation or positive correlation) is conducted to divide them into 5 clusters. And we get the center of gravity for each cluster.

According to the defined distance formula, the most representative stock is selected from each cluster, which is mathematically the point closest to the center of gravity of each cluster. They are 180 Degree Capital Corp, Eros International Plc-A, Diamondback Energy Inc, Fox Factory Holding Corp, Markel Corp.



Using machine learning, I made a portfolio of these five stocks, established the weights and reached the best sharpe ratio. The vector of weights is (0.0065%, 31.19%, 9.0%, 0.023%, 59.772%)



After deducting the risk-free return, the expected return of the portfolio is 5.7818%, the standard deviation is 2.45494% and the sharpe ratio is 2.315.