Crypto Analysis: What a Cluster

Aquiba Benarroch
Dariush Ruch-Kamgar
Tim Moriarty
David Dannenberg

Objectives

1- Utilize unsupervised learning algos to identify relationships between different cryptocurrencies based on the daily movement of their market capitalization

2- Identify relationships between news headlines and changes in cryptocurrency market capitalization utilizing unsupervised learning and sentiment analysis.

(Note, we focus on daily changes in market capitalization and not price to normalize for the varying rates of supply emissions between cryptos)

Data Sources

Chose CoinGecko API due to the breadth of tokens available.

- Data availability forced us to limit data lookback to one year
- Rejected CoinMetrics API due to limited data breadth.

Manual collection and sensitivity assessment of relevant news articles

- Subject matters: China, institutional adoption, and regulation.

Objective #1

Identify relationships between tokens based on valuation

Data Collection and Preprocessing

Retrieved “ids” for the Top 500 cryptocurrencies by market capitalization.

- API limit of 250 results per page.

Retrieved the market capitalization data for selected tokens

- Limit of 50 calls per minute - created a workaround to avoid the rate limit
- Pulled daily data for all 500
- When cleansing the data for NaN values this fell to 220
- Many tokens were obscure/niche/lacked trading volume
- Ultimately selected 52 tokens across various categories: L1/L2 Chains, DEXs, Play-to-earn, Meme, and middleware

Calculated daily pct_change as final DataFrame for clustering analysis

At a Glance

52 cryptocurrencies

64 news articles

1-year lookback period

CoinGecko API

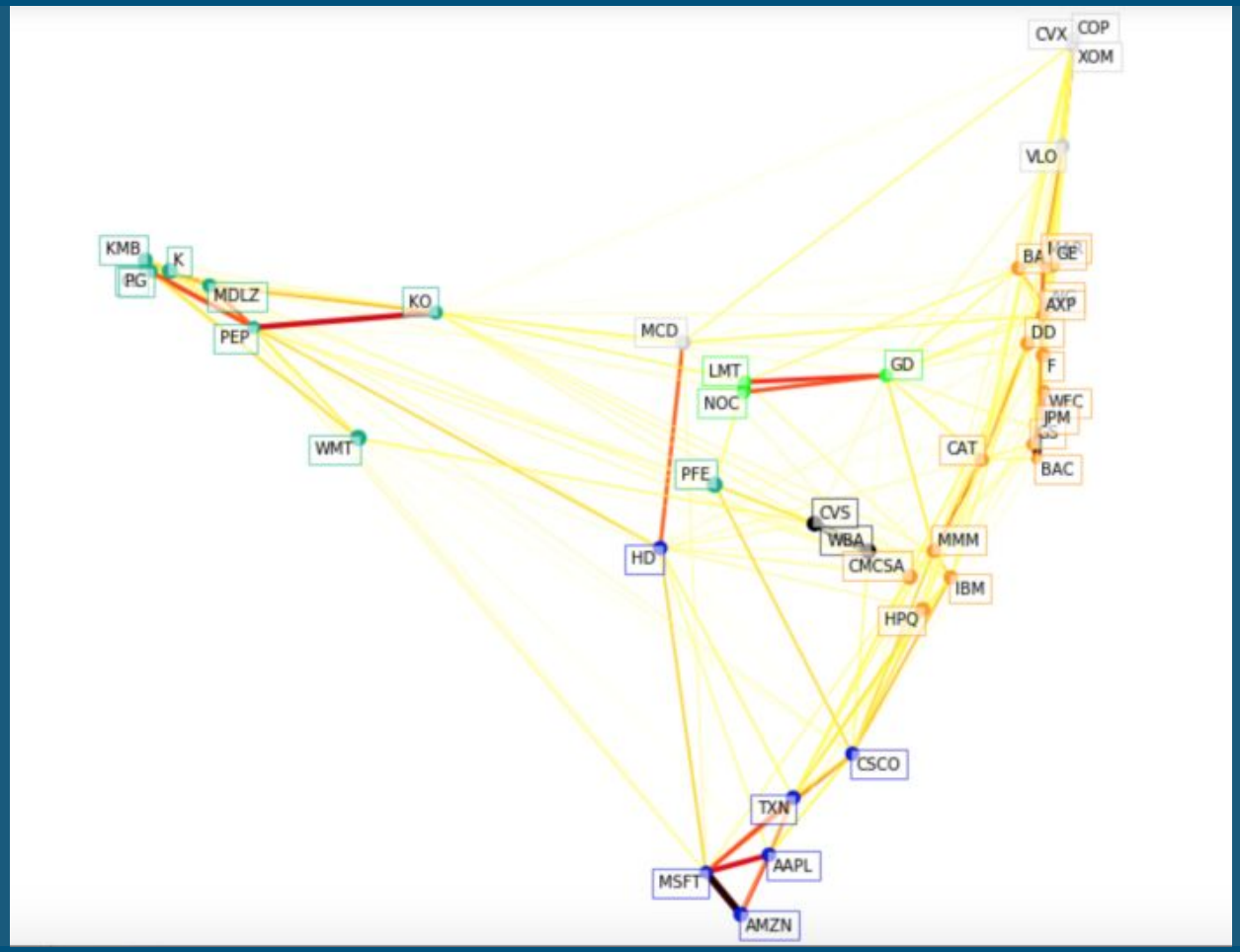
Market Capitalization analysis

- Affinity Propagation (clustering)
- Graphical Lasso (partial correlation)
- Locally Linear Embeddings (visualization)

News analysis

- KMeans

Clustering Analysis Inspiration



- MSFT, AAPL, AMZN, TXN all have same color nodes, bold lines connecting, and positioned closely together
- Indicates a close relationship based on variations in historical stock prices
- Potential pairs trading opportunity
- Diversified portfolio creation

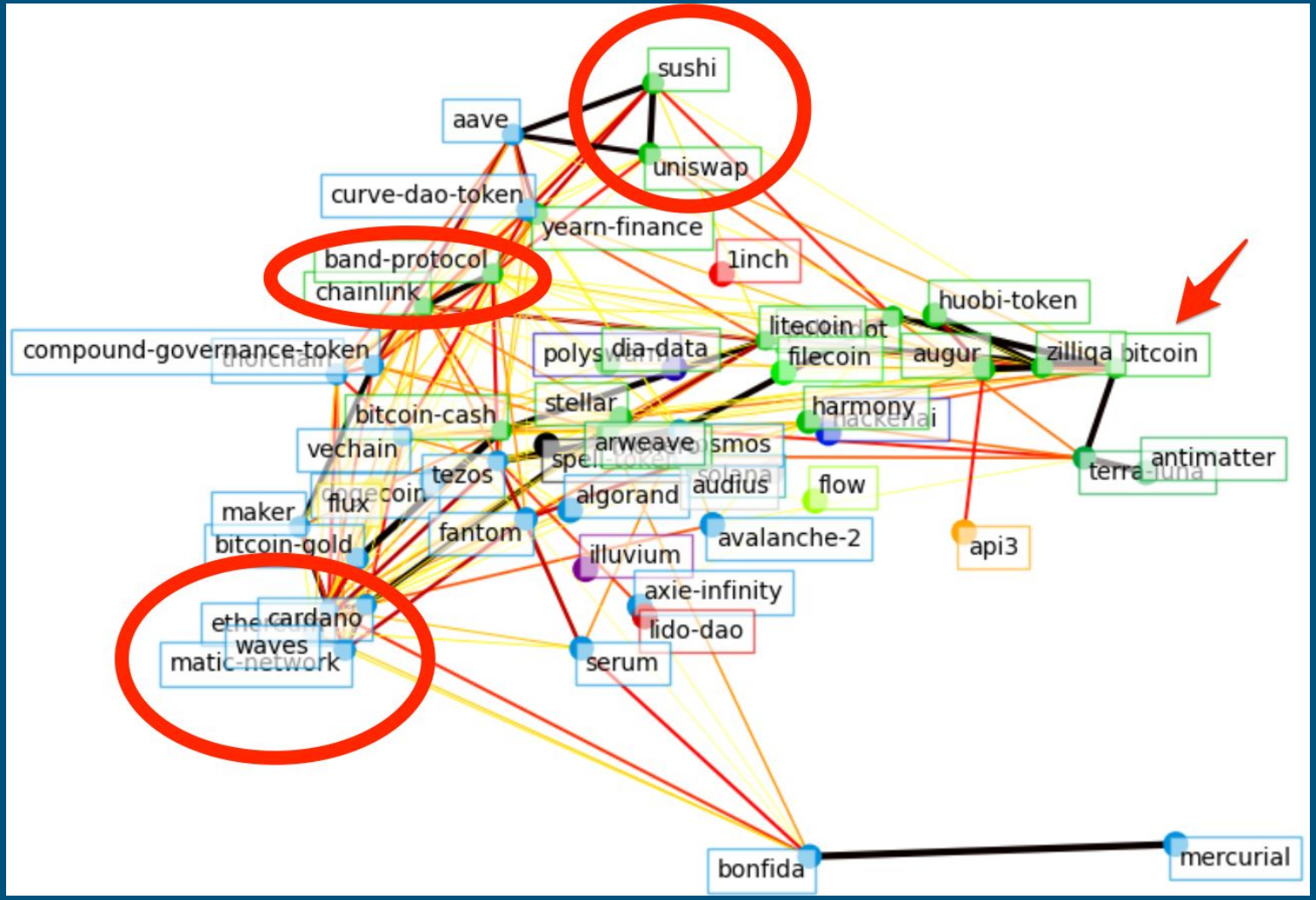
Three-Model Clustering Approach

Graphical Lasso (lines between the nodes) is a sparse inverse covariance estimate to determine the partial correlation between variables and control for the impact of confounding variables in typical correlation calculations

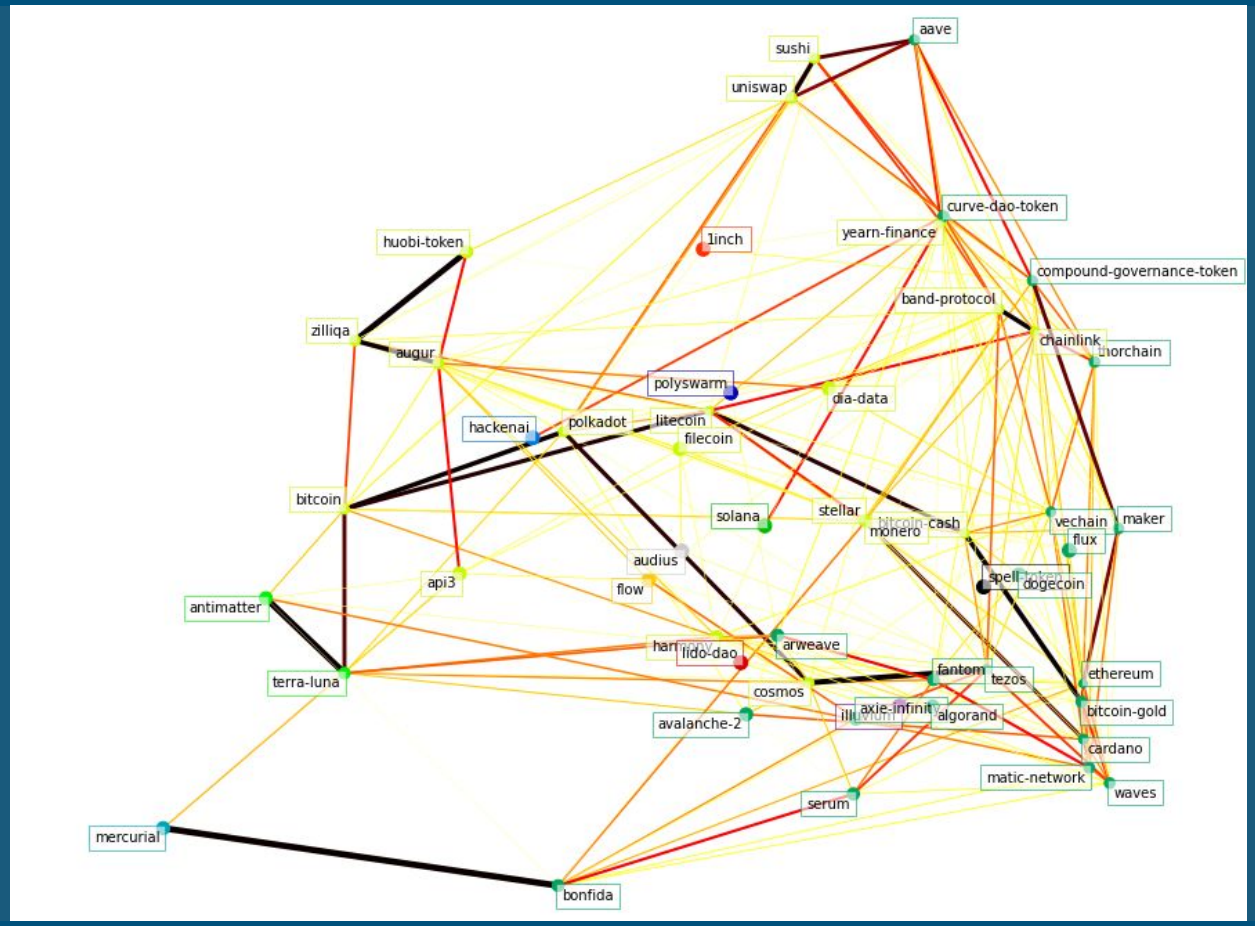
Affinity Propagation (color of the nodes) is a clustering analysis similar to KMeans with the advantage of self-optimizing the number of clusters

Manifold learning - Locally Linear Embedding (closeness of the nodes) is similar to a nonlinear PCA analysis, allowing 2D visualization

Model Output: Tokens clustered across all three models have strongest relationship



Model Output – Price Action Clustering

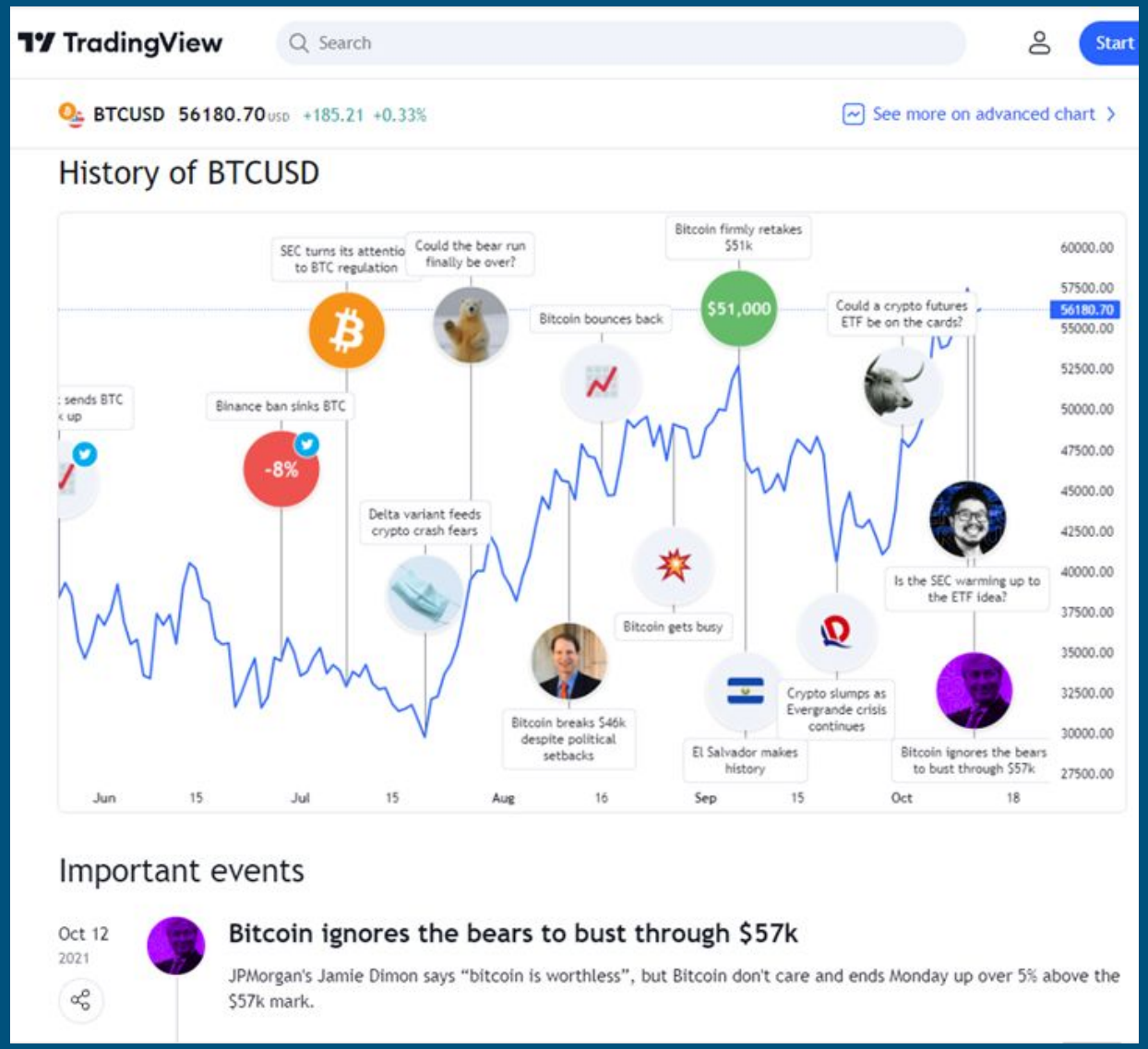


- 3 model cluster:
  - Strength of lines - Lasso Cross-Validation (sparse inverse covariance estimation)
  - Color of nodes - Affinity Propagation (clustering that doesn't enforce equal size clusters andcan automatically determine the number of clusters - analogous to k-means)
  - Positioning of nodes - Manifold learning
- Insights:
  - Tokens clustered across all three models have strongest relationship: yearn and curve, sushi and uniswap, Chainlink and Band etc
  - Smart contract platforms were well clustered and distinct from BTC

Objective #2

Identify relationships between tokens using crypto news

Data Collection - Crypto News Sources



Data Collection - Crypto News Assessment

2021 CRYPTO NEWS				
DATE	**NEWS**	**NEWS CATEGORY**	**COIN**	**SENTIMENT**
1/5/2021	JP Morgan: BTC Could Hit $146K	Banking	BTC	1
1/11/2021	UK FCA Warns: Prepare to Lose All Your Money	Government	CRYPTO	-1
1/14/2021	Bitcoin Core Major Upgrade (0.21.0)	Infrastructure	BTC	1
1/25/2021	UK FCA Warns: Bitcoin's Days Are Numbered	Government	BTC	-1
1/28/2021	Musk Positive BTC Tweet	Musk Tweet	BTC	1
1/28/2021	Musk Positive DOGE Tweet	Musk Tweet	DOGE	1
1/28/2021	Dorsey Positive BTC Tweet	Dorsey Tweet	BTC	1
2/5/2021	Elon Musk Tweet: Dogecoin to the Moooonn	Musk Tweet	DOGE	1
2/8/2021	SEC Filing: Tesla Bought $1.5 Billion in BTC	Government	BTC	1
2/8/2021	Tesla Accepts BTC as Payment	Corporate	BTC	1
2/10/2021	Mastercard: Bringing Crypto onto its Network	Corporate	CRYPTO	1

Data Preprocessing and Approach

Defined a new dataframe with the dates when news signals happened.

- Data is daily change in market capitalization for the same 52 tokens.
- Included 64 dates with news events

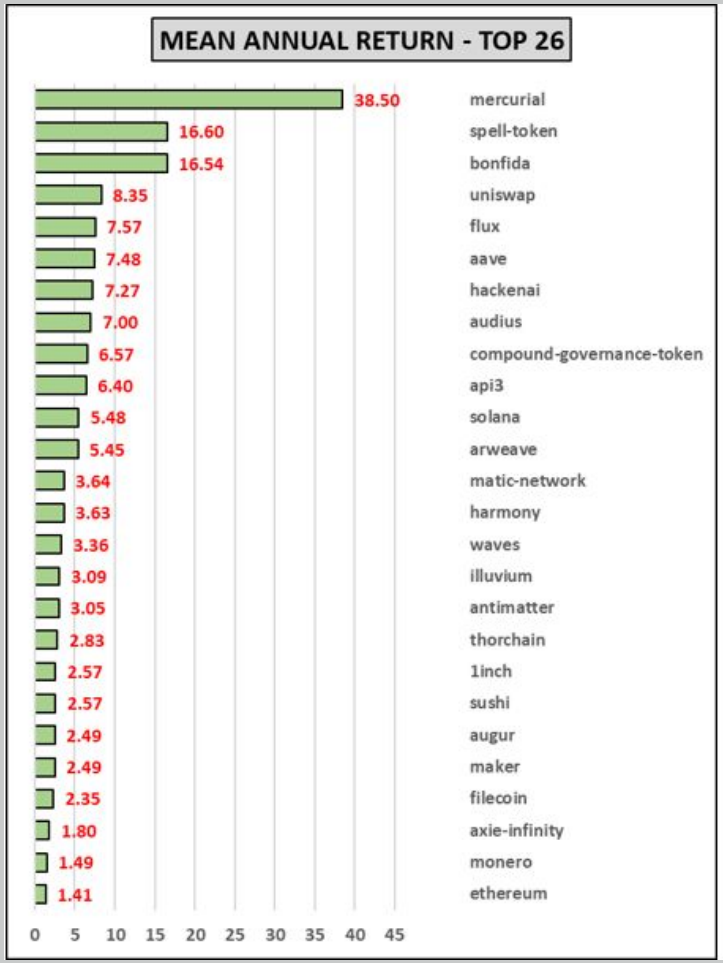
Calculated annual returns and variances for the selected dates.

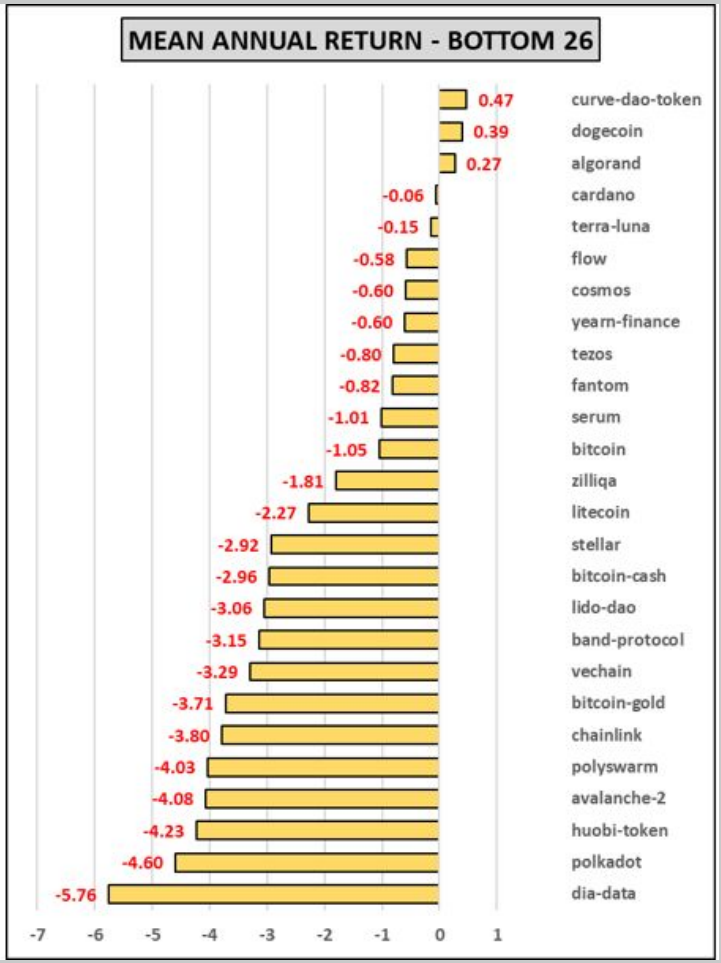
Used the KMeans algorithm for clustering.

Crypto price movements on days with a major crypto-related headline appear well-clustered

Annual Return on News Days

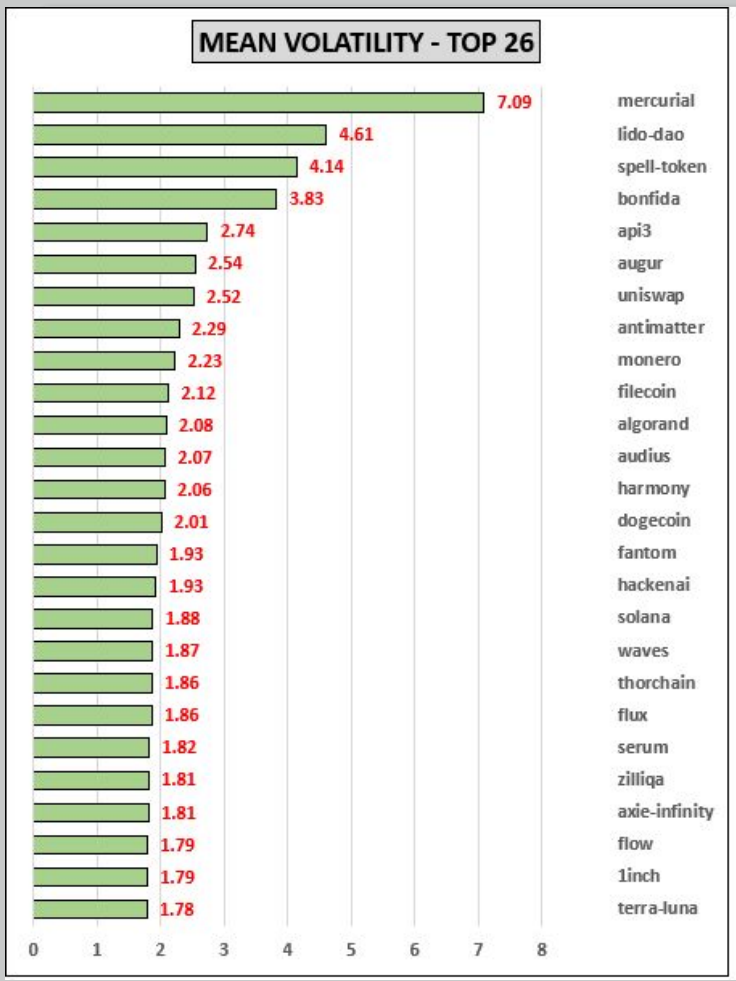
On days with news, tokens on the left outperformed tokens on the right

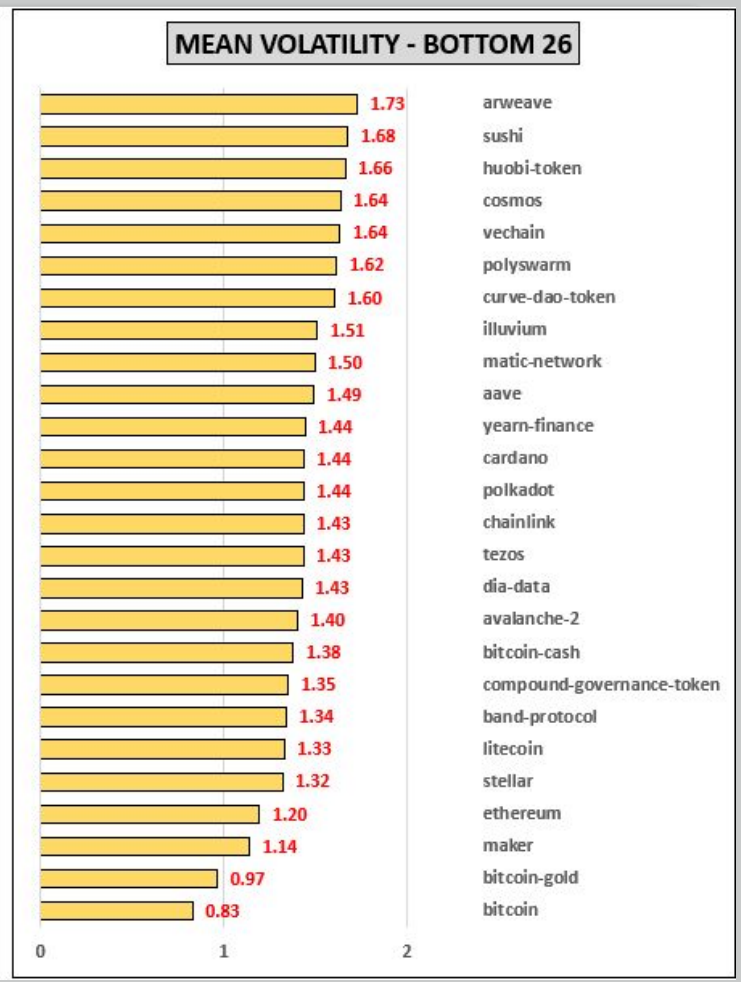




Volatility on News Days

On days with news, tokens on the left were more volatile than tokens on the right



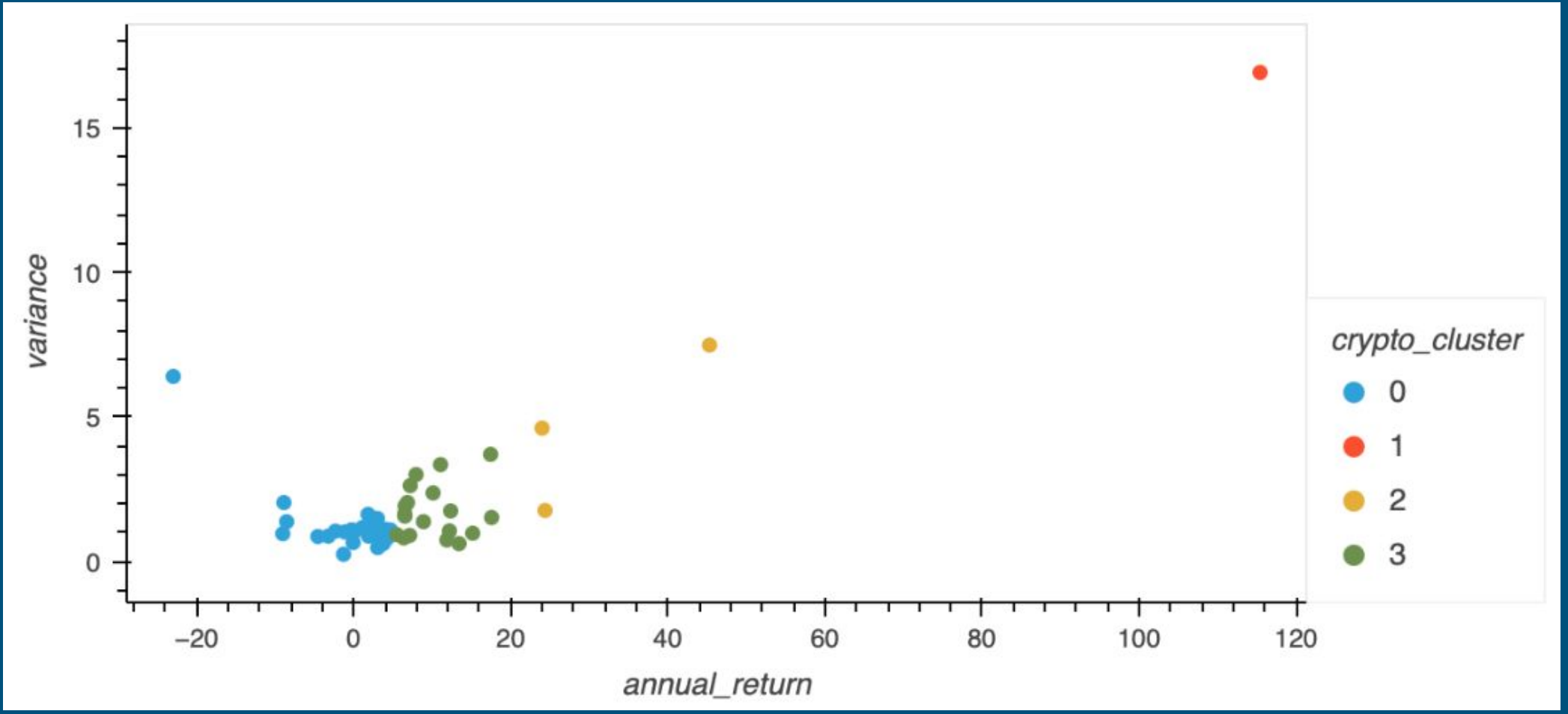


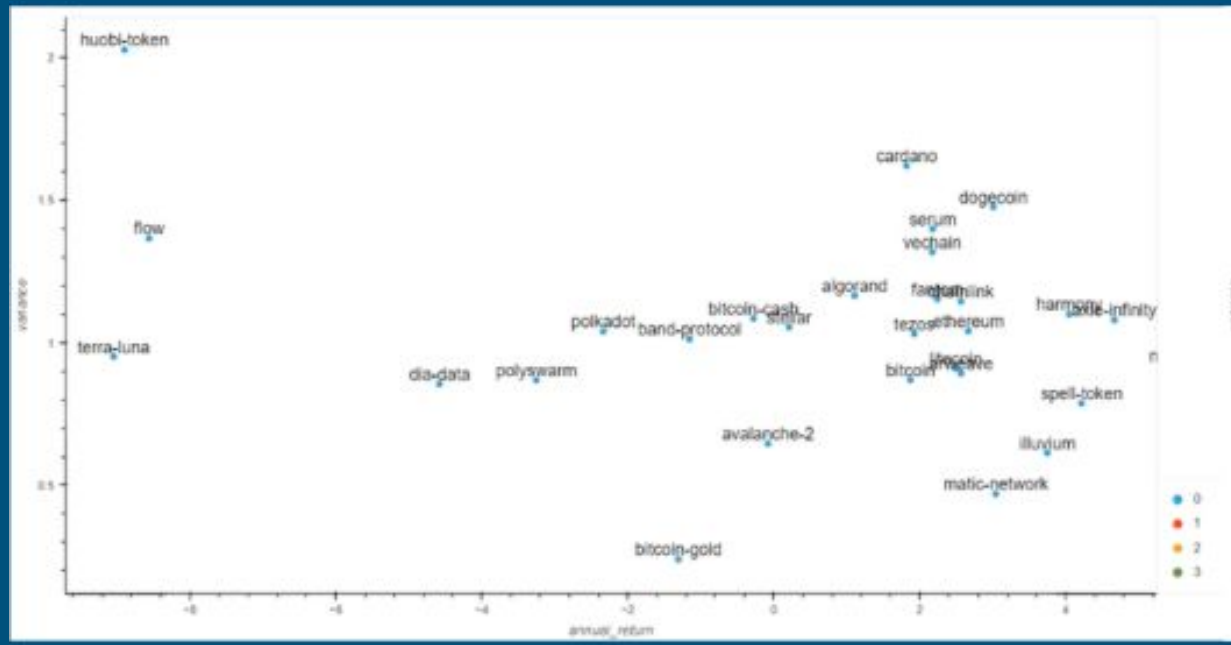
CHINA NEWS - 2021

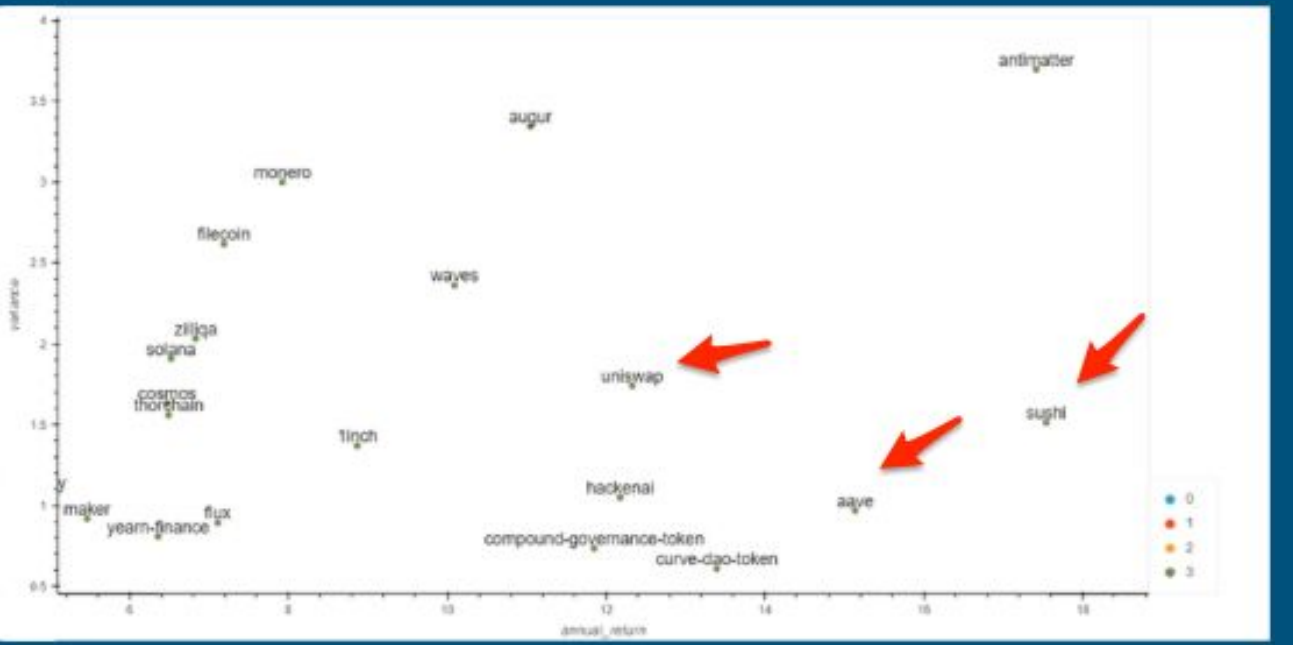
DATE	NEWS	CATEGORY	INDICATOR
5/18/2021	China Bans Institutions From Crypto Business	China	-1
5/21/2021	China Crackdown on Mining & Trading	China	-1
6/21/2021	China Intensifies Mining Crackdown	China	-1
9/20/2021	Evergrande Crisis Hits	China	-1
9/27/2021	China: All Crypto Business is Illegal	China	-1

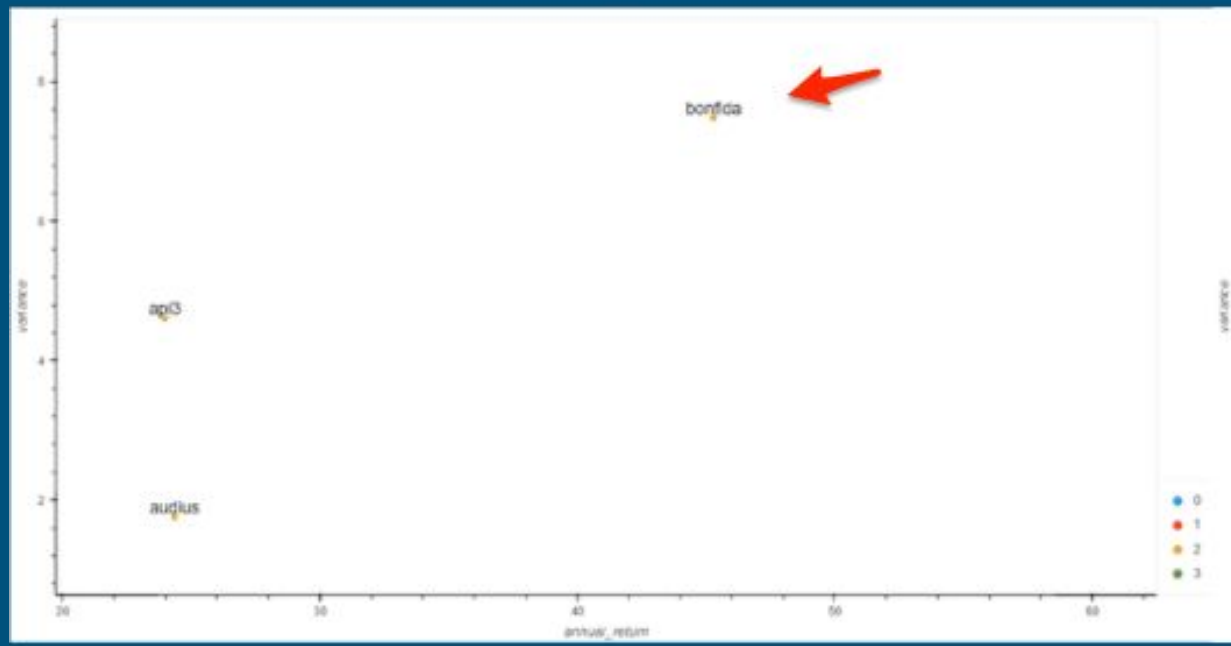
China News

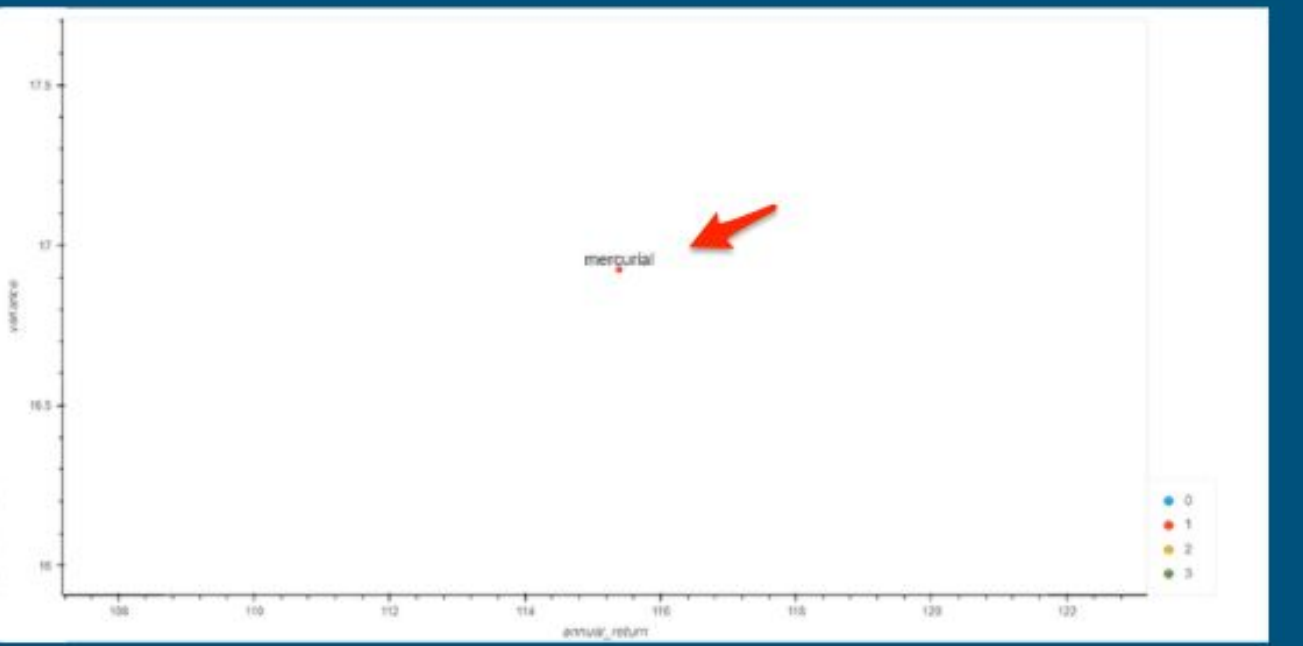
Decentralized applications (DEX) meaningfully outperformed other tokens




huobi-token
flow
terra-luna
cardano
dogecoin
serum
vechain
algorand
polkadot
band-protocol
bitcoin-cash
tezos
ethereum
harmony
dia-data
polyswarm
bitcoin
spell-token
avalanche-2
illuvium
matic-network
bitcoin-gold
variance
annual_return
0
1
2
3

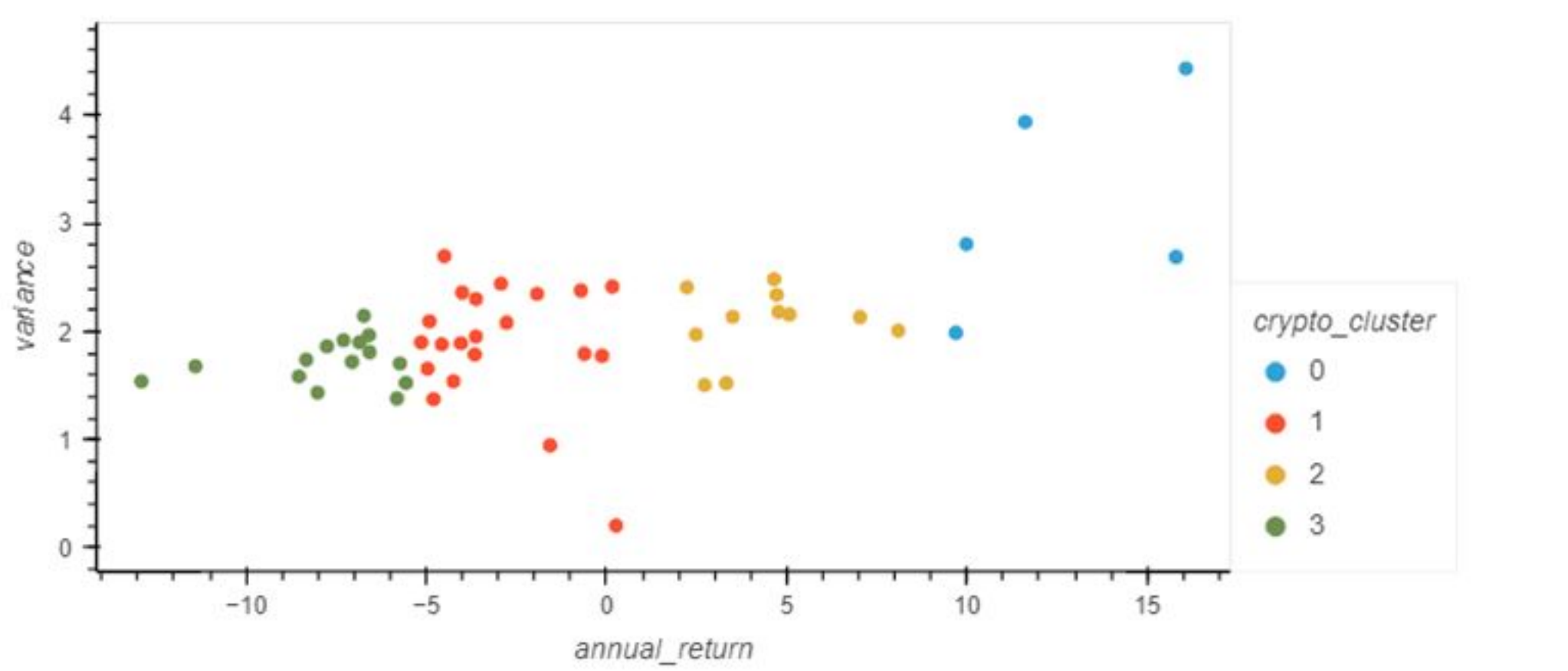

antimatter
augur
mogero
filecoin
waves
ziliqa
solana
uniswap
cosmos
thorchain
sushi
1inch
hedgenai
aave
maker
yearn-finance
compound-governance-token
curve-dao-token
variance
annual_return
0
1
2
3

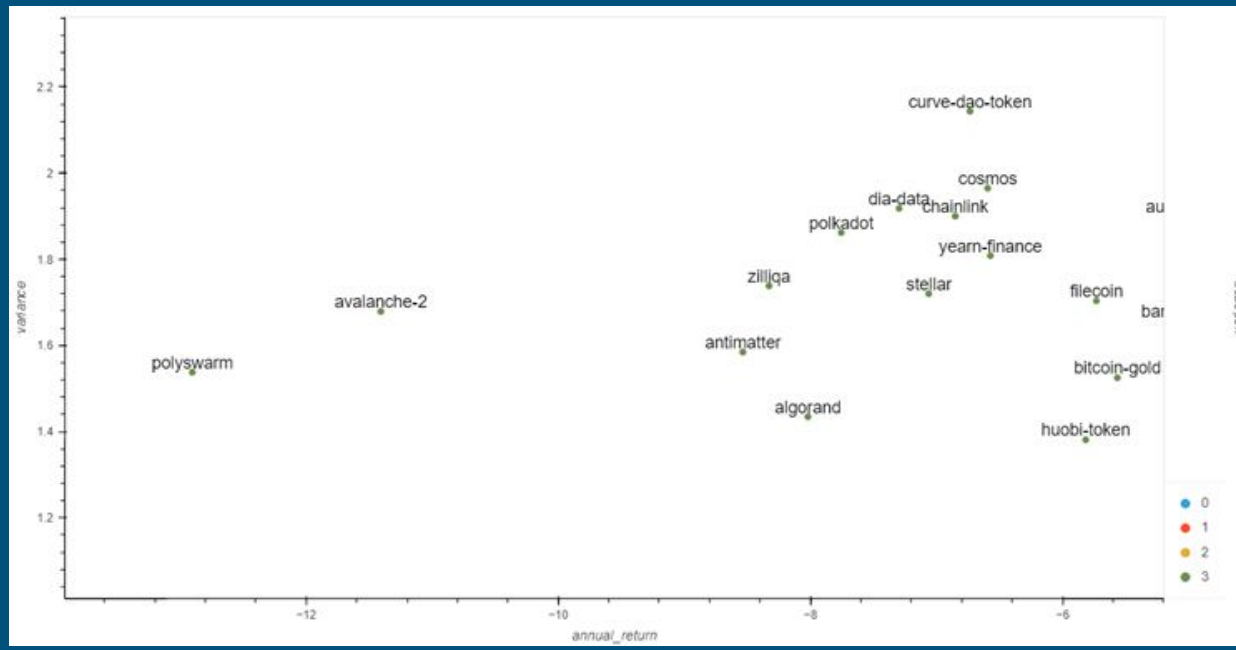

bonfida
api3
audius
variance
annual_return
0
1
2
3

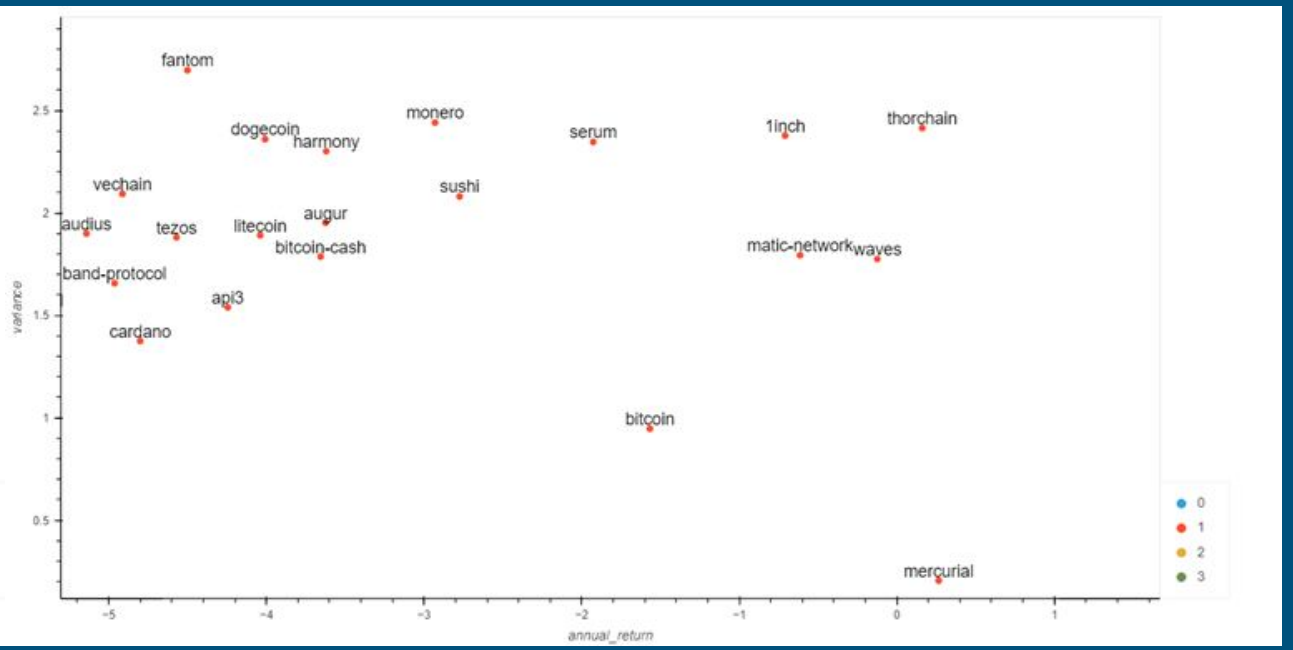

mercurial
variance
annual_return
0
1
2
3

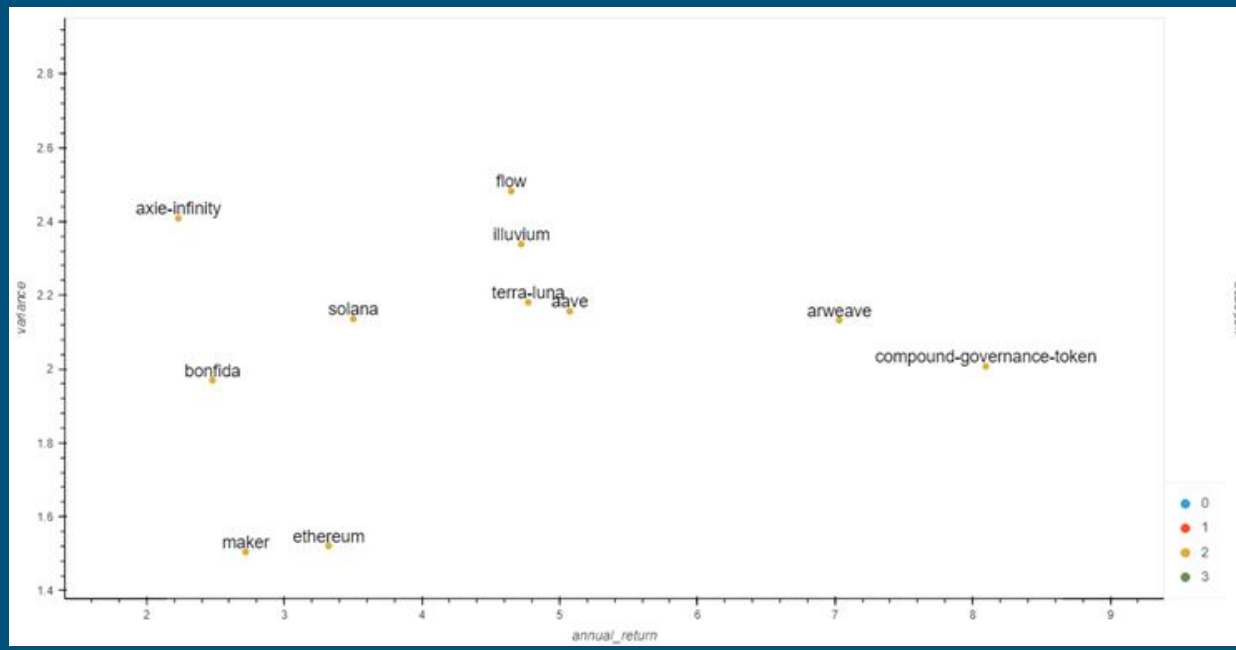
Regulatory News (EU and USA)

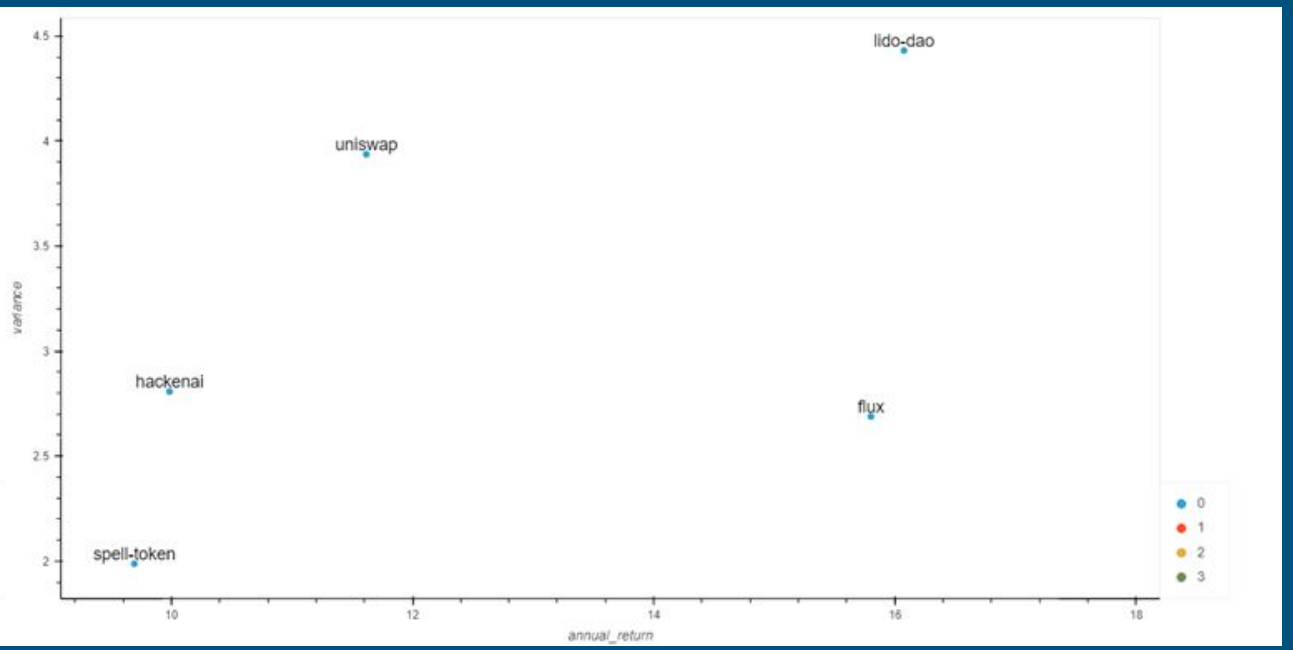
Greater dispersion of returns and volatility may indicate surprise news




curve-dao-token
cosmos
dia-data
chainlink
polkadot
yearn-finance
zilliqa
stellar
filecoin
avalanche-2
antimatter
polyswarm
bitcoin-gold
algorand
huobi-token
variance
annual_return
0
1
2
3

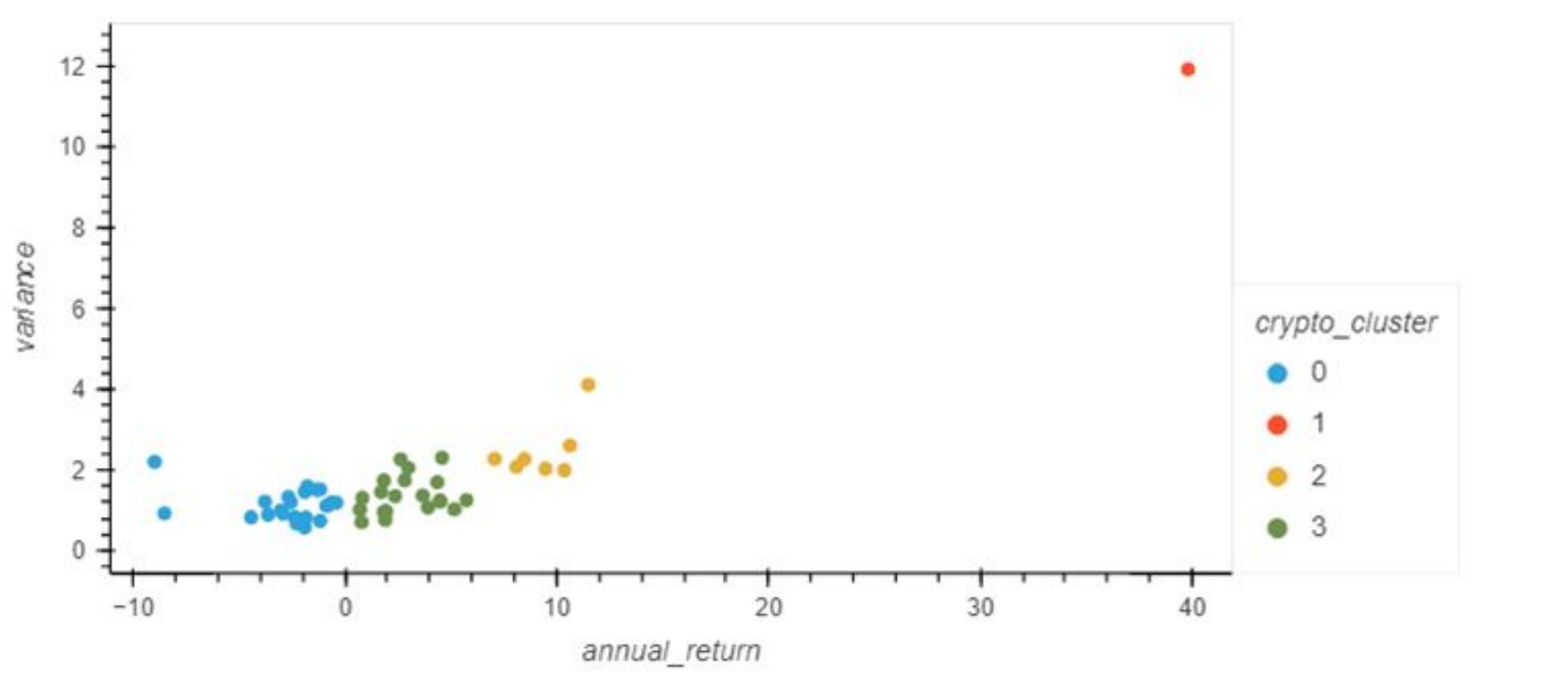

fantom
monero
dogecoin
harmony
serum
1inch
thorchain
vechain
sushi
audius
tezos
litecoin
augur
bitcoin-cash
matic-network
waves
band-protocol
api3
cardano
bitcoin
mercurial
variance
annual_return
0
1
2
3

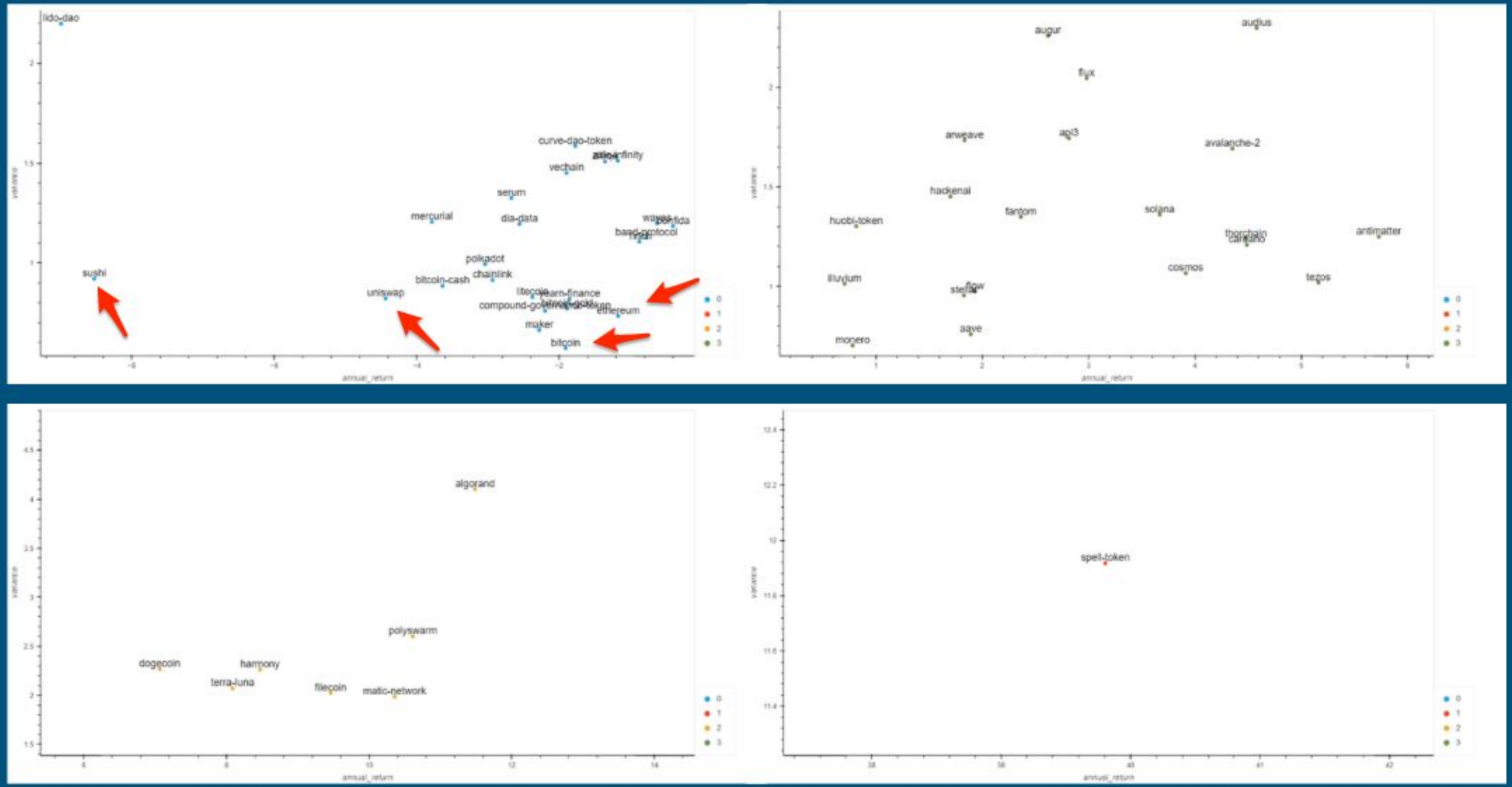

flow
axie-infinity
illuvium
terra-luna
aave
solana
arweave
compound-governance-token
bonfida
maker
ethereum
variance
annual_return
0
1
2
3


lido-dao
uniswap
hackenai
flux
spell-token
variance
annual_return
0
1
2
3

Positive Institutional News

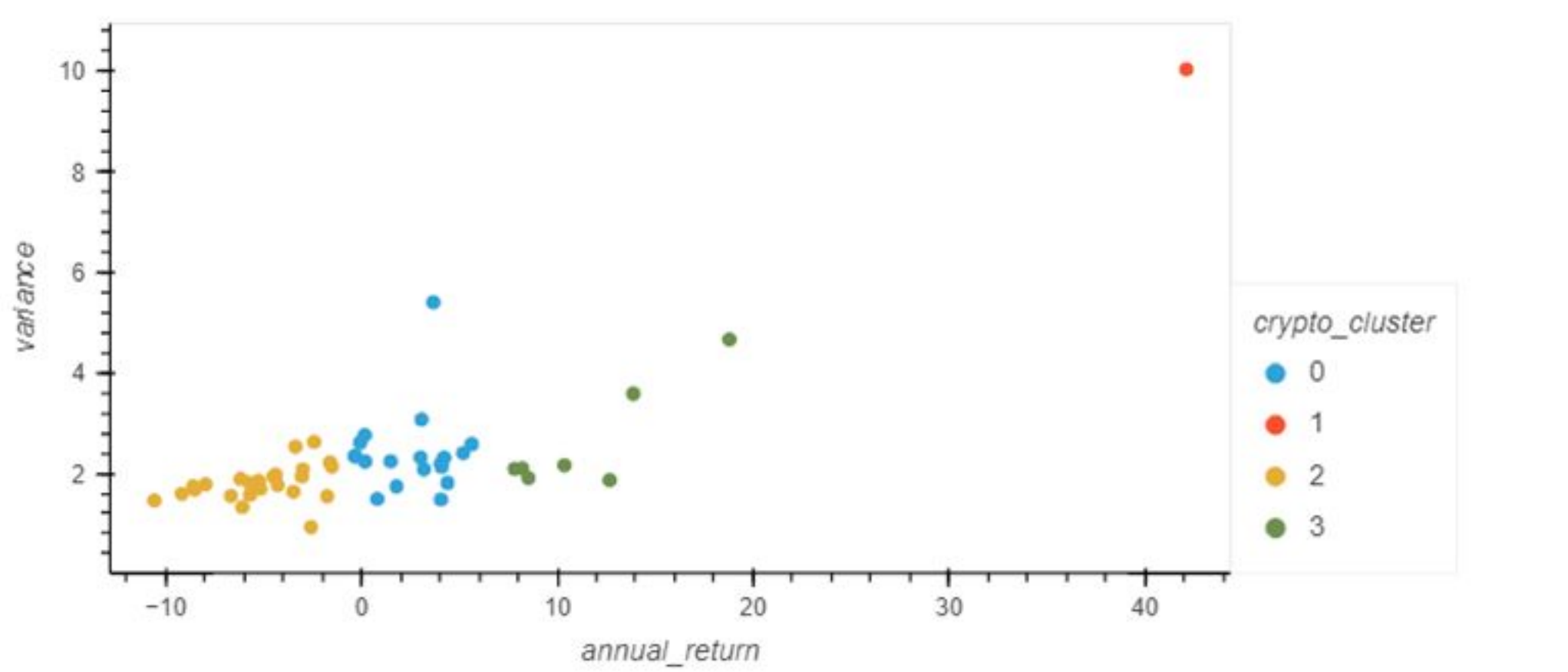
Large cap tokens (ETH, BTC, UNI, SUSHI) reacted negatively to the news.

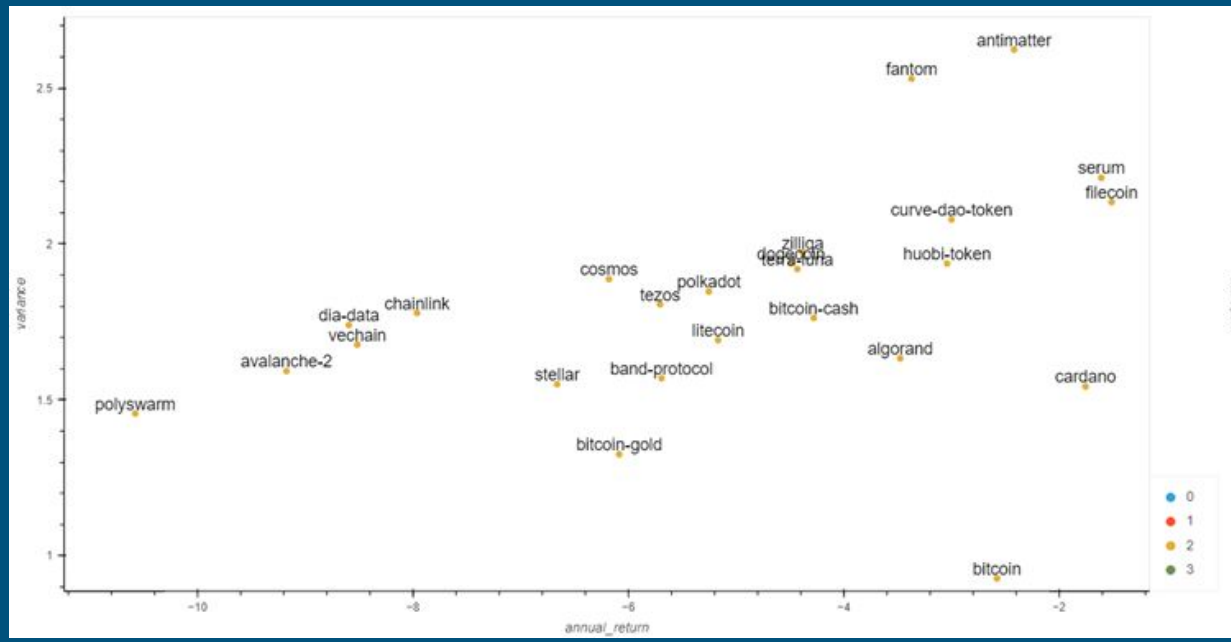


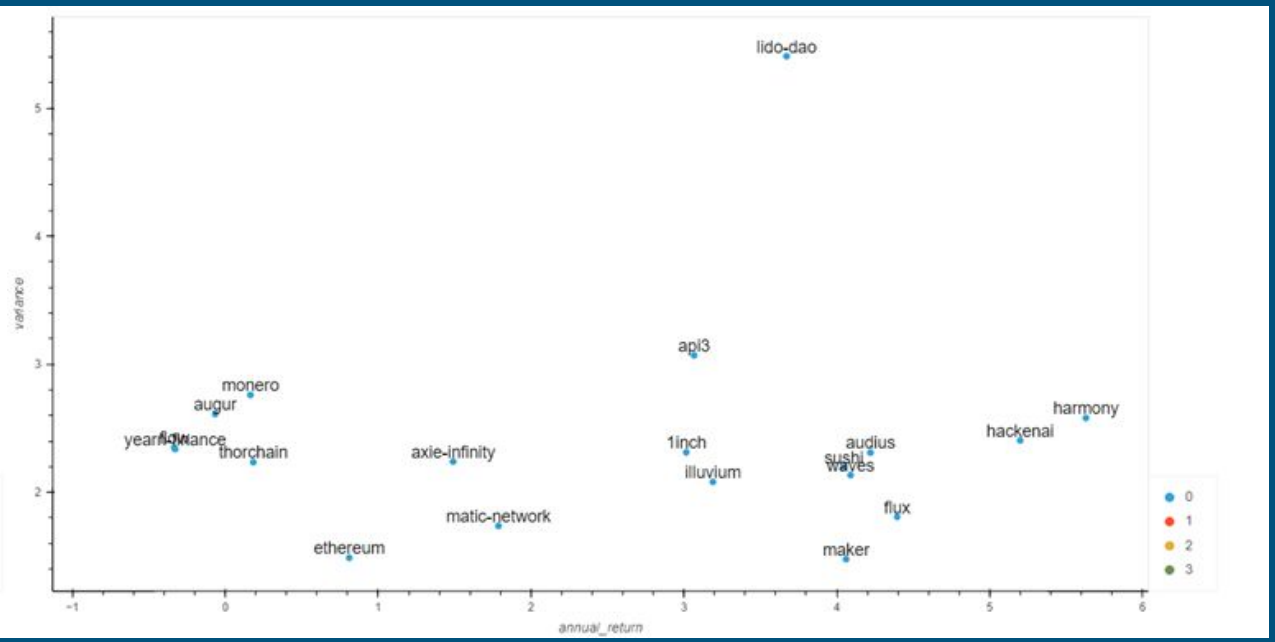

lido-dao
curve-dao-token
vechain
serum
mercurial
dia-data
band-protocol
polkadot
chainlink
bitcoin-cash
sushi
uniswap
yearn-finance
ethereum
maker
bitcoin
augur
audius
flux
arweave
api3
avalanche-2
hackenai
fantom
solana
huobi-token
thorchain
antimatter
cosmos
tezos
illuvium
flow
aave
monero
algorand
polyswarm
dogecoin
harmony
terra-luna
filecoin
matic-network
spell-token
variance
annual_return

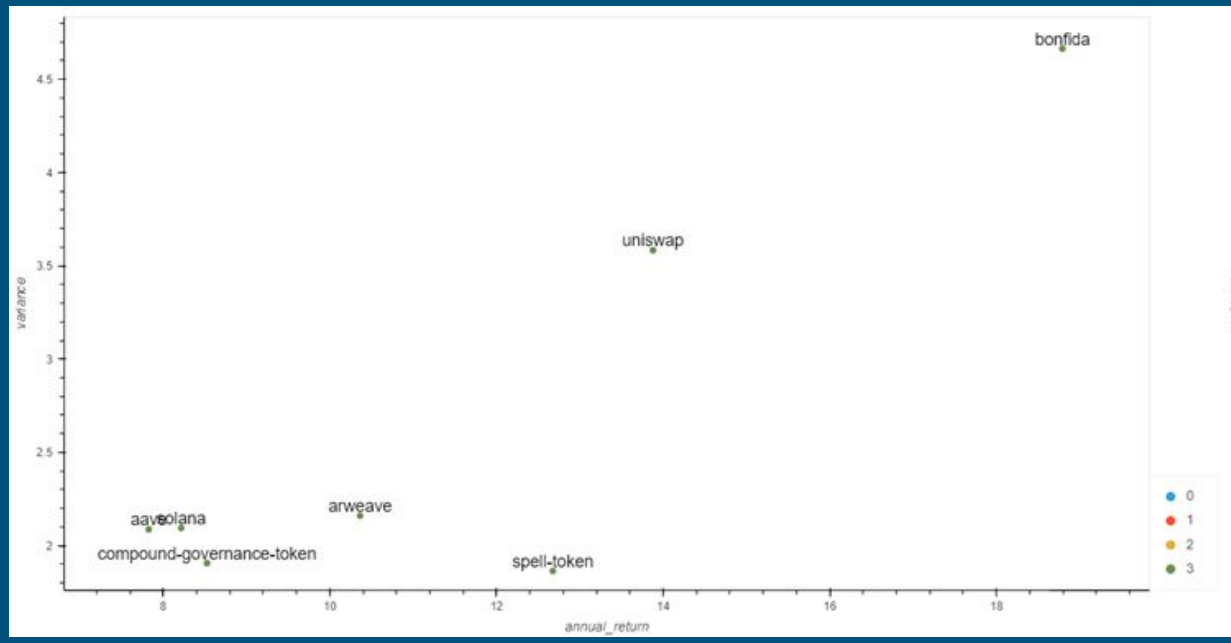
Negative Institutional News

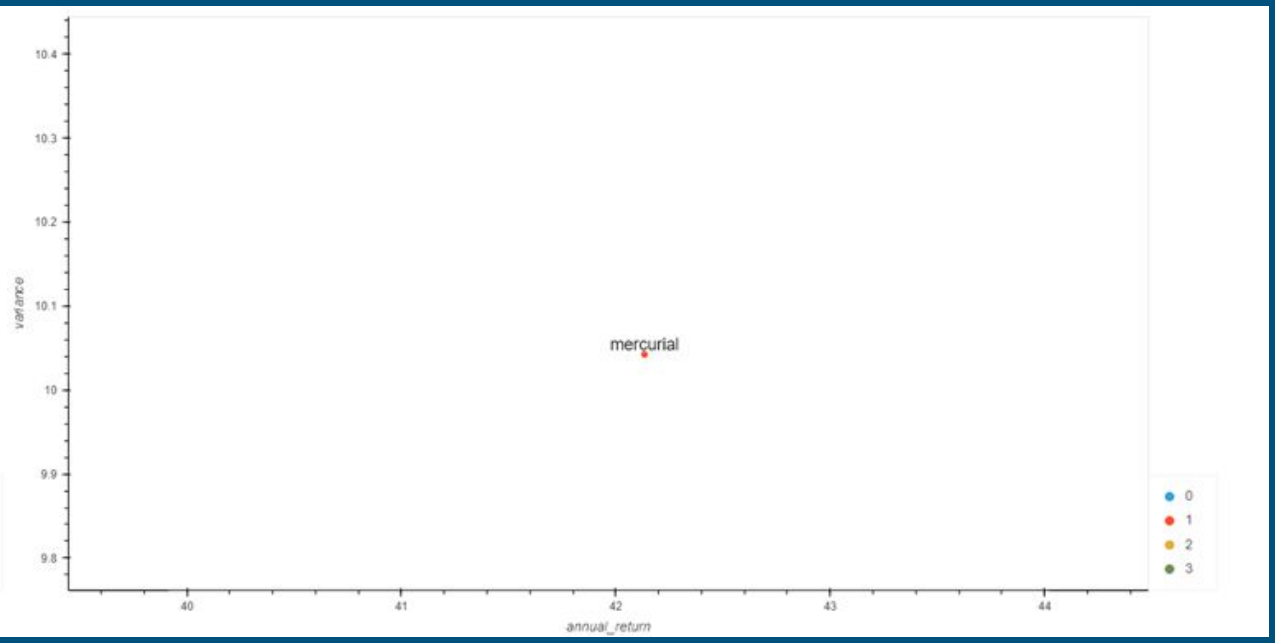
Greater dispersion of returns and volatility compared to positive news




antimatter
fantom
serum
filecoin
curve-dao-token
huobi-token
cosmos
polkadot
tezos
chainlink
dia-data
vechain
avalanche-2
litecoin
bitcoin-cash
algorand
stellar
band-protocol
cardano
polyswarm
bitcoin-gold
bitcoin
variance
annual_return
0
1
2
3


lido-dao
api3
monero
augur
harmony
hackenai
1inch
audius
sushi
waves
illuvium
axie-infinity
thorchain
flux
matic-network
ethereum
maker
variance
annual_return
0
1
2
3


bonfida
uniswap
arweave
aave
solana
compound-governance-token
spell-token
variance
annual_return
0
1
2
3


mercurial
variance
annual_return
0
1
2
3

Key Insights

- Crypto price movements on days with a major crypto-related headline were well-clustered
- On days where China bans made headlines, decentralized exchanges meaningfully outperformed
- On days when European and American regulatory headlines were negative, decentralized exchanges and yield management platforms slightly outperformed
- Among cryptocurrencies, we found smart-contract names to be well clustered and payments names to be well-clustered according to price action

Opportunities for Future Work

1) Cluster model - could add to data set or adjust model methodology
   a) Different methods could be used to confirm model output (e.g. different manifold learning instead of locally linear embedding, etc)
   b) Could combine equities to determine relationship between equity and crypto price action
   c) Could formally categorize cryptocurrencies and incorporate into model to determine relationships across different types (e.g. DEX vs. yield)

2) News analysis - could code the news with additional metrics or coded additional stories and observe model changes
   a) Compare price behavior on news events to dates without news events
   b) Compare impact of news events over time