

Assignment Session 3

Student ID: 20202027
Name: Thu Phuong Nguyen

1 K-mean clustering

The results of clustering are as below:

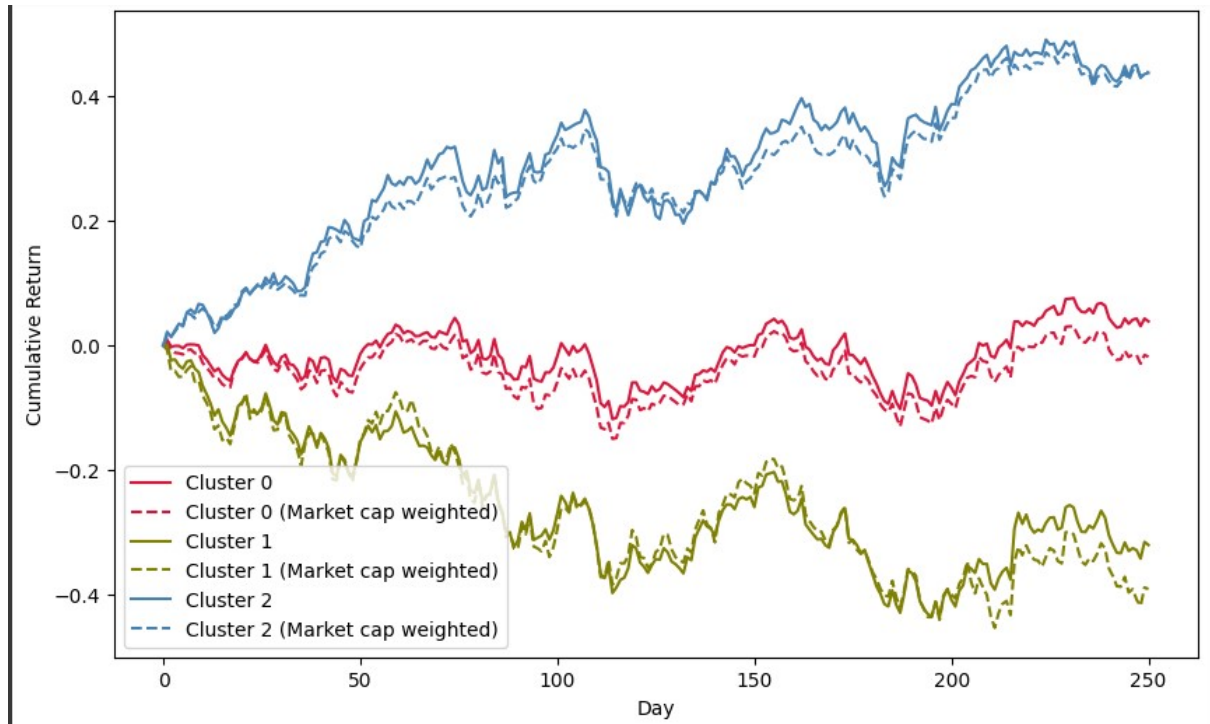


Figure 1: K-mean clustering

The results of F-F 3 Factors are as below:

1.1 Cluster 0

Cluster 0 comprises of established companies in traditional industries such as manufacturing and utilities. These companies tend to have stable growth rates and are considered less volatile compared to the technology companies in Cluster 1. It comprises companies with significant market capitalizations, indicating their substantial presence in the market. Representative stocks in this cluster include Apple Inc. (AAPL) with a market capitalization of \$ 2908.55 billion, Johnson & Johnson (JNJ) with a market capitalization of \$ 426.78 billion, Walmart (WMT) with a market capitalization of \$ 418.81 billion, and Coca-Cola Company (KO) with a market capitalization of \$ 266.69 billion. With their diversified product portfolio, they have managed to establish a strong market position and attract a large customer base. Hence, these companies are typically recognized for their stability and market presence. These companies have a history of stable performance and may be less prone to market fluctuations compared to smaller-cap or more volatile stocks. When comparing Cluster 0 with the F-F 3 factors, it is likely that the stocks exhibit lower market risk (beta) due to their established nature. Cluster 0 resembles the SMB (Small Minus Big) factor, which oscillates around zero during the given period. Furthermore, given their large market capitalizations, these stocks may have a smaller size factor, suggesting that

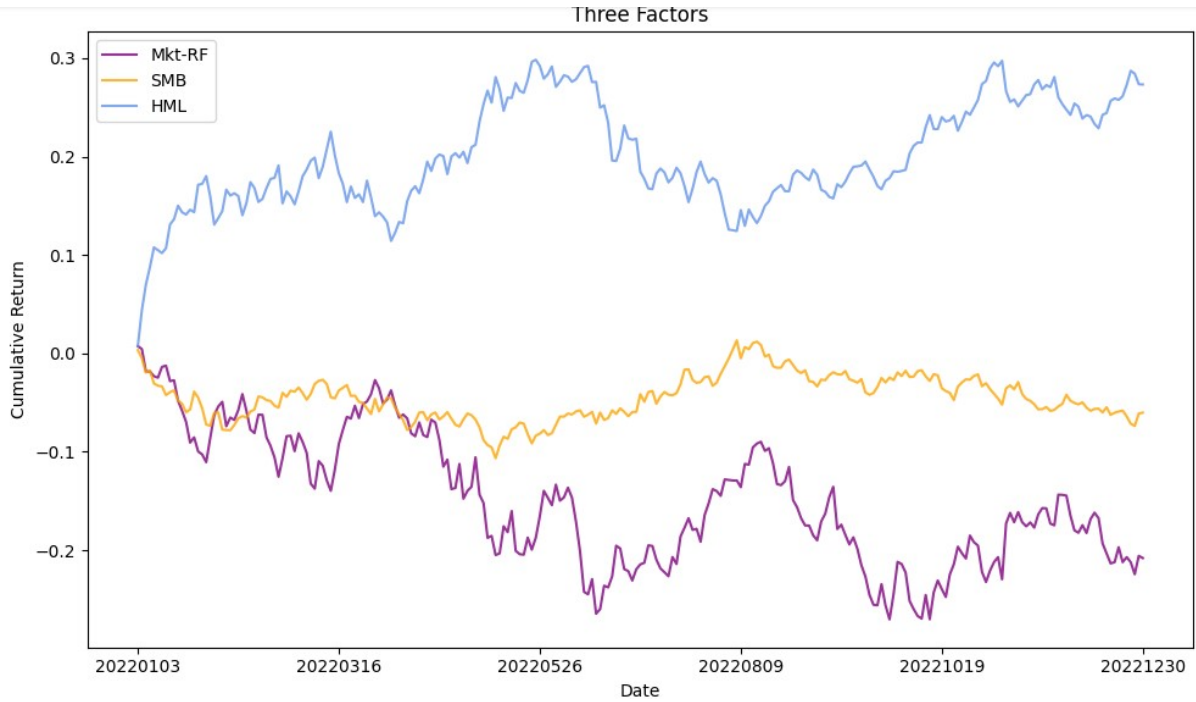


Figure 2: Fama-French 3 Factors

their performance may be influenced by other factors such as profitability, dividends, or industry-specific dynamics.

1.2 Cluster 1

Cluster 1 consists of technology companies with higher market capitalizations. These companies are predominantly from the technology and entertainment sectors. Representative stocks in this cluster include Nvidia (NVDA) with a market capitalization of \$ 1054.49 billion, Tesla, Inc. (TSLA) with a market capitalization of \$ 825.78 billion, Meta Platforms (META) with a market capitalization of \$ 720.12 billion, and Netflix (NFLX) with a market capitalization of \$ 192.02 billion. It is likely that these stocks exhibit higher market risk (beta) due to their association with the technology sector and their potential for volatility. These companies are frequently characterized by their potential for significant growth and the significantly high market capitalizations of the stocks in this cluster are because of their substantial market presence. In relation to the Fama-French 3 factors, Cluster 1's resemblance to the Mkt-RF factor suggests that the cluster's performance is driven by overall market movements. The stocks within Cluster 1 tend to follow the trends of the broader market during the given period, reflecting their sensitivity to market-wide factors and sentiment.

1.3 Cluster 2

Cluster 2 comprises established companies in traditional industries. These companies represent diverse industries such as energy, telecommunications, and biotechnology and tend to have stable growth rates, and are considered less volatile compared to the technology companies. This cluster consists of companies such as Chevron Corporation (CVX) with a market capitalization of \$ 297.95 billion, T-Mobile US (TMUS) with a market capitalization of \$ 157.10 billion, Vertex Pharmaceuticals (VRTX) with a market capitalization of \$ 89.59 billion, and Marathon Petroleum (MPC) with a market capitalization of \$ 48.71 billion. Within Cluster 2, the companies' performances are likely influenced by factors specific to their sectors, such as energy prices, consumer demand for mobile services, advancements in biotech, and refining margins. When compared to the F-F 3 factors, the cluster shows a consistent upward trend during the given period, indicating that the stocks within this cluster with higher book-to-market ratios have outperformed those with lower book-to-market ratios. These companies may represent a diverse range of industries, and their market capitalizations suggest a mix of larger and mid-sized companies.

2 Other clustering methods

2.1 Kshape clustering

The results of clustering are as below 3 and we achieved a small Silhouette score of Kshape (0.06) as we can see in the figure 4:

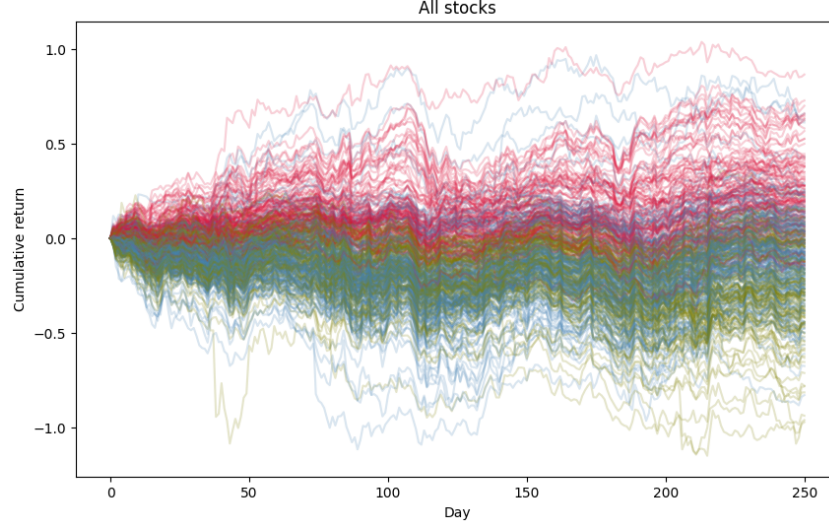


Figure 3: Kshape clustering

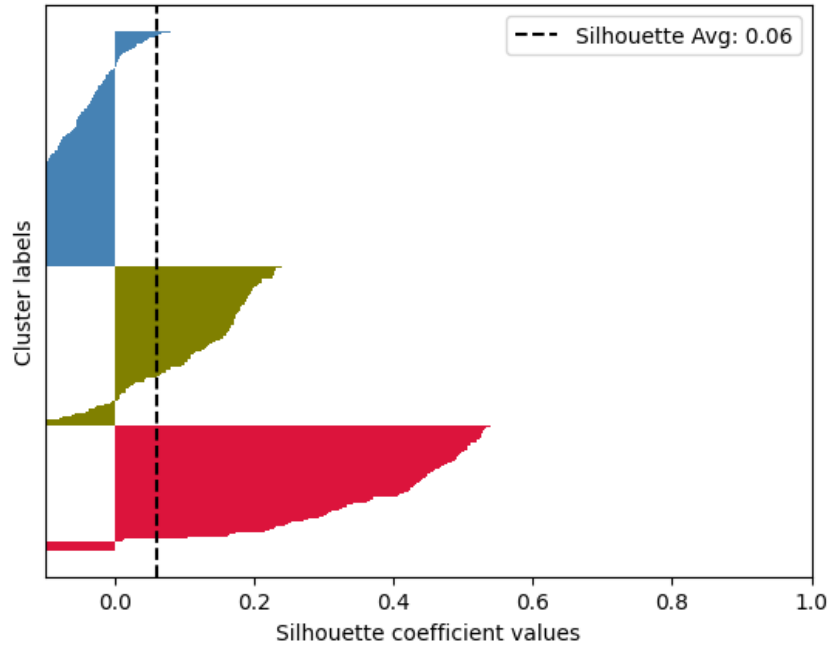


Figure 4: Silhouette scores of Kshape

When performing clustering based on S&P 500 cumulative return data, the choice between KShape is not appropriate to use for clustering this data. KShape is a clustering algorithm specifically designed for time series data. It utilizes the Dynamic Time Warping (DTW) distance measure to capture temporal dependencies and variations in the shape of time series. KShape can be effective when the temporal alignment and shape similarity between time series are crucial factors in the clustering process.

However, in the case of S&P 500 cumulative return data, K-means clustering can be a suitable choice for several reasons:

- **Linearity:** The cumulative return data for stocks in the S&P 500 index is typically represented as a linear sequence of returns over time. The Euclidean distance metric, which assumes linear relationships between dimensions, can adequately capture similarities and dissimilarities between the stock return sequences.
- **Robustness to Noise:** K-means clustering is generally more robust to noise and outliers in the data compared to KShape. Since cumulative return data can exhibit fluctuations and noise due to market volatility, K-means clustering may produce more stable and reliable results.

2.2 DTW Metrics

DTW metric is another appropriate metric for K-Means clustering other than Euclidian metric. We can see that in figure 5, the cluster centers were visualized based on the three clusters obtained from clustering, even though it has quite a big variance from the one with market cap weighted, it still can show the trend of 3 cluster groups and achieve an appropriate score (figure 6).

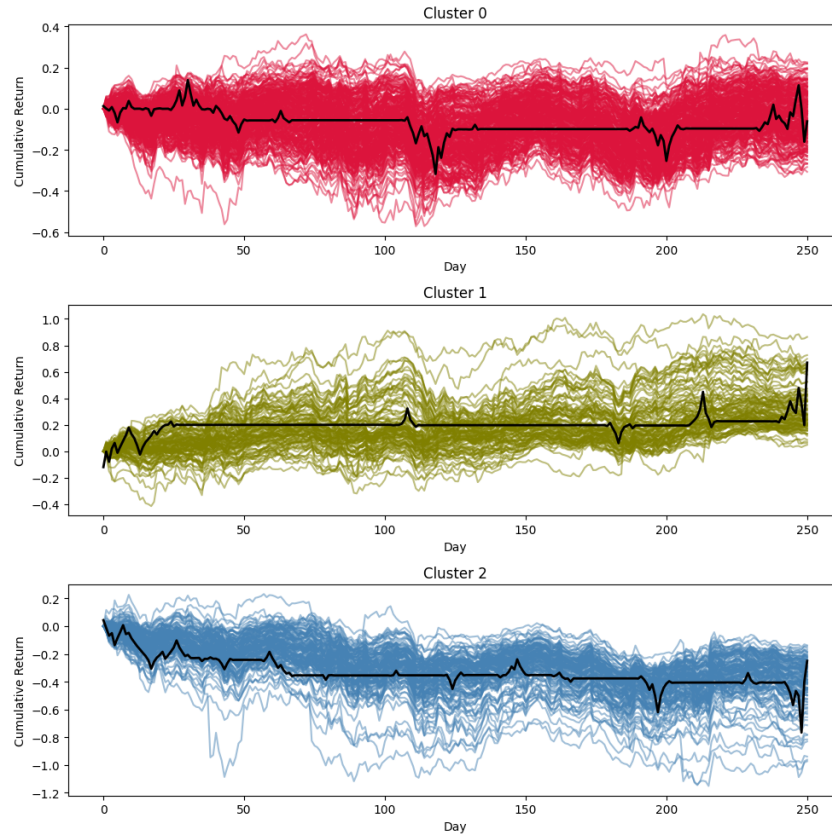


Figure 5: K-Means Clustering with DTW metric

The Euclidean metric would yield better results in clustering this cumulative return data compared to DTW for the following reasons:

- Euclidean distance calculations are less computationally intensive compared to DTW, allowing for faster processing.
- The Euclidean metric assumes linear relationships between dimensions, which can be suitable for capturing similarities in cumulative return patterns.
- The Euclidean metric measures direct point-to-point distances, enabling intuitive cluster assignments based on proximity.

3 Regression

We choose Tesla as the stock to do the analysis.

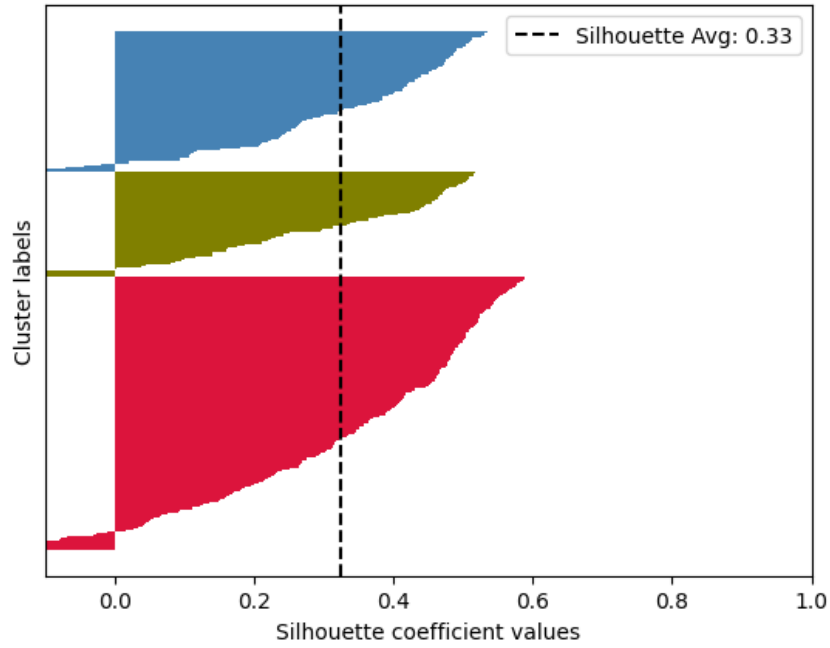


Figure 6: Silhouette scores of K-Means Clustering with DTW.

3.1 K Mean Clusters

The results of regression fitting clustering are as figure 7:

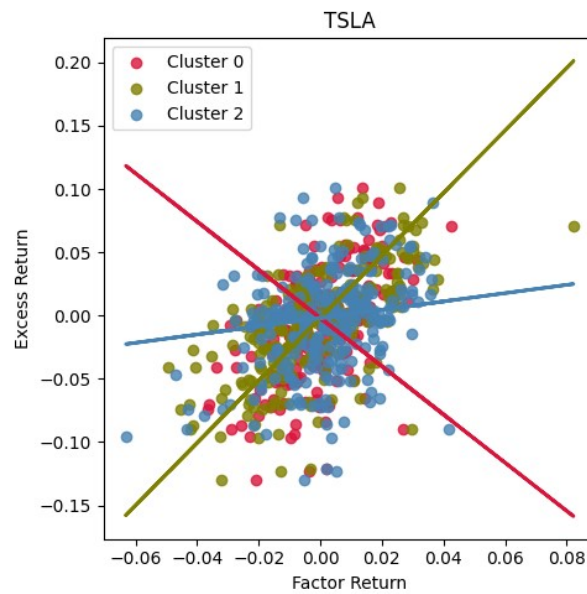


Figure 7: Regression fitting K-mean clustering

Statistics:
 Stock: TSLA
 Alpha: -0.0019
 Beta for factor 1 (Cluster 0): -1.9054
 Beta for factor 2 (Cluster 1): 2.4690
 Beta for factor 3 (Cluster 2): 0.3269
 R^2 : 0.4173

Based on the statistics, the following results were obtained:

- Beta for factor 1 (Cluster 0): The beta for Cluster 0 is -1.9054. This suggests an inverse relationship between Tesla stock and the factor represented by Cluster 0, which resembles the SMB (Small Minus Big) factor.
- Beta for factor 2 (Cluster 1): The beta for Cluster 1 is 2.4690. This indicates that Tesla stock has a positive sensitivity to the factor represented by Cluster 1, which exhibits a movement similar to the Mkt-RF (Market Risk-Free) factor. When the factor represented by Cluster 1 increases, Tesla stock is likely to experience an upward movement.
- Beta for factor 3 (Cluster 2): The beta for Cluster 2 is 0.3269. This suggests a positive relationship between Tesla stock and the factor represented by Cluster 2, which resembles the HML (High Minus Low) factor with a consistent upward trend.

The price movement of Tesla's stock will be both represented by market both SMB and Mkt-RF factors.

3.2 F-F 3 Factors

We obtained regression fitting: 3 factors figure as 8.

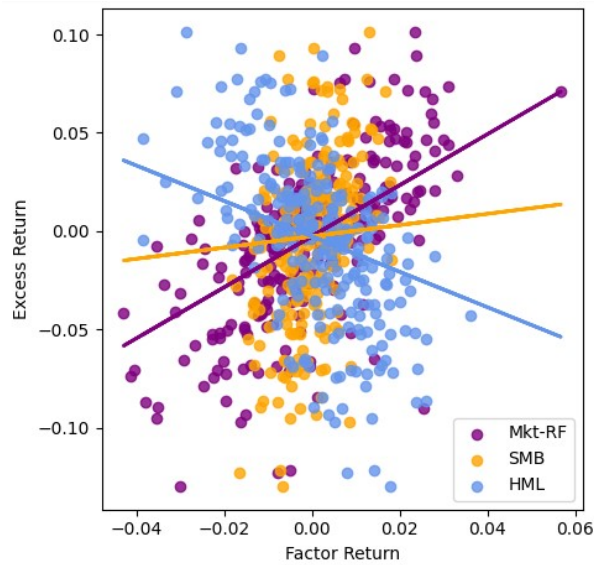


Figure 8: Regression fitting Fama-French 3 Factors

Statistics:

Stock: TSLA

Alpha: -0.0026

Beta for factor 1 (Mkt-RF): 1.2973

Beta for factor 2 (SMB): 0.2859

Beta for factor 3 (HML): -0.9002

R^2 : 0.5047

Based on the statistics, the following results were obtained:

- Beta for factor 1 (Mkt-RF): The beta for the Mkt-RF (Market Risk-Free) factor is 1.2973. This indicates that Tesla stock is positively sensitive to overall market movements represented by the Mkt-RF factor. Tesla stock is more volatile than the overall market.
- Beta for factor 2 (SMB): The beta for the SMB (Small Minus Big) factor is 0.2859. This suggests that Tesla stock has a relatively low sensitivity to the performance difference between small-cap and large-cap stocks.
- Beta for factor 3 (HML): The beta for the HML (High Minus Low) factor is -0.9002. This indicates an inverse relationship between Tesla stock and the performance difference between high book-to-market ratio stocks and low book-to-market ratio stocks (value stocks vs. growth stocks). When the HML factor increases, Tesla stock is likely to experience a decrease, and vice versa.

- R^2 : The R-squared value indicates the degree to which the factors explain Tesla stock's price movement, with 50.47% of the variation in Tesla stock being accounted for by the factors analyzed.

In this fitting progress, the components that contribute to the movement of Tesla's stock price are Mkt-RF and the High Minus Low factors, which are slightly different than the insights we got with the KMean Cluster method.