Cryptocurrencies

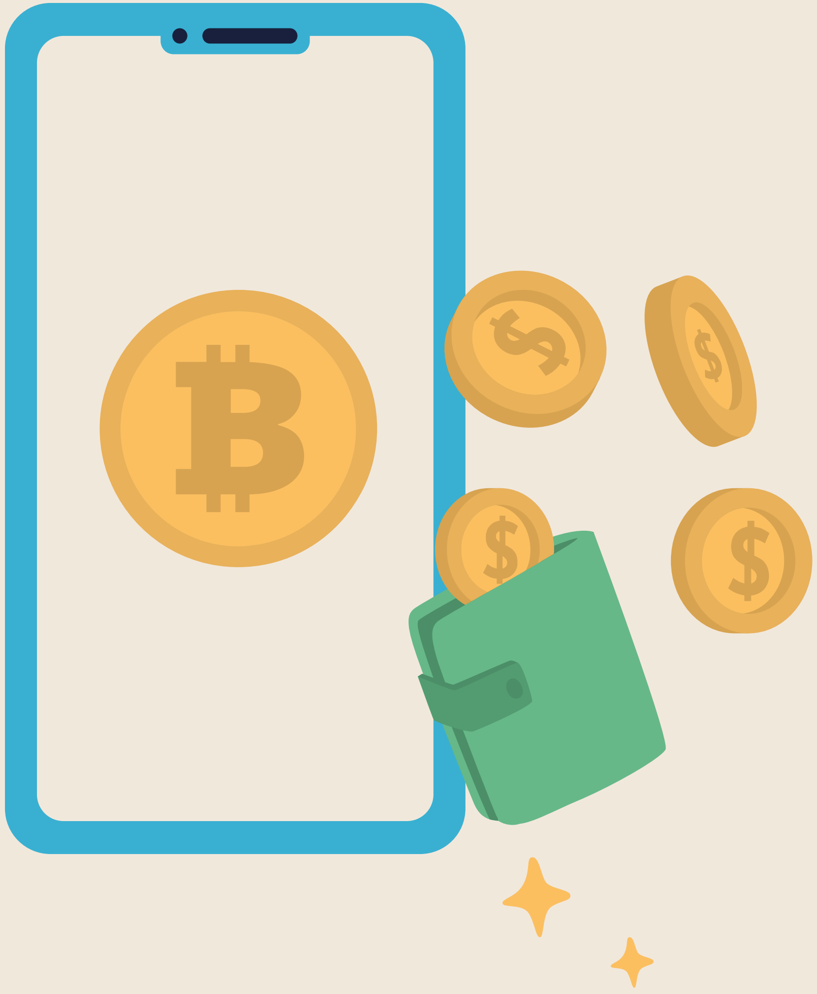
A Socialogical Approach

Jordan Ong Zhi Rong
Kenneth Low Yan Wei
Peace Tay Jiunn Ching
Tan Yu Yan, Rachel
Widya Tantiya Yutika
Yap Pin Yaw

Hypothesis

“Prices of cryptocurrencies have a **strong correlation** with the **sentiments** of certain individuals, institutions or the general population”

Objectives



1. To explore if sentiments on social media have a correlation with crypto prices
2. To analyse the extent of influence selected individual/Institutions have on crypto prices based on centrality score
3. To gain insights on the topics on social media during days of large price movements

Table of Contents

01

Dataset

Data Sets

Tweets (Influencers)

Elon Musk

- Founder of Tesla, SpaceX
- Richest man in the world
- Known to influence price in the market

Vitalik Buterin

- Co-founder of Ethereum
- Strong advocate for creating decentralized money

Nayib Bukele

- President of El Salvador
- First country in the world to adopt cryptocurrency as legal tender

Changpeng Zhao

- Founder and CEO of Binance
- World largest cryptocurrency exchange by trading volume

Tweets (Institutes)

- Leading global investment bank and financial services company
- Widely followed by institutional and retail investors and market participants

- Delivers business and market news
- Also widely followed by institutional and retail investors and market participants

Tweets (General Pop)

- In 2019, around 18k bitcoin related tweets per day
- In recent years, as bitcoin gains more popularity, around 100k+ bitcoin related tweets per day

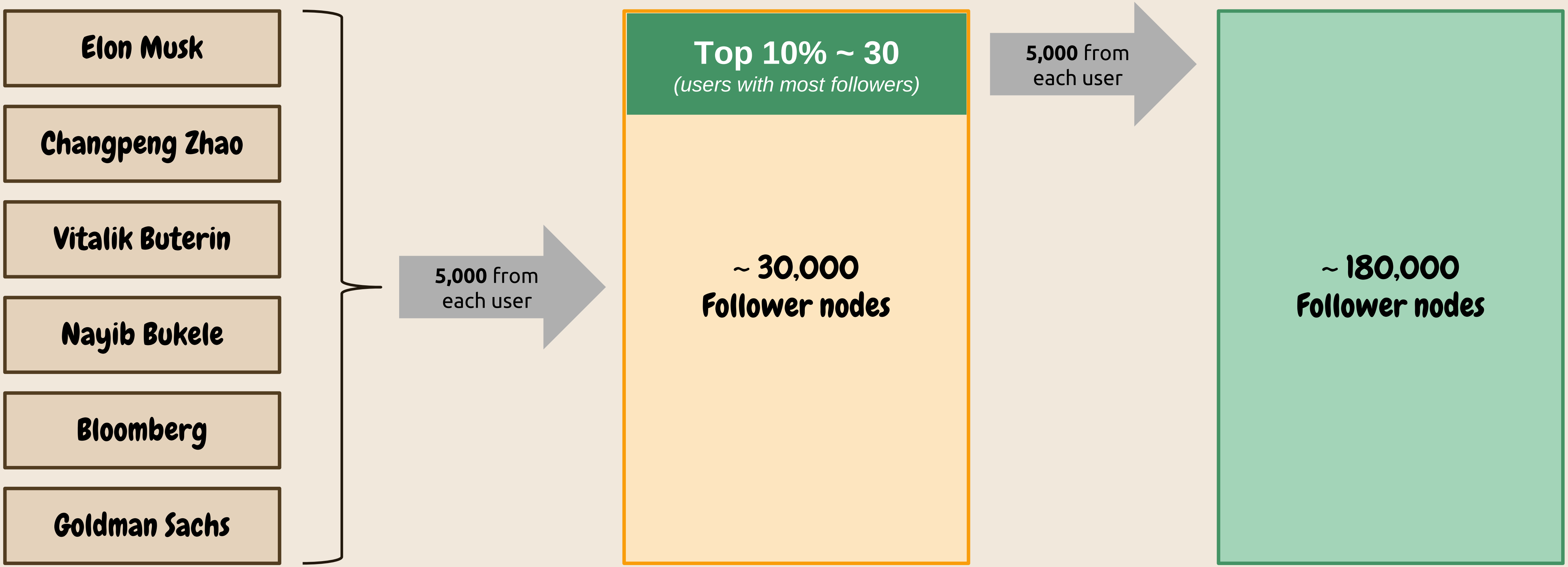
Forum Discussions

- Online social forum for discussion
- Bitcoin subreddit has 4,300,000 members with average of 3,600 active online
- Data of posts in the subreddit

02

Network Visualisation & Centrality Analysis

Total Number of Nodes ~ 180,000

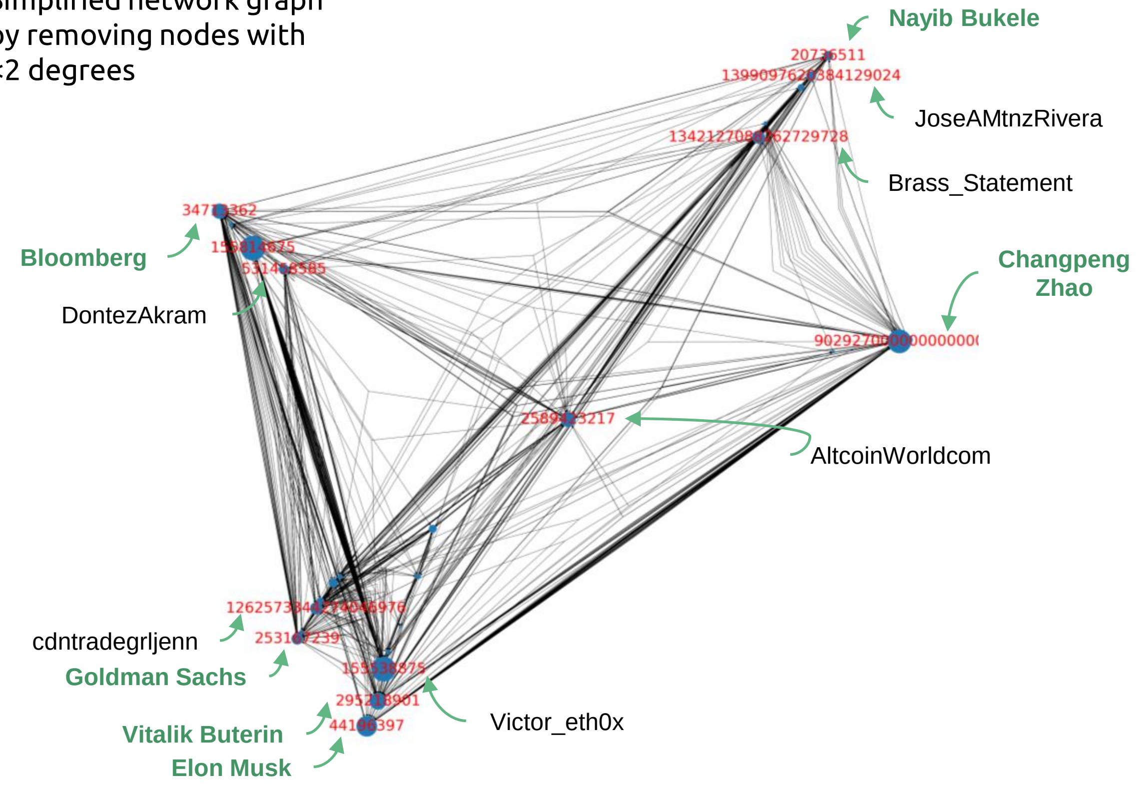


Network Visualisation

Level 0	Level 1	Level 2	Results
4 influencer + 2 institutes	5000 followers from each Level 0 users	5000 followers from each of the top 10% Level 1 users	Network Diagram of ~180,000 nodes, removing nodes with <2 degrees

Network Graph Visualisation

Simplified network graph by removing nodes with <2 degrees



In-Degrees of Nodes

Influencers:

Elon Musk	470
Changpeng Zhao	616
Vitalik Buterin	302
Nayib Bukele	100

Institutions:

Bloomberg	252
Goldman Sachs	144

Notable Others:

Brass_Statement	Defi payment network	182
JoseAMtnzRivera	Co-founder of XoloToken	102
Dontez Akram	Entrepreneur	112
Victor_eth0x	Cryptocurrency promoter	671
AltcoinWorldcom	Altcoins promoter	251
cdntradegrljenn	Equity Investor	162

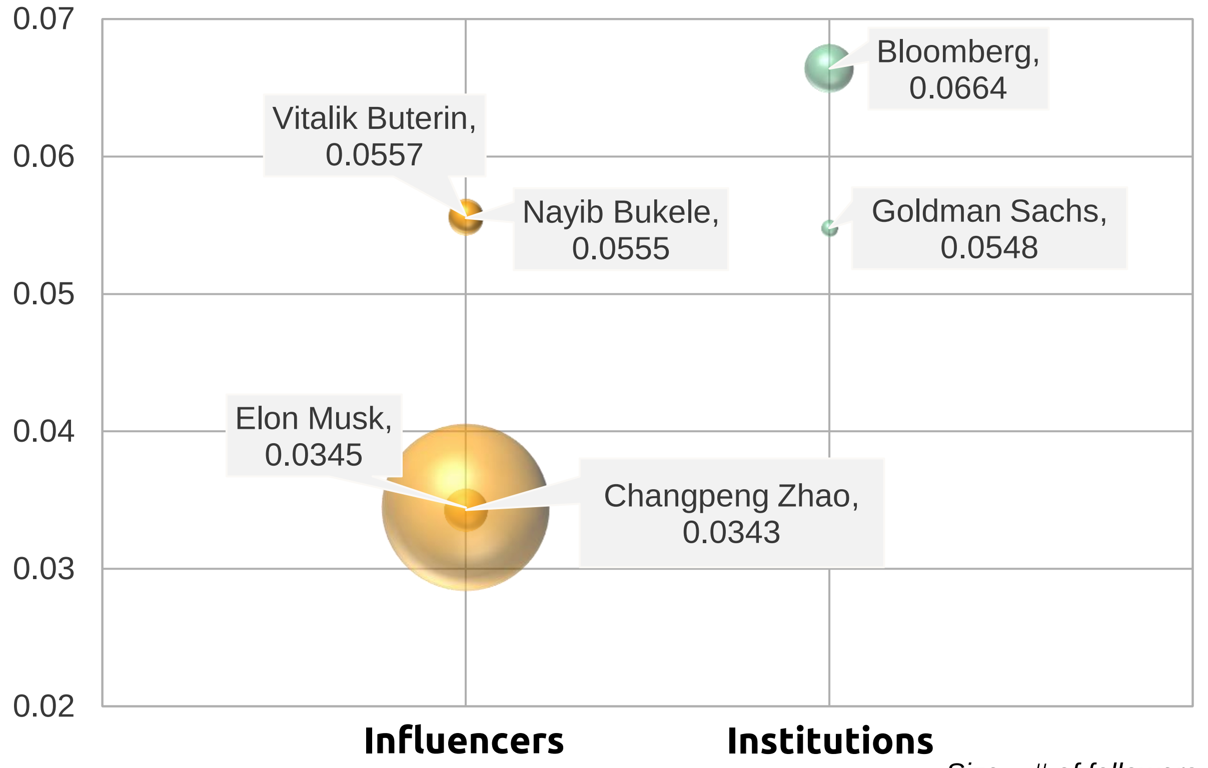
Observation:

Followers of 6 selected influencers + institutes tend to follow other accounts with similar focus.

Network Centrality Analysis

Level 0	Level 1	Level 2	Results
4 influencer + 2 institutes	5000 followers from each Level 0 users	5000 followers from each of the top 10% Level 1 users	PageRank is calculated based on a ~180,000-node network

PageRank Scoring



Size = # of followers

No. of Followers

Influencers:

Elon Musk	~ 98,000,000
Changpeng Zhao	~ 6,400,000
Vitalik Buterin	~ 4,000,000
Nayib Bukele	~ 4,000,000

Institutions:

Bloomberg	~ 8,300,000
Goldman Sachs	~ 937,700

Observation:

High follower count ≠ High PageRank Score

- Retail investors & speculators follow influencer accounts such as Elon Musk → **Lower PageRank Score**
- Finance professionals follow news channels and Financial Institutes such as Bloomberg and Goldman Sachs → **Higher PageRank score**

03

Sentiment Analysis

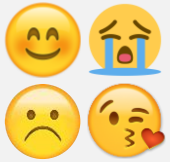
Step by Step Process of Analysis

Extracting Data

Twitter API
Python Library: Snscrape

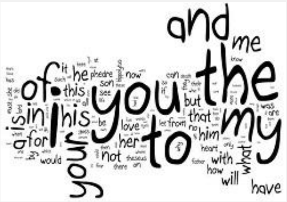
Pushshift Reddit API

Pre-Processing



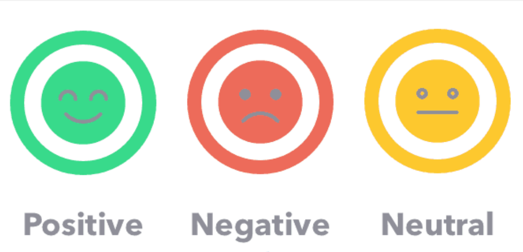
Emoji Handler

Language Detect: English



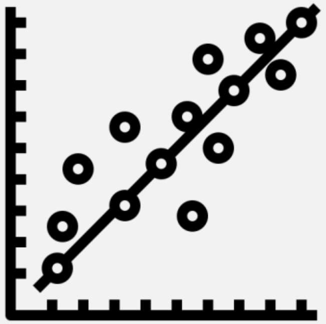
Stop Words Removal

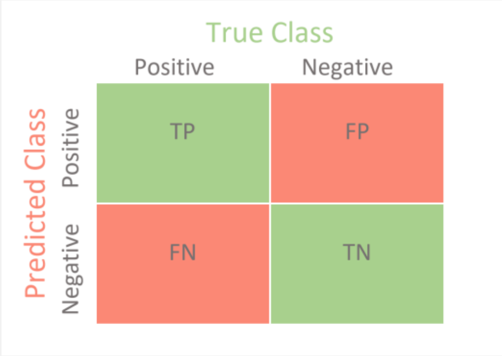
Sentiment Analysis



Compound Score

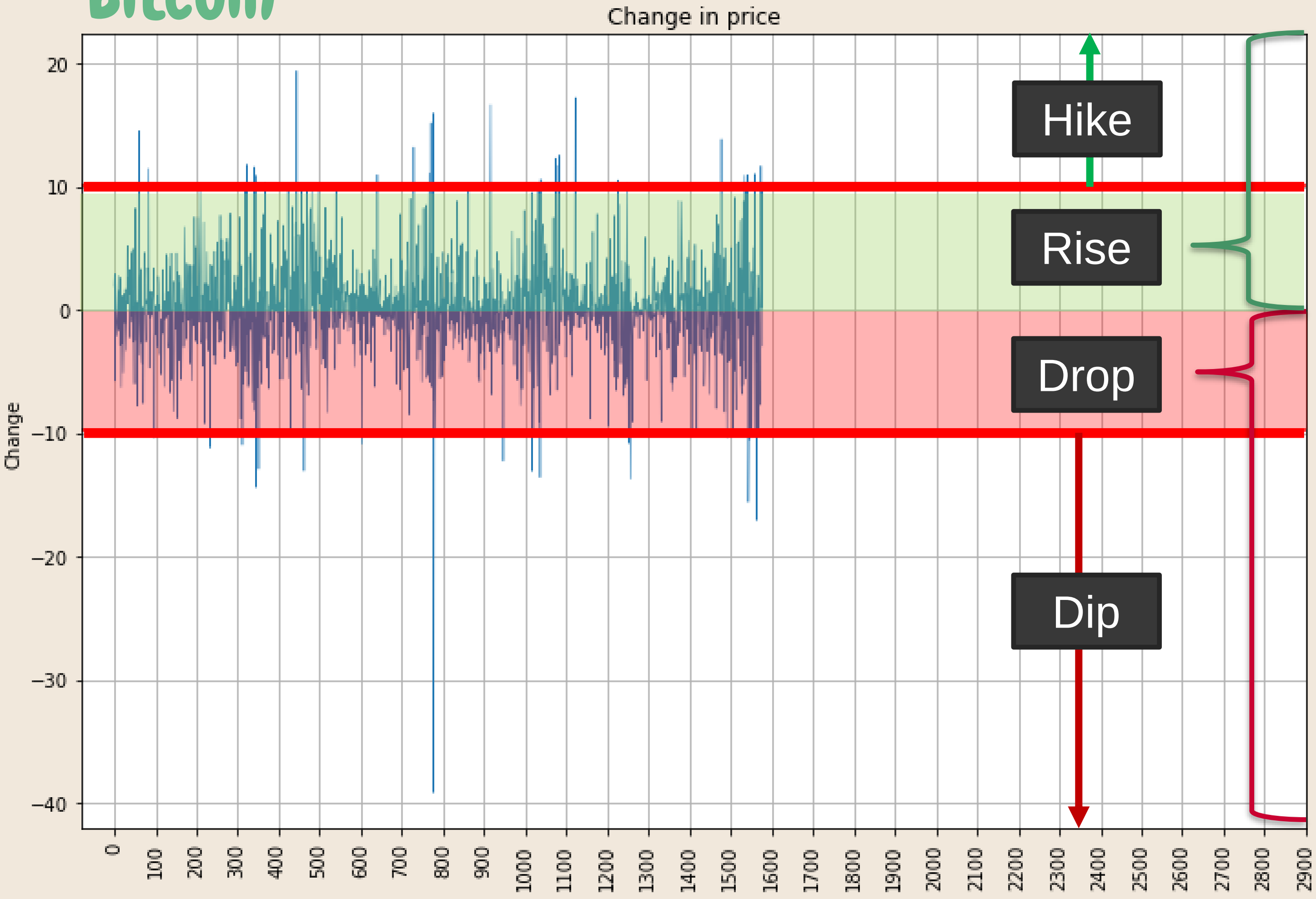
Results





Price Fluctuations

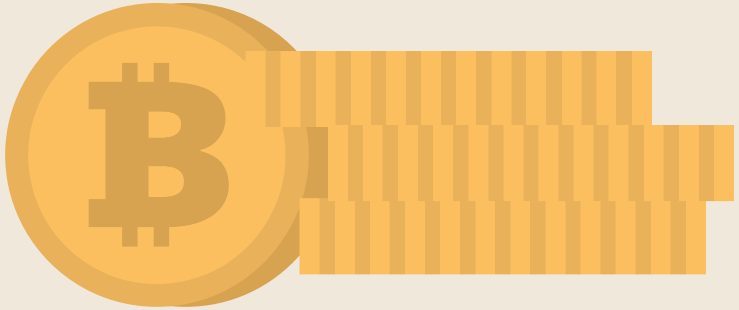
Bitcoin



- Any positive % change increase = Price Rise
- Positive % change > 2.5 SD = Price Hike

- Any negative % change increase = Price Drop
- Negative % change > 2.5 SD = Price Dip

- +-2.5 standard deviation from the mean [+10% / -9.8%]
- 27 days with large increase in price, 24 days with large decrease in price



On 9th June Changpeng Zhao Tweeted

- Bitcoin price increased by 11.83%
- Compound score of 0.89

^time shown here is Singapore time

Main idea!

Price Hike/Dip Days

Extract crypto related tweets on those days

Price Rise/Drop Days

Extract tweets on those days

Compound score

- **Price hike/rise** days should have **positive compound score**
- **Price dip/drop days** should have **negative compound score**

Influence Scoring (Computation)

	Actual Label = 0 (Price Dip/Drop)	Actual Label = 1 (Price Hike/Rise)
Predicted Class = 0 (-ve Compound Score)	True Neg **(TN)**	False Neg **(FN)**
Predicted Class = 1 (+ve Compound Score)	False Pos **(FP)**	True Pos **(TP)**

- Individual analysis of each influencer's tweet on days of price hike/dip with their relative compound score (+ve or -ve).

Influence Scoring (Bitcoin – Hike/Dip)

	Actual Label = 0 (Price Dip/Drop)	Actual Label = 1 (Price Hike/Rise)
Predicted Class = 0 (-ve Compound Score)	True Neg **(TN)**	False Neg **(FN)**
Predicted Class = 1 (+ve Compound Score)	False Pos **(FP)**	True Pos **(TP)**

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

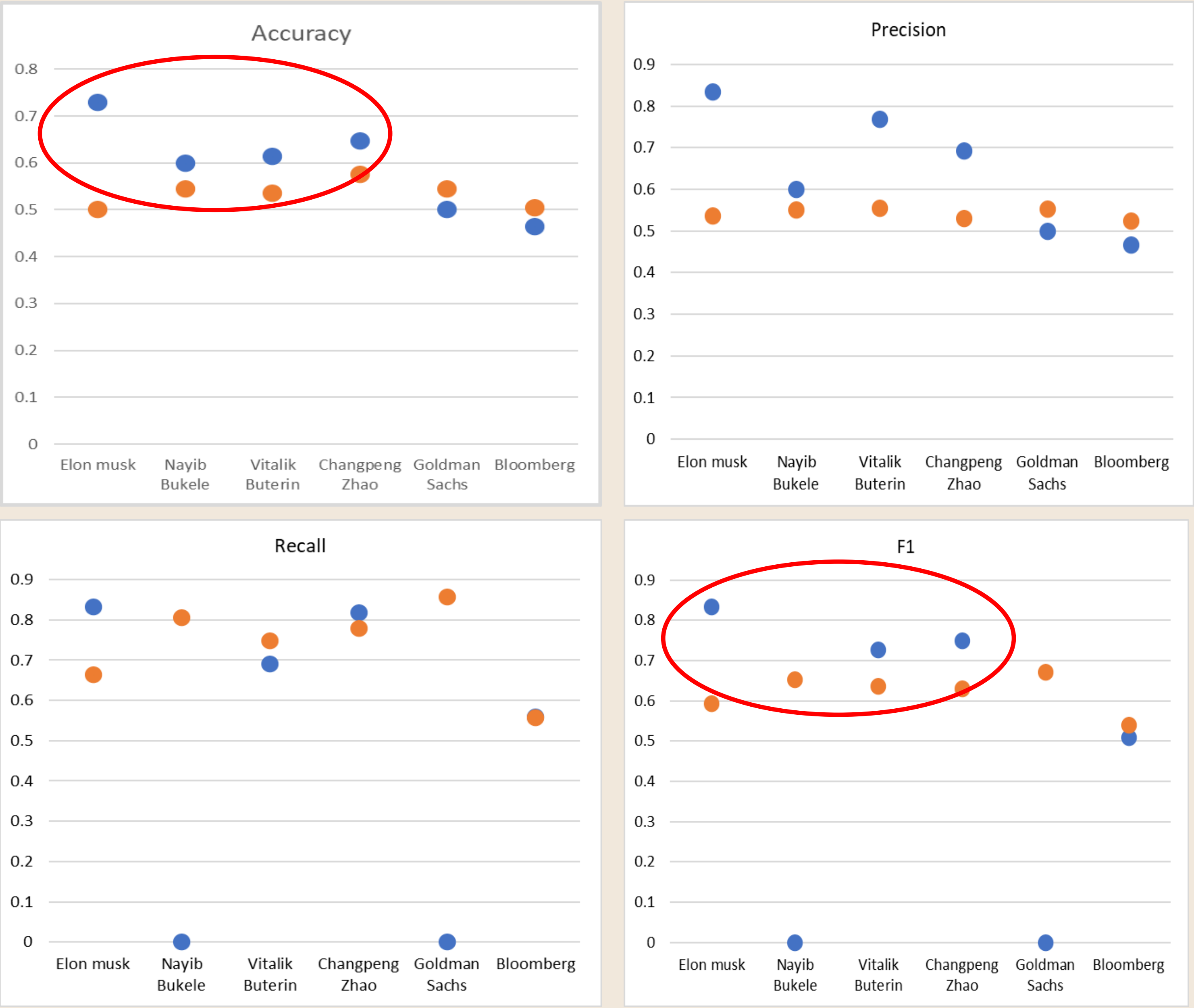
$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}}$$

	Elon Musk		Nayib Bukele		Vitalik Buterin		Changpeng Zhao		Goldman Sachs		Bloomberg	
	S.D	ALL DATES	S.D	ALL DATES	S.D	ALL DATES	S.D	ALL DATES	S.D	ALL DATES	S.D	ALL DATES
Accuracy	0.73	0.5	0.6	0.545	0.615	0.536	0.647	0.576	0.5	0.544	0.464	0.504
Precision	0.833	0.537	0.6	0.55	0.769	0.555	0.692	0.531	0.5	0.552	0.466	0.524
Recall	0.833	0.663	1*	0.805	0.690	0.748	0.818	0.778	1*	0.857	0.559	0.557
F1-Score	0.833	0.593	1*	0.653	0.727	0.637	0.750	0.631	1*	0.671	0.508	0.540

Confusion matrix



Hike/Dip

Rise/Drop

Generally, modelling based on hike or dip in price scored higher for accuracy and F1-scores.

Social Text Analysis (Compound Score w/ Confusion Matrix)

	Overall compound score (Price Dip)	Overall compound score (Price Hike)
Vitalik Buterin	0.99	0.96
Elon Musk	0.77	0.99
Nayib Bukele	0.82	0.92
Changpeng Zhao	-.10	0.99
Bloomberg	0.99	0.98
Goldman Sachs	0.80	0.81

During days of both price dips and price hikes, influencers’ tweets are generally very positive

What does this tell us?

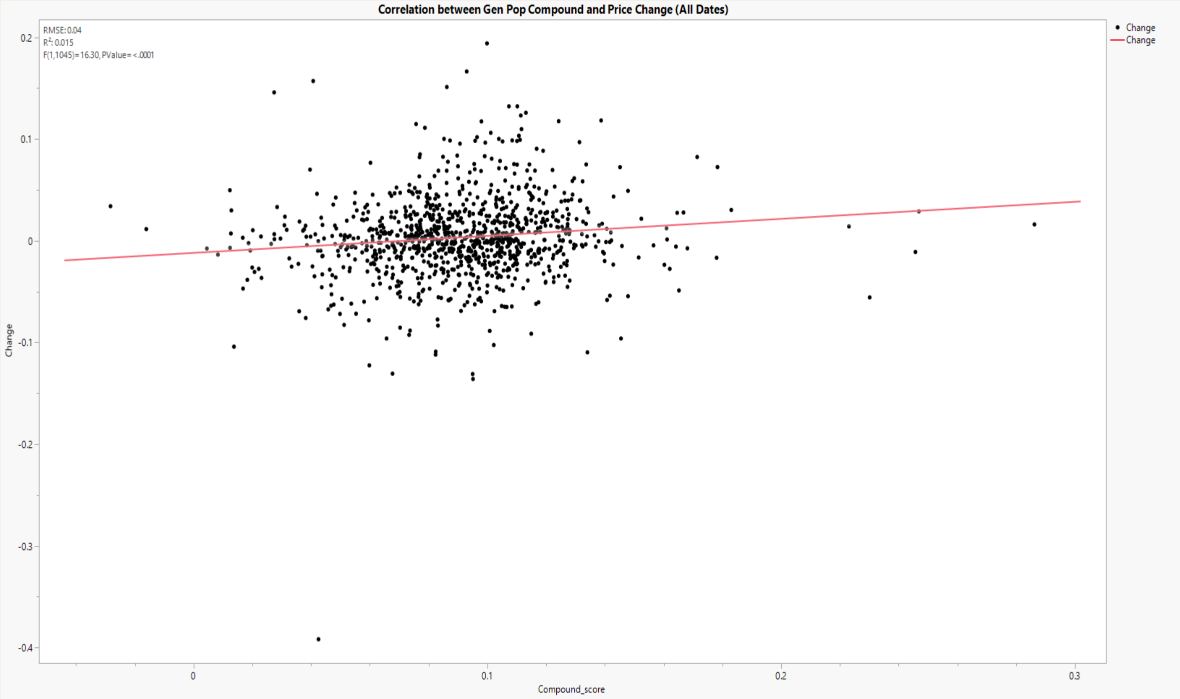
There is better accuracy and F1-scores between sentiments of crypto-related tweets by influencers/institutions to explain large changes in Bitcoin price, as compared to non-significant Bitcoin price changes.

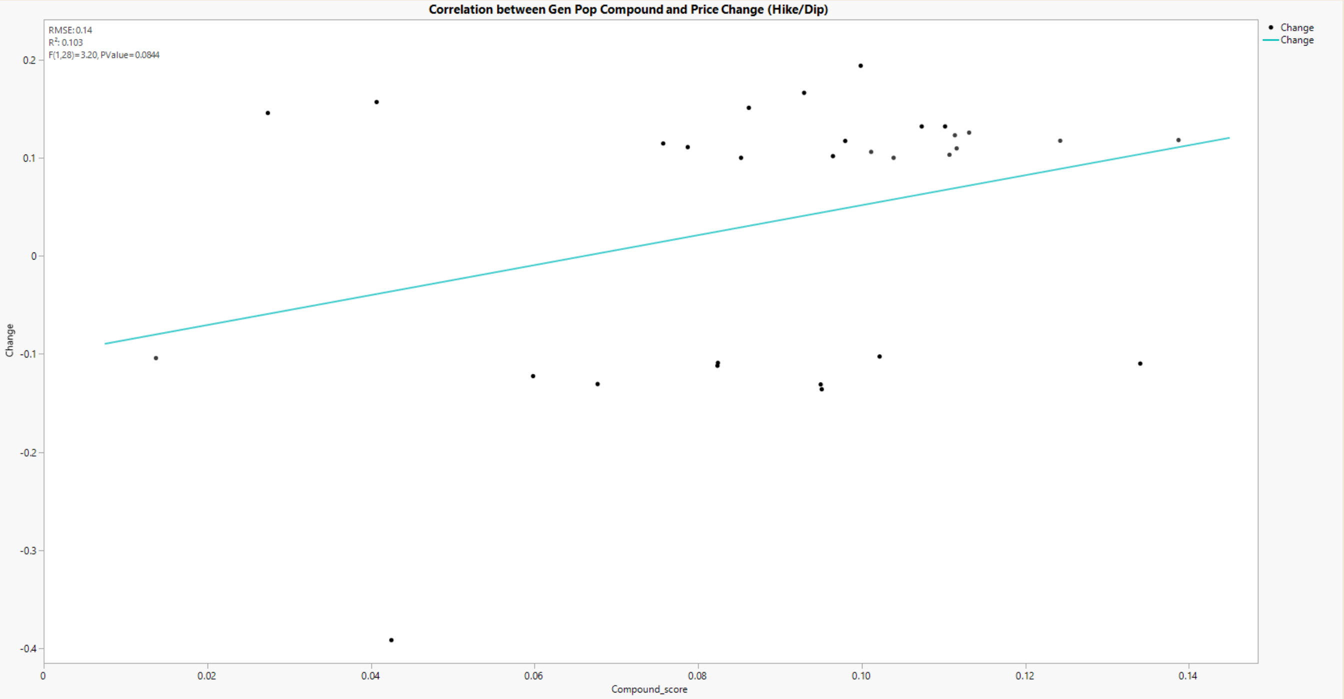
What does this tell us?

Influencers will still tweet positivity during days of large decrease, probably because of personal interest in the coin.

Twitter GenPop Insights

Hypothesis: “Prices of cryptocurrencies have a strong correlation with the sentiments of certain individuals, institutions or the general population”





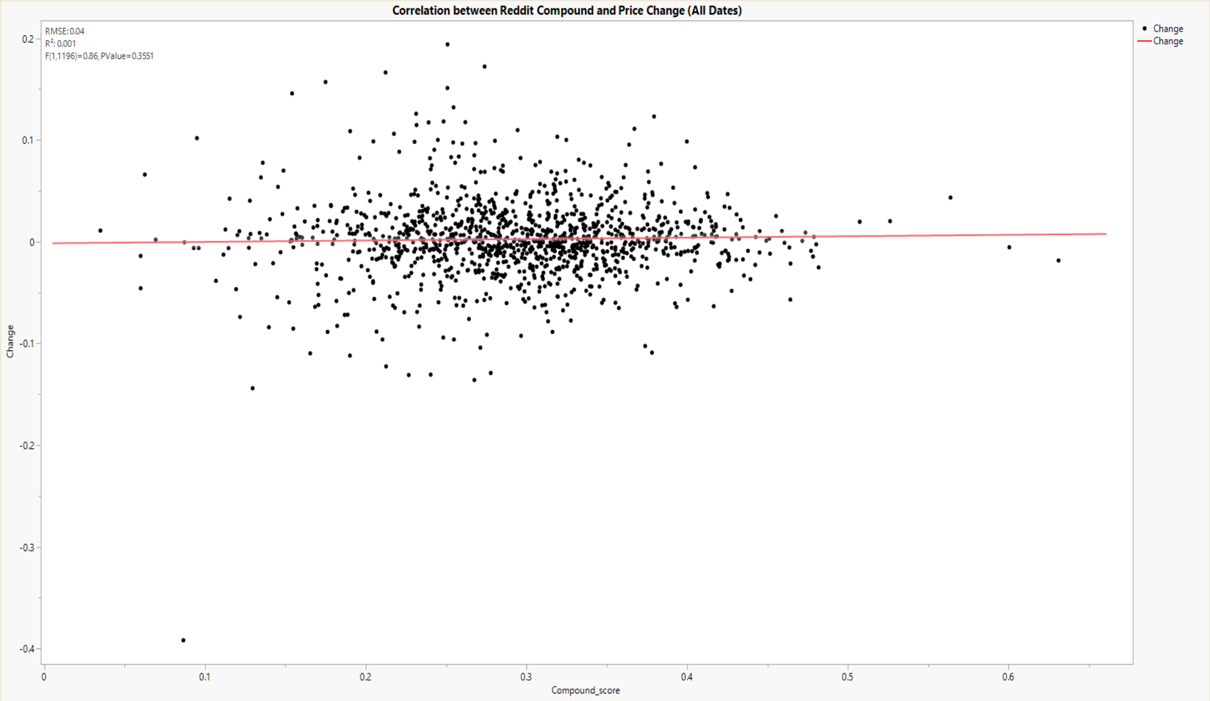
GenPop
$R^2 = 0.015$

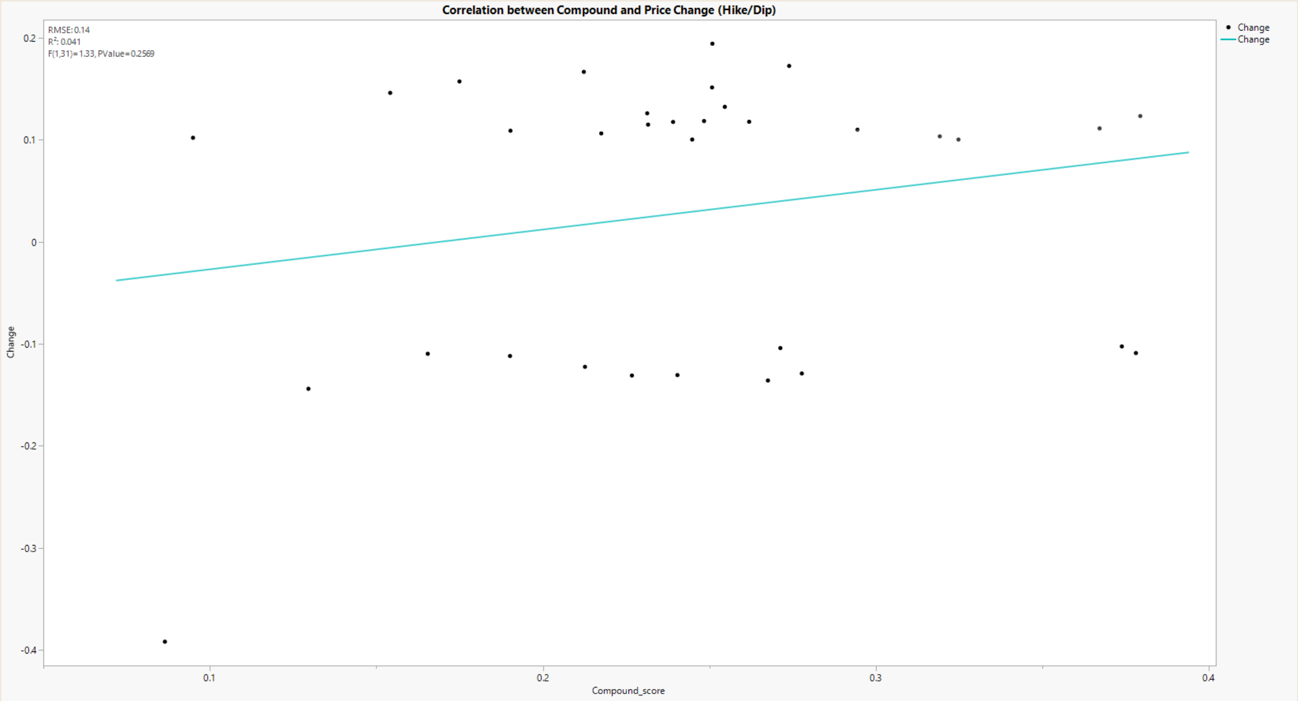
GenPop (Hike/Dip)
$R^2 = 0.103$

There is weak/no correlation between prices and the sentiments of twitter general population

Reddit Insights

Hypothesis: "Prices of cryptocurrencies have a strong correlation with the sentiments of certain individuals, institutions or the general population"





GenPop
$R^2 = 0.001$

GenPop (Hike/Dip)
$R^2 = 0.041$

There is weak/no correlation between prices and the sentiments of reddit general population

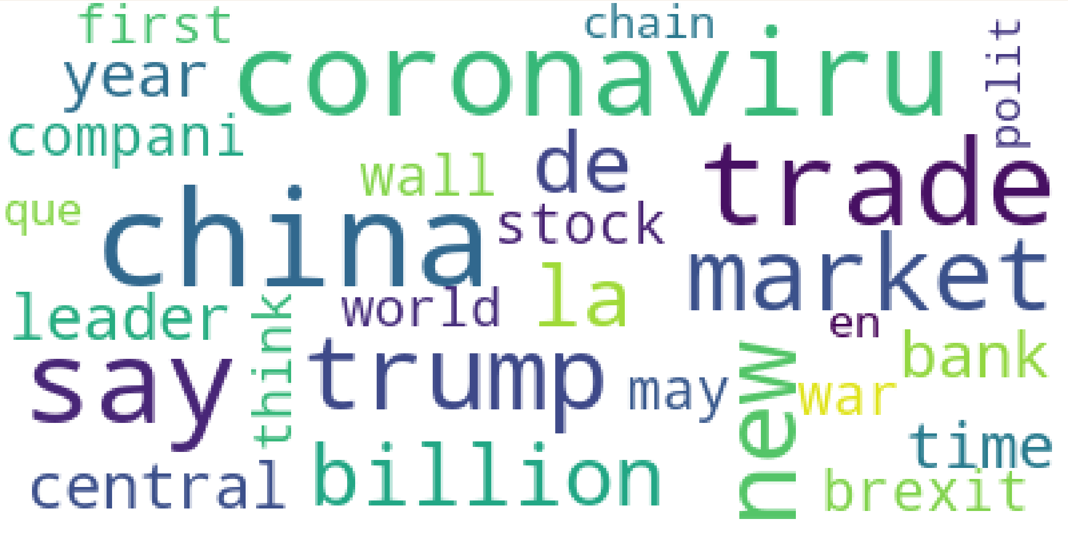
04

Topic modelling

For Bitcoin

Topic Modeling [Bitcoin days]

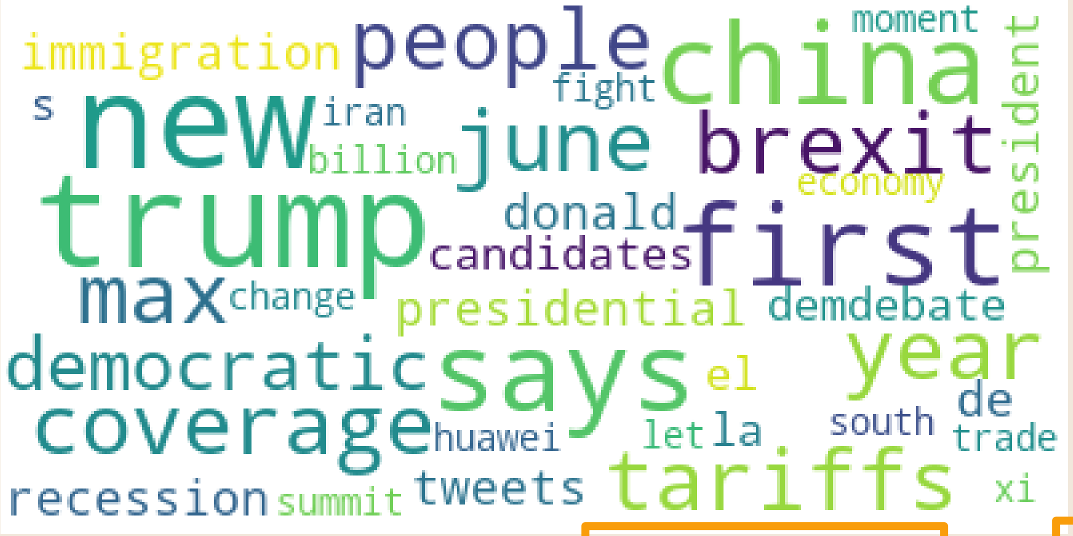
High days



Topics revolve around

- Geopolitics [Trump]
- Markets and finance
- Coronavirus

Low days



Topics similar to high days, with exception of

- Xi Jin Ping
- Recession
- Tariffs
- Economy

Trade war !

We can infer that crypto prices dropped significantly on days where news about trade war, tariffs, speculation about recession broke and spread to Twitter

The New York Times

Trump's Trade War With China Is Officially Underway

Give this article

632

05

Conclusion

Analysis and room for improvements

Analysis

- Influencers and Institutions do not **always** affect price change of Bitcoin.

- There is no correlation between general population sentiments and Bitcoin price change.

- You can never be **RICH** just by analyzing tweets and reddit post to predict Bitcoin price change!

Room for improvements for sentiment analysis

Model does not understand <u>context and meaning</u>

```
Tweet: 🎶 who let the doge out 🎶


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 0.0%
Negative: 0.0%
Neutral: 100.0%
Compound score (-1 to 1): 0.0
--------------------
```

```
Tweet: how much is that doge in the window?


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 0.0%
Negative: 0.0%
Neutral: 100.0%
Compound score (-1 to 1): 0.0
--------------------
```

- **Elon musk is hyping up Doge here = Very Positive**

Room for improvements for sentiment analysis

Model does not understand community created acronyms, words and phrases

```
Tweet: hodl the rainforests!!


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 0.0%
Negative: 47.4%
Neutral: 52.6%
Compound score (-1 to 1): -0.2023
--------------------
```

HODL stands for “Hold On For Dear Life”, which means to never sell your holdings

```
Tweet: doge barking at the moon


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 0.0%
Negative: 0.0%
Neutral: 100.0%
Compound score (-1 to 1): 0.0
--------------------
```

Doge barking to the moon implies a significant increase in price of Doge

Room for improvements

Variety of Influencers and Institutions

that have influence over the S&P 500

Correlation between the S&P 500 and BTCUSD

correlation

Jim Cramer

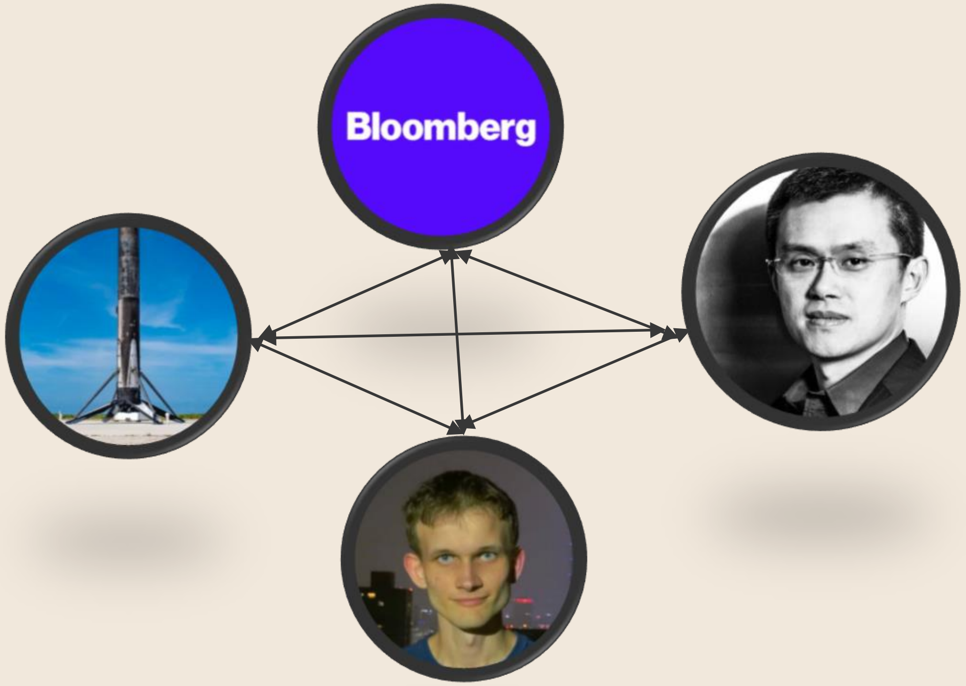
Fed Reserves

Influencers & Institutions

Room for improvements

Full Centrality Model for Influencers/Institution

with high computational CPU



Influencers & Institutions

Thank You

ANNEX

Influence Scoring (Bitcoin – Hike(Rise)/Dip(Drop))

$$\text{Recall} = \frac{TP}{TP+FN}$$

Elon Musk

	Actual Label = 0		Actual Label = 1	
	Dip	Drop	Hike	Rise
Predicted Class = 0 (-ve Compound Score)	1	24	2	33
Predicted Class = 1 (+ve Compound Score)	2	56	10	65

Changpeng Zhao

	Actual Label = 0		Actual Label = 1	
	Dip	Drop	Hike	Rise
Predicted Class = 0 (-ve Compound Score)	2	62	2	72
Predicted Class = 1 (+ve Compound Score)	4	222	9	252

Bloomberg

	Actual Label = 0		Actual Label = 1	
	Dip	Drop	Hike	Rise
Predicted Class = 0 (-ve Compound Score)	70	2295	82	2492
Predicted Class = 1 (+ve Compound Score)	119	2854	104	3138

Nayib Bukele

	Actual Label = 0		Actual Label = 1	
	Dip	Drop	Hike	Rise
Predicted Class = 0 (-ve Compound Score)	0	9	0	8
Predicted Class = 1 (+ve Compound Score)	2	27	3	33

Vitalik Buterin

	Actual Label = 0		Actual Label = 1	
	Dip	Drop	Hike	Rise
Predicted Class = 0 (-ve Compound Score)	4	174	9	185
Predicted Class = 1 (+ve Compound Score)	6	440	20	548

Goldman Sachs

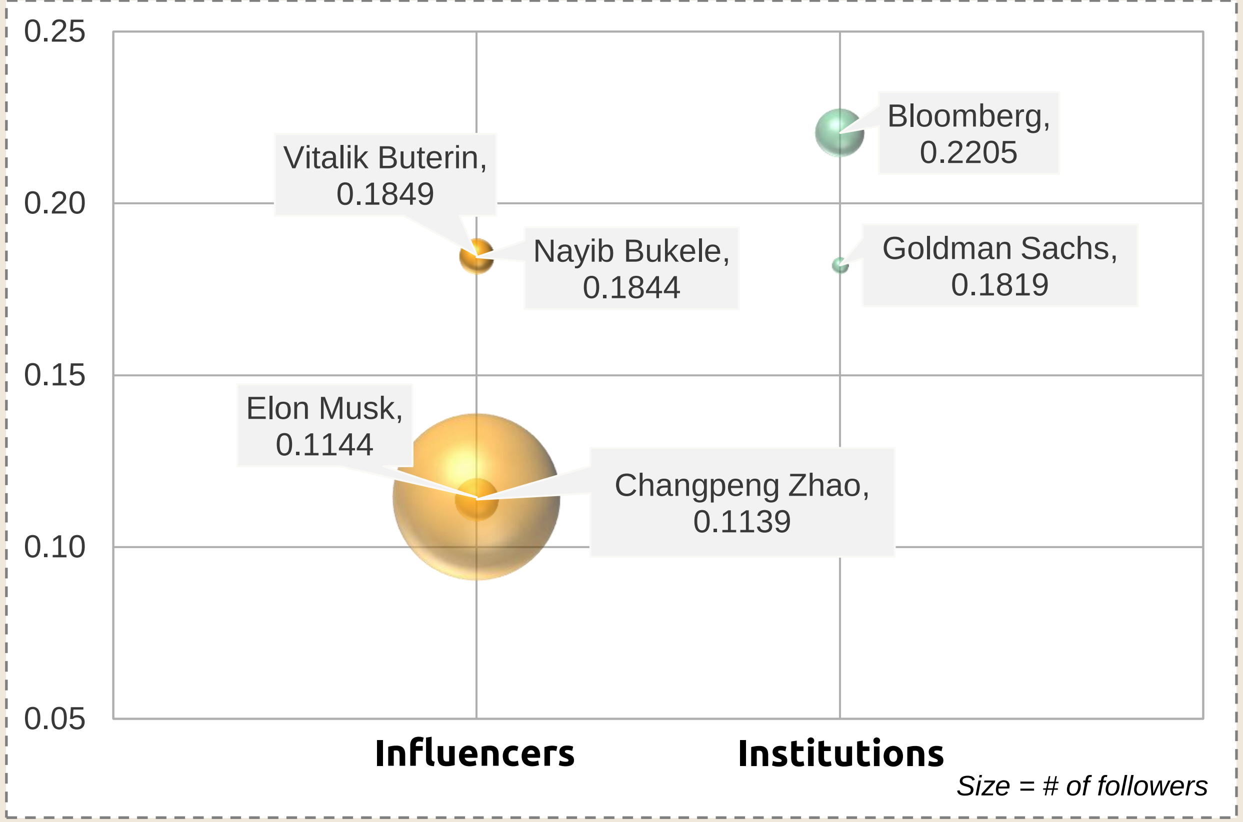
	Actual Label = 0		Actual Label = 1	
	Dip	Drop	Hike	Rise
Predicted Class = 0 (-ve Compound Score)	0	8	0	8
Predicted Class = 1 (+ve Compound Score)	1	39	1	48

Normalised PageRank

Network Centrality Analysis

Level 0	Level 1	Level 2	Results
4 influencer + 2 institutes	5000 followers from each Level 0 users	5000 followers from each of the top 10% Level 1 users	PageRank is calculated based on a ~180,000-node network, normalised by sum of all scores

PageRank Scoring



No. of Followers

Influencers:

Elon Musk	~ 98,000,000
Changpeng Zhao	~ 6,400,000
Vitalik Buterin	~ 4,000,000
Nayib Bukele	~ 4,000,000

Institutions:

Bloomberg	~ 8,300,000
Goldman Sachs	~ 937,700

Observation:

High follower count ≠ High PageRank Score

- Changpeng's PageRank score is similar to Elon Musk's despite having only 10% of the follower count

Topic Modelling data

```
Topics
0 : 0.009*"economi" + 0.008*"want" + 0.008*"minist" + 0.007*"econom" + 0.007*"china"
1 : 0.012*"market" + 0.010*"money" + 0.010*"get" + 0.009*"new" + 0.007*"cut"
2 : 0.018*"bank" + 0.013*"china" + 0.012*"coronaviru" + 0.010*"market" + 0.010*"trump"
3 : 0.008*"say" + 0.008*"new" + 0.006*"nation" + 0.006*"airlin" + 0.006*"leader"
4 : 0.012*"year" + 0.011*"coronaviru" + 0.010*"could" + 0.009*"new" + 0.008*"trade"
5 : 0.016*"coronaviru" + 0.012*"new" + 0.011*"china" + 0.008*"say" + 0.008*"market"
6 : 0.017*"trump" + 0.010*"war" + 0.009*"new" + 0.008*"trade" + 0.006*"china"
7 : 0.011*"billion" + 0.010*"stock" + 0.010*"market" + 0.009*"know" + 0.009*"day"
8 : 0.053*"de" + 0.031*"la" + 0.014*"en" + 0.014*"lo" + 0.013*"el"
9 : 0.017*"say" + 0.013*"trade" + 0.013*"tariff" + 0.009*"china" + 0.008*"war"
```

Topic distribution for some documents

```
--------------------
['volatil', "that'", 'push', 'nasdaq', 'futur', 'twice', 'last', 'year', 'rate', 'make', 'queasi', 'among', 'buy-and-hold', 'set', 'kind', 'market', 'trader', 'wait']
[(2, 0.18668051), (5, 0.18617712), (8, 0.590223)]
--------------------
```

```
--------------------
['hospit', 'overwhelm', 'understaf', 'rage', 'omicron', 'variant', 'swept', 'holiday', 'univers', 'hospit', 'new', 'jersey', 'public', 'acute-car', 'hospit', "state'", 'busiest', 'emerg', 'depart']
[(2, 0.8339531), (5, 0.12596214)]
--------------------
```

Bloomberg's Tweets

During days when price of bitcoin dropped, bloomberg still have positive tweets related to other topics, because it is a news network

Low Days

```
Tweet: retail traders are giving smaller coins like $eth a second look during $btc's volatile spring


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 32.9%
Negative: 0.0%
Neutral: 67.1%
Compound score (-1 to 1): 0.5994
--------------------
```

```
Tweet: here's why most of the world doesn’t use bitcoin as legal tender or a store of wealth


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 43.9%
Negative: 0.0%
Neutral: 56.1%
Compound score (-1 to 1): 0.5719
--------------------
```

High Days

```
Tweet: live: january jobs data is imminent.   consensus is a gain of 125,000 but there are whispers it could zero. follow for
analysis


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 23.6%
Negative: 0.0%
Neutral: 76.4%
Compound score (-1 to 1): 0.5267
--------------------
```

CZ_Binance Tweets during dip days

CZ_Binance tweets are sarcastic, which model cannot compute accurately.

```
Tweet: bitcoin rallies, "damn, i wish i bought in earlier"  a dip, "omg, it's all crushing to 0, let me get out. "  rallies aga
in. "damn,  i wish i bought that dip"  a dip, "omg, it's crushing again, let me get out"  ...


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 17.5%
Negative: 33.8%
Neutral: 48.7%
Compound score (-1 to 1): -0.6124
--------------------
```

Model also cannot pick up context

```
Tweet: if you panic about #bitcoin's current price, just don't.


Crypto Tweets Sentiments from this Influencer
--------------------
Positive: 0.0%
Negative: 52.4%
Neutral: 47.6%
Compound score (-1 to 1): -0.5106
--------------------
```