

Stock Superclusters

Samy Palaniappan
Metis Project-4
Feb 2020, Chicago



Goal

Identify stocks that have a yield better than the S&P 500.

Goal

Identify stocks that have a yield better than the S&P 500.

Why ?



Dolla dolla bill ya'll...
- Akon



Background Story

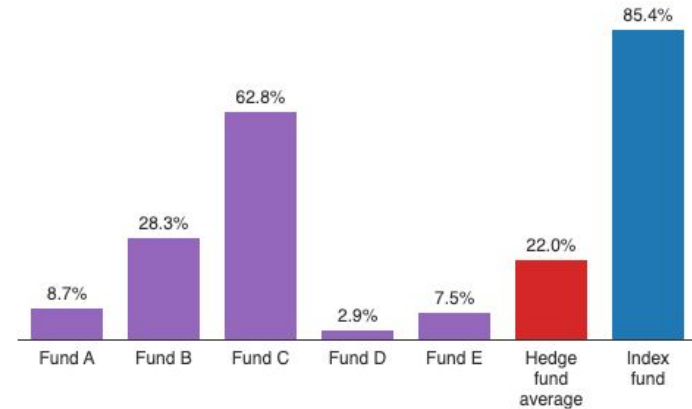
Background Story

The Bet of the Decade..



Year 9: cumulative returns

2008 through 2016



The funds of funds have not been named publicly; the index fund is Vanguard's S&P 500 Admiral fund

Goal - why not ? we get the satisfaction of proving Buffet wrong...



SOoo Much Data



► Methodology



Topic Analysis on Fin. News

CountVectorizer +
LDA

TF-IDF + LSA / NMF



Senti Analysis on Relevant Articles

Vader Sentiment +

Loughran and
McDonald, 2011

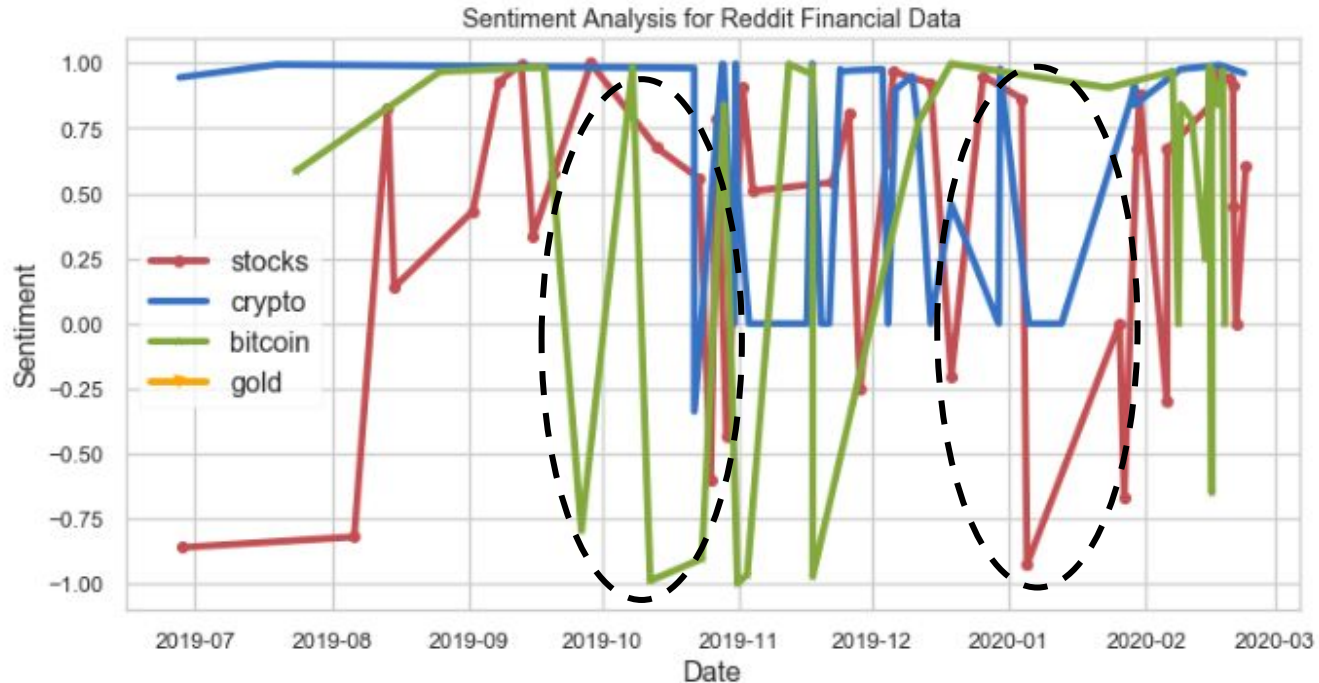


Cluster Average Daily Price + SA

KMeans

DB Scan

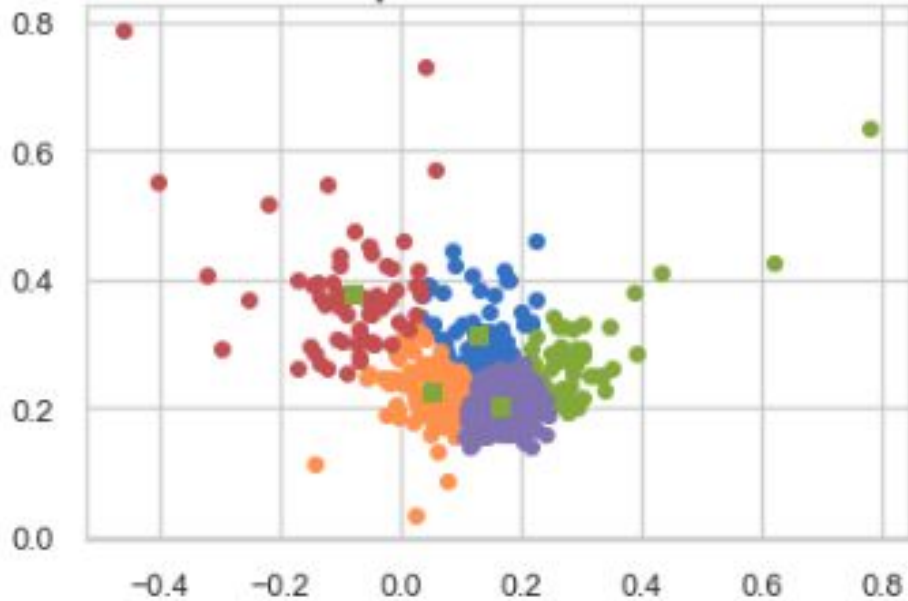
Results - Sentiment Analysis (SA)



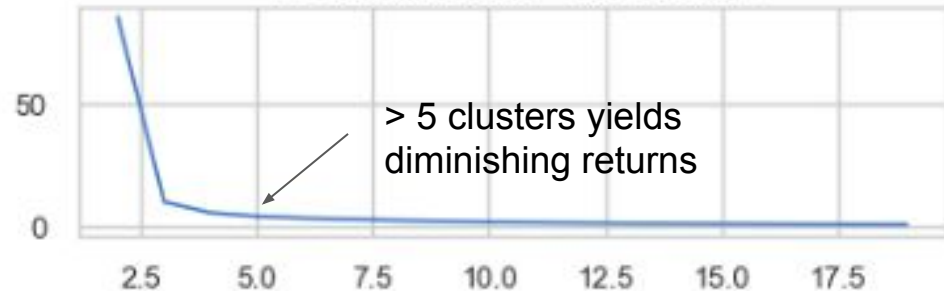
Expected behavior (when stock sentiments are low, sentiment of bitcoin and gold should be high. Some areas, have good prediction (marked). Need better rules and data for SA.

Results - KMeans Clustering

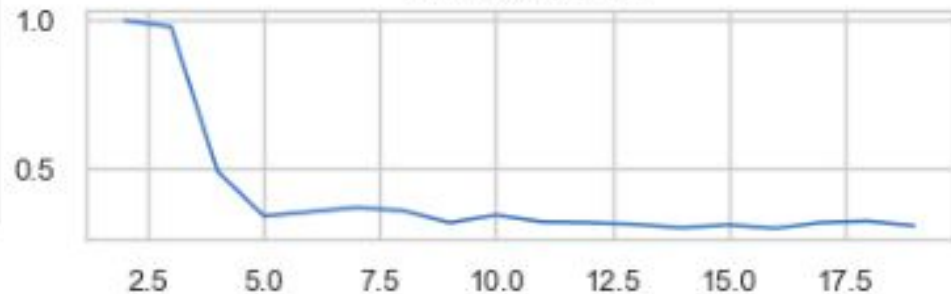
Clusters Separation from K-Means



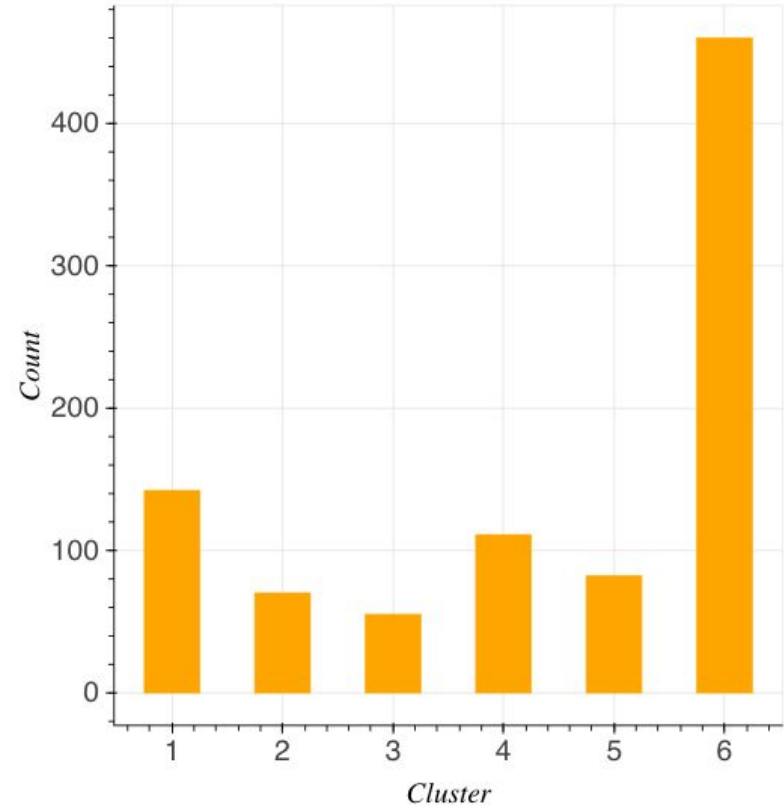
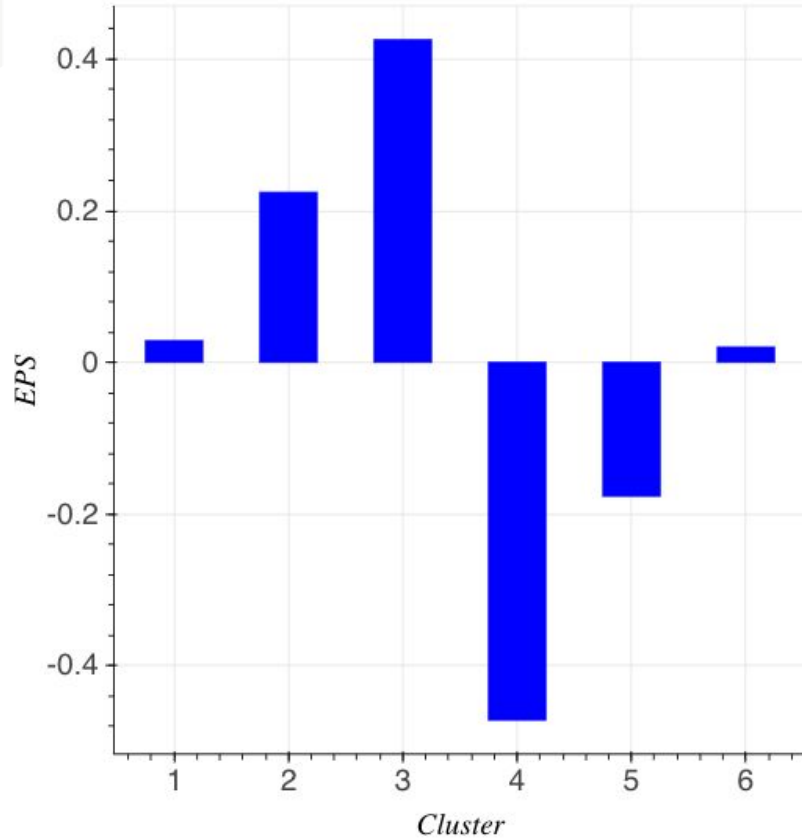
SSE within cluster - Elbow curve



Silhouette Score



Results - KMeans Clustering (1 yr Period)



x 10 Earnings

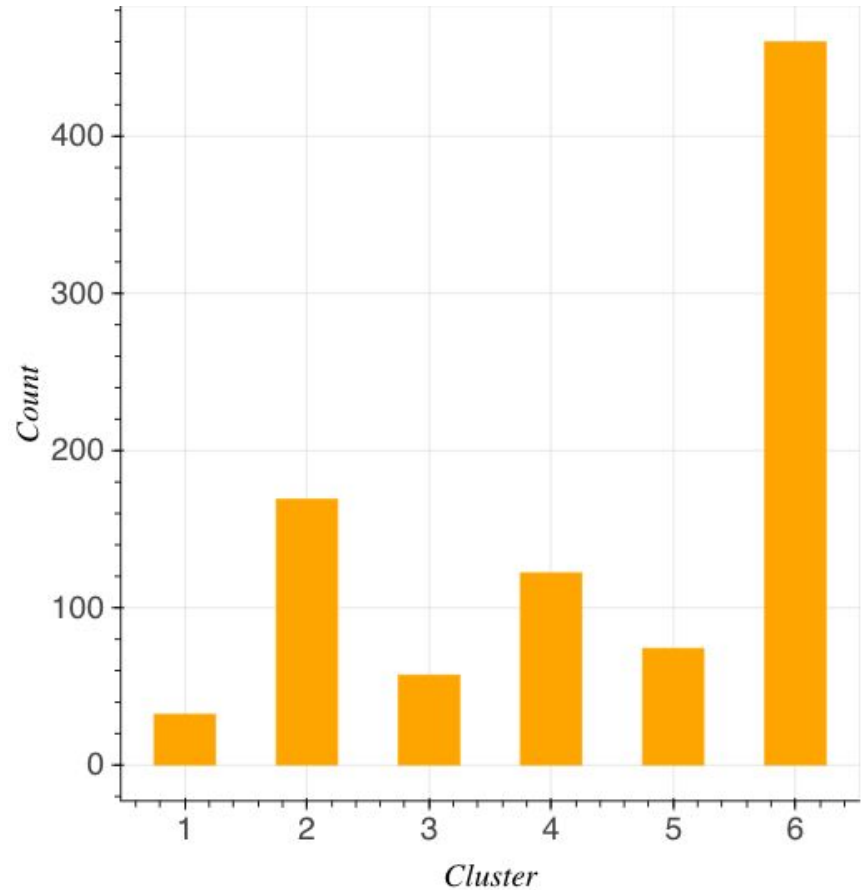
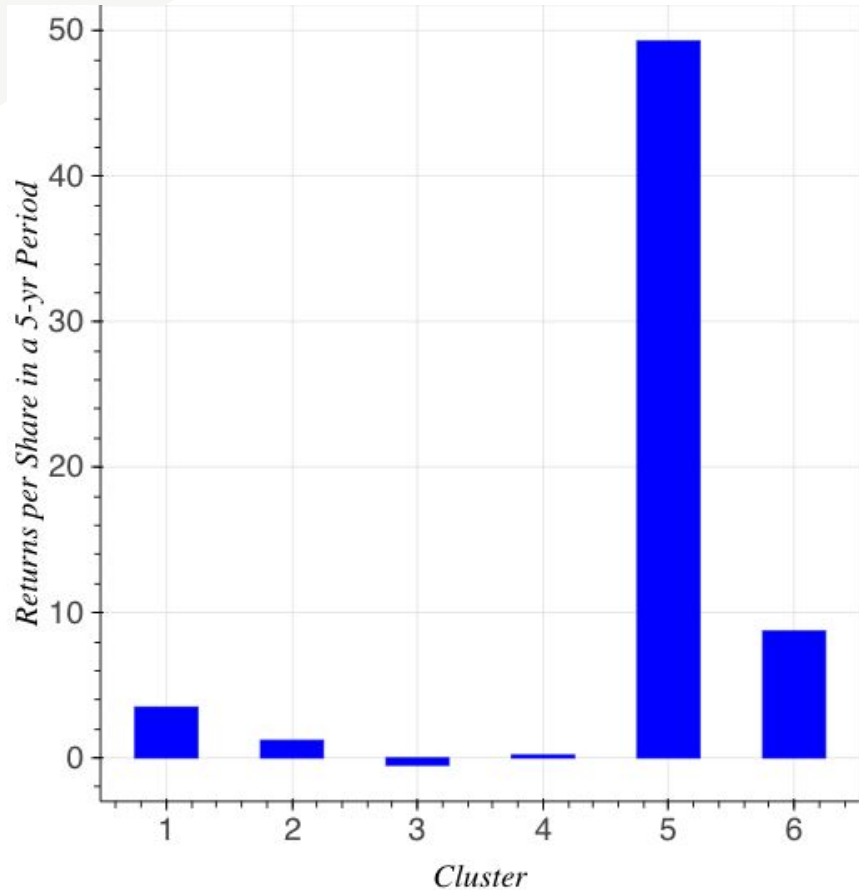
Our Best Cluster Earned 10 times more than S&P500 Last Year.

Results - The holy grail cluster (1 yr period)

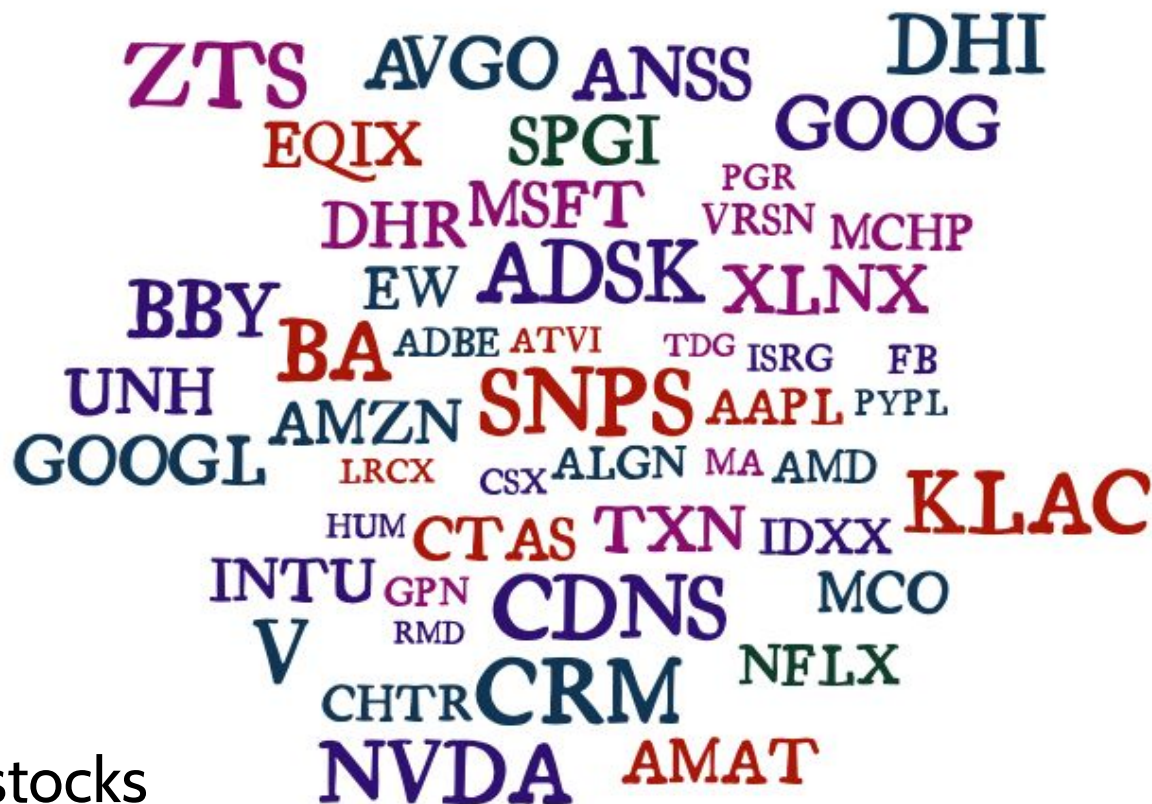


Cluster - 35 stocks

Results - KMeans Clustering (5 yr Period)



Results - The holy grail cluster (5 yr period)



Cluster - 50 stocks



Conclusions

1. Good Clustering Performance from KMeans.
2. At 5 years, the rate of return is slightly diminished when compared to S&P 500 from 10x to about 5x. Number of stocks in the top cluster is increased from 35 to 50.
3. Annual earnings → 10% more than S&P 500.
4. Sentiment analysis (SA) - lots of room for improvement.

Future Work

1. Advanced financial rules for Sentiment Analysis.
2. Data for SA - Include 8-k and 10-q data from SEC. Include better data for Gold and Crypto sentiments.
3. Use fake news detection to eliminate bot generated / trolls sentiment.
4. Try HMM, RNN for better buy/sell predictability future price of stocks.

Thank You !

Questions ?