Twitter NLP & Crypto Coin Recommender

Version 0.01.0

Table of Contents

01

Intro

You can enter a subtitle here if you need it
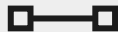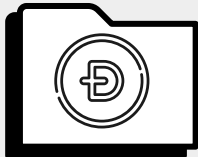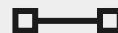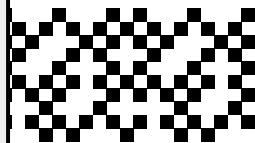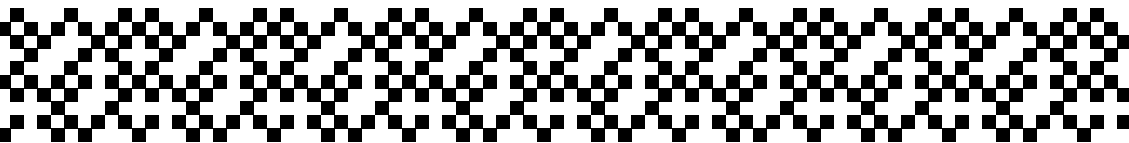
Who Am I?

Nicholas (Nick) Nguyen

- Data Scientist/Analyst
- Crypto Expert/Consultant (6+ years experience)
- Investor/Miner

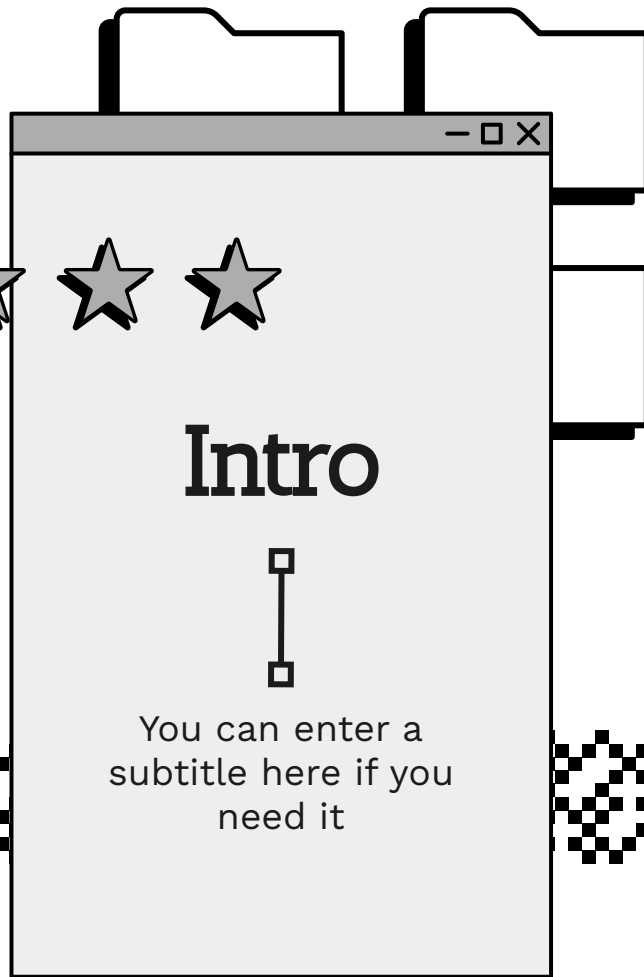Can there be a quicker and easier way to identify undervalued coins on Twitter?

API's

Tweepy

Following

200 counts per 15 minutes

CoinGecko

Coin Price

30 calls per min rate limit

SNScrape

Tweets & More!

No Rate Limit

Methodology:

User inputs: @GemsOfRa (can be anyone)

tweepy ● Scrape #1. All users that user is following
snscrape ● Scrape #2. All tweets within 30 days from
users following list
● Identify coins from those tweets ($ex.)

snscrape ● **Scrape #3**. All tweets involving top 20
coins within 90 days from following list

snscrape ● **Scrape #4**. All tweets involving top 20
coins within 90 days from entire twitter

snscrape ● Scrape #5. All chinese tweets involving top
20 coins within 90 days from entire twitter

Box of $coins
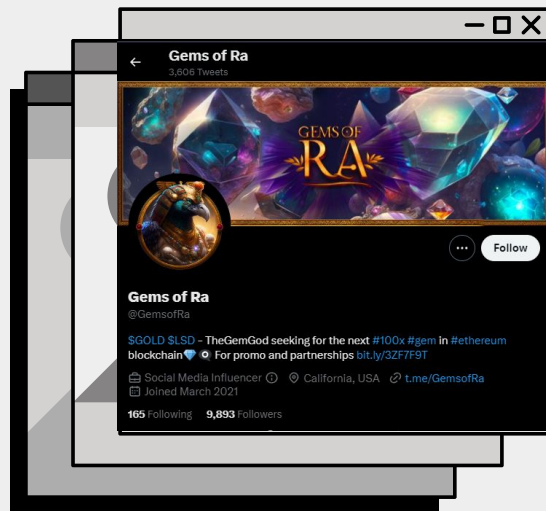
Data Dictionary

Date	Username	Followers
Tweet	Sentiment	Coins Extracted
Sentiment Rank	Language	Translated

02

Cleaning

Data engineering, cleaning, and processing

Translate for NLP

Google Translate

Identify Language

Create Translate Column

Tweets

5,000 characters per request

Language Column

Translated Tweets
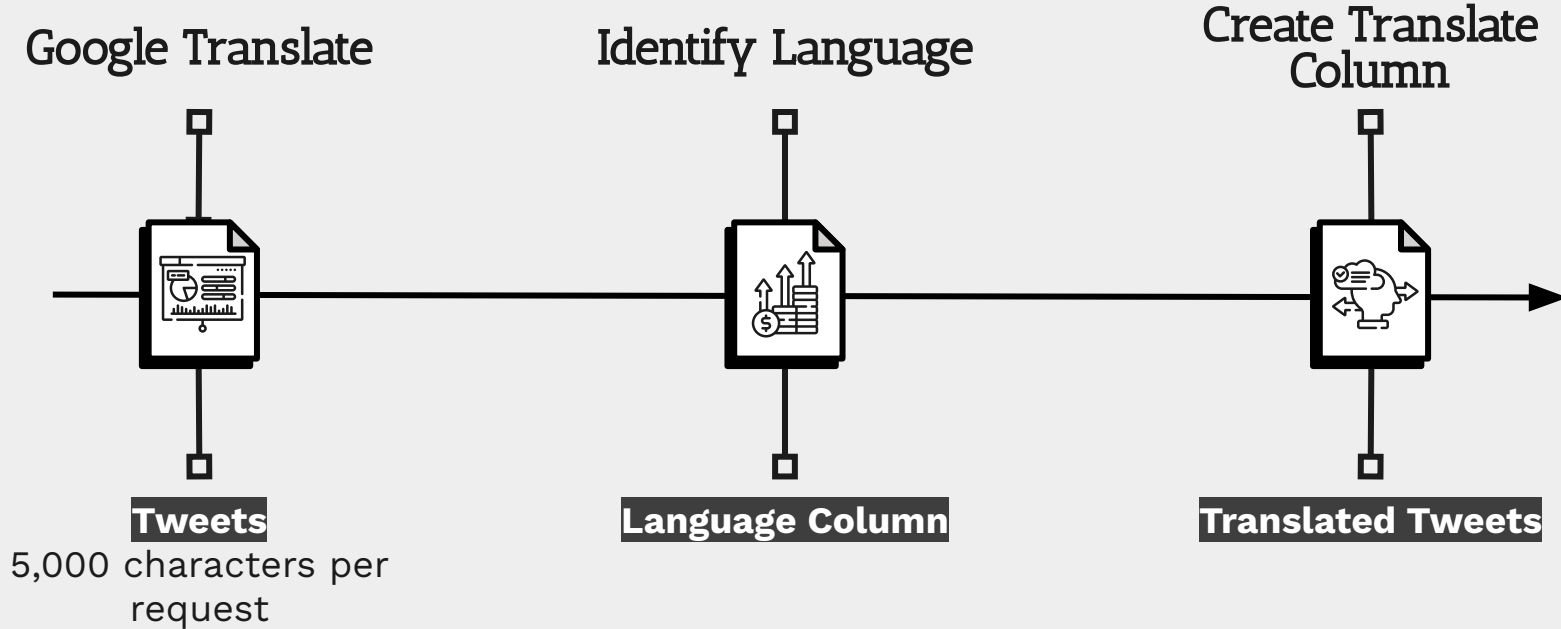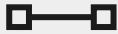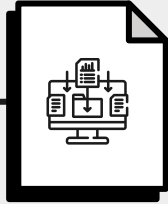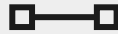
Data Engineering for NLP

Tweets

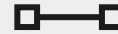Remove: special characters, links, hashtags, emojis

Find Coins

Extract words that begin with '$' & special function for those without

Drop Tweets

If they don't involve a coin, drop em!

.explode()

Assign tweets with multiple coins to separate tweets

Preprocessing Tweets w/Lemmatization
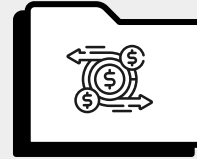
Take out:

- Words that involve the coin regarding tweet

- English Stopwords

Raw Tweet

"i'm so sick of this i sold all my $ARB and shorted bc of shanghai and now im about to get liquidated at $20"

Result

"sick sold shorted shanghai get liquidated"

Sentiment Analysis

< 0

Negative Sentiment

"the arb for $usdc to $usdt is insane along with the $eth gas fees yikes"

> 0

Positive Sentiment

"just look at that weekly $sani chart and look at the macro environment there is no way you can't be bullish on this. 100m+ is programmed people have 0 clue about fomo. i've experienced it in 2 bull cycles there is nothing comparable."

= 0

Neutral Sentiment

"given the changes in stable coins and banks, will convert the remaining of the $1 billion industry recovery initiative funds from busd to native crypto, including , and eth. some fund movements will occur on-chain. transparency."

Sentiment Ranking (experimental)

This algorithm allows for those who have more followers to have a more important sentiment.

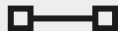I have made this calculation by the formula:

- **Sentiment = Sentiment * (0.1 * log(# of followers))**
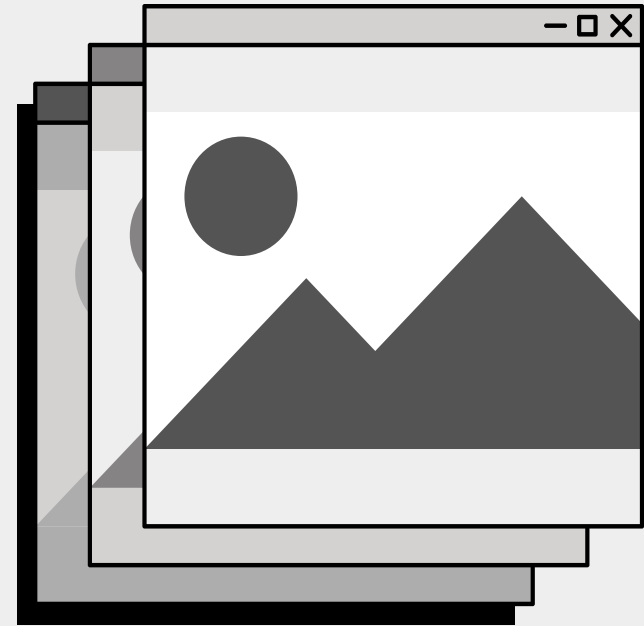
TFIDF Vectorizer

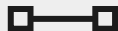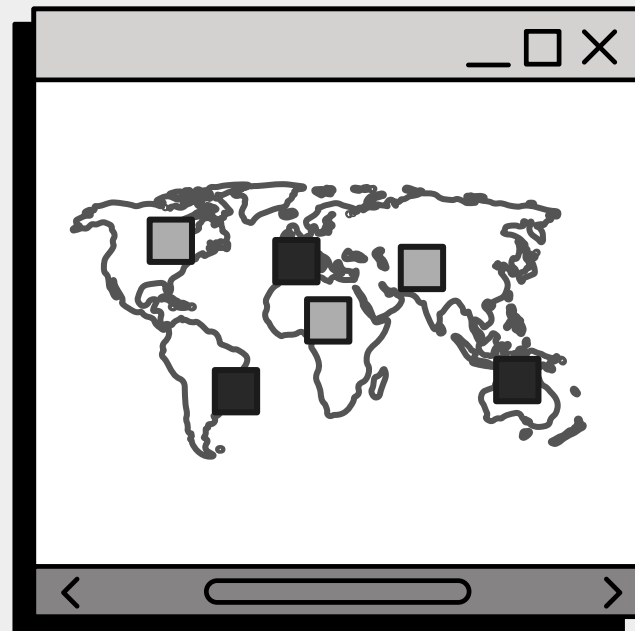**Positive**

◻—◻

Sentiment > 0.3

Negative

◻—◻

Sentiment < 0.3

Tweets from Following List

10 Most Important Words for Positive Sentiment Tweets

10 Most Important Words for Negative Sentiment Tweets

Importance Level

Tweets from Entire Twitter

10 Most Important Words for Positive Sentiment Tweets

10 Most Important Words for Negative Sentiment Tweets

Importance Level

03

EDA
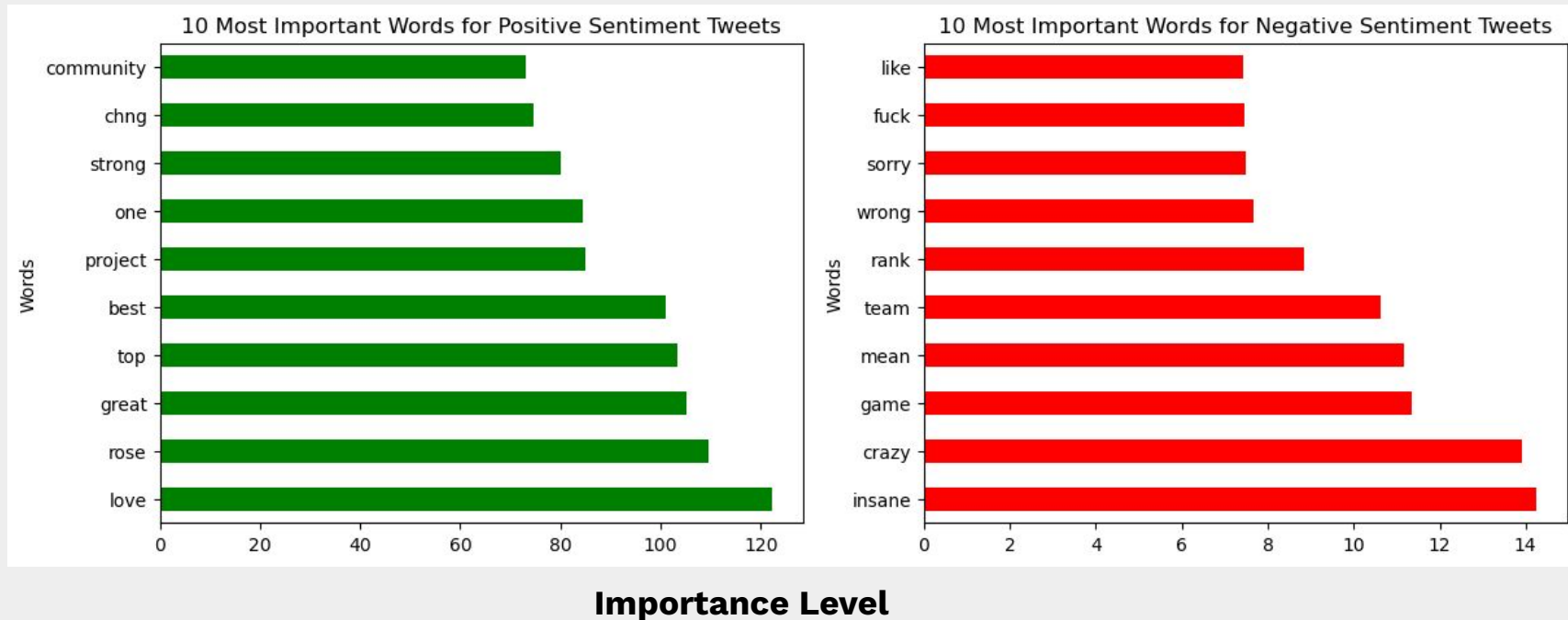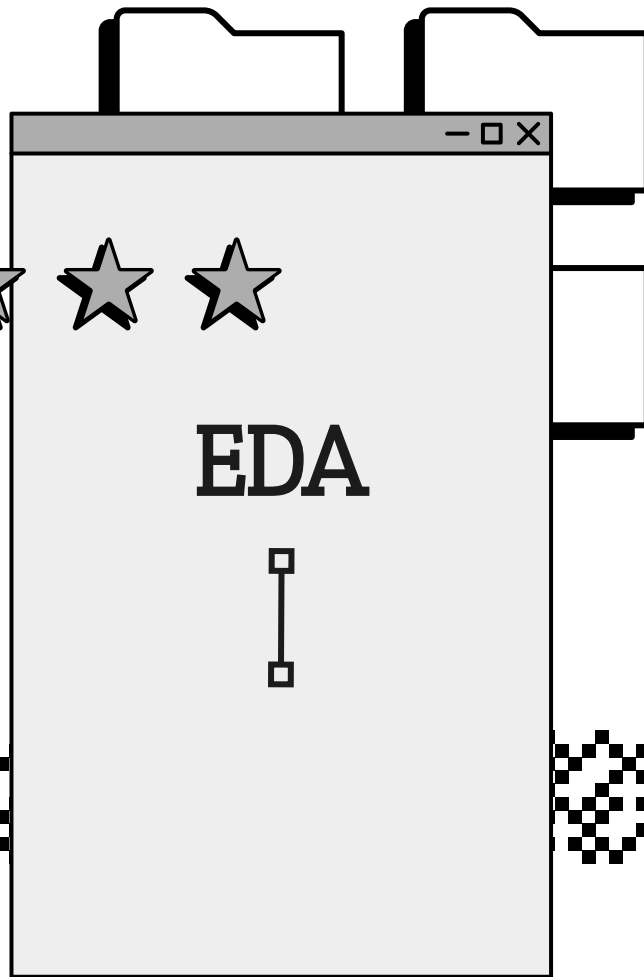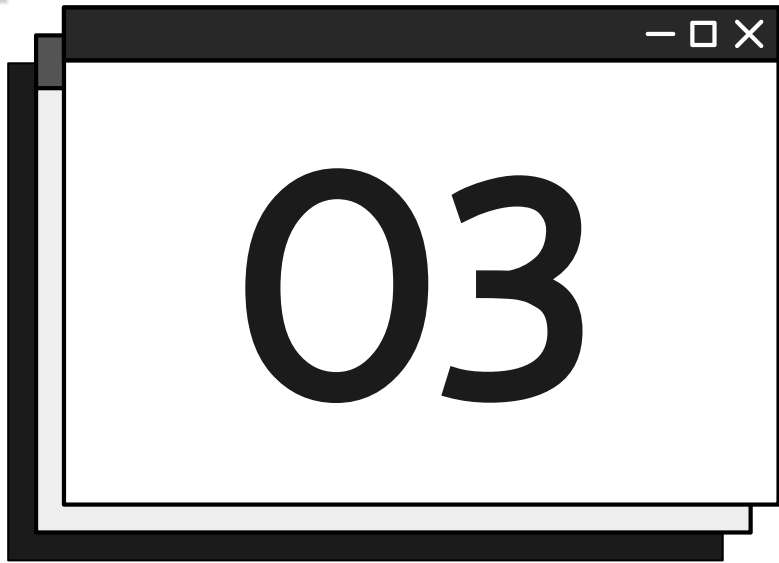
Top 10 Most Talked About Coins

Entire Twitter

Coins	Counts
$gold	~1200
$rxd	~1380
$atlas	~1400
$opti	~2100
$bag	~2120
$swirl	~2150
$arb	~2220
$sani	~2270
$ic	~2280
$eth	~2550

Based Off Following

Coins	Counts
$gold	~140
$arb	~150
$atlas	~150
$lsd	~330
$sani	~390
$toshi	~400
$ic	~460
$swirl	~650
$bag	~680
$eth	~750

Counts

To Tableu!

04

Recommender
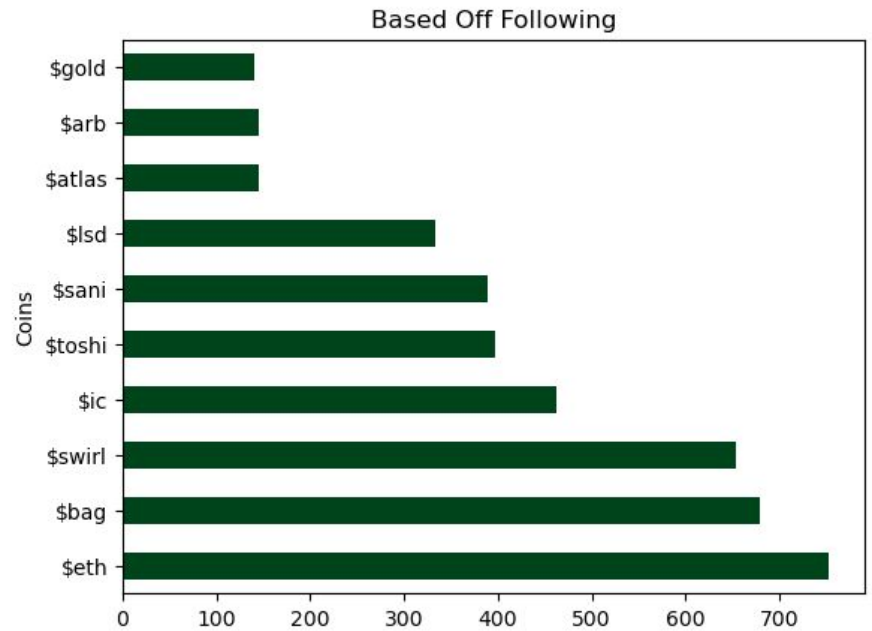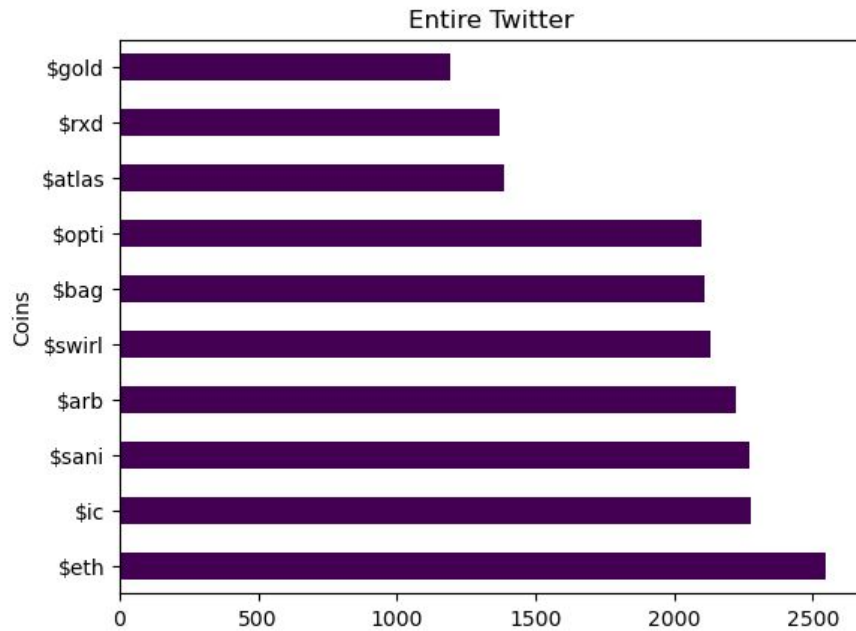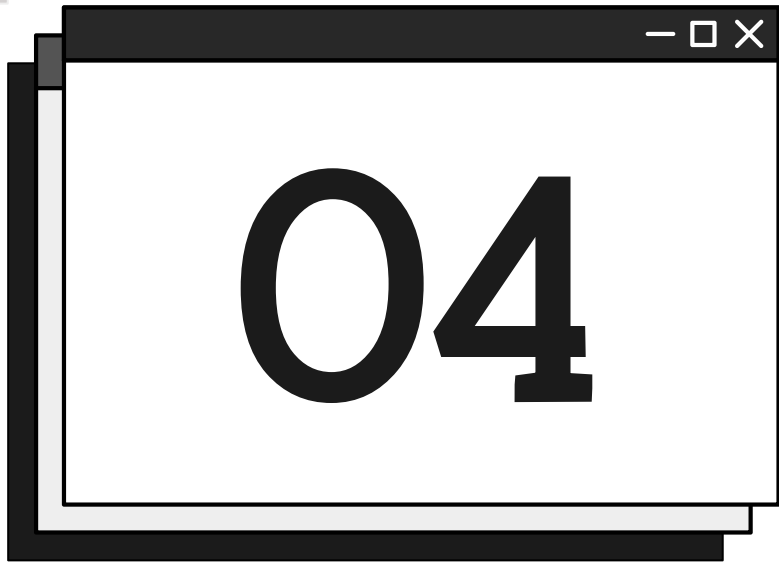
How Do We Identify Top 3 Coins

Easy

Top 3 most talked about based on following

Medium

Highest overall mean sentiment based on who you follow

Hard

Mean Sentiment from following - Mean Sentiment from entire twitter

This shows coins that are liked by who you follow but disliked by the public
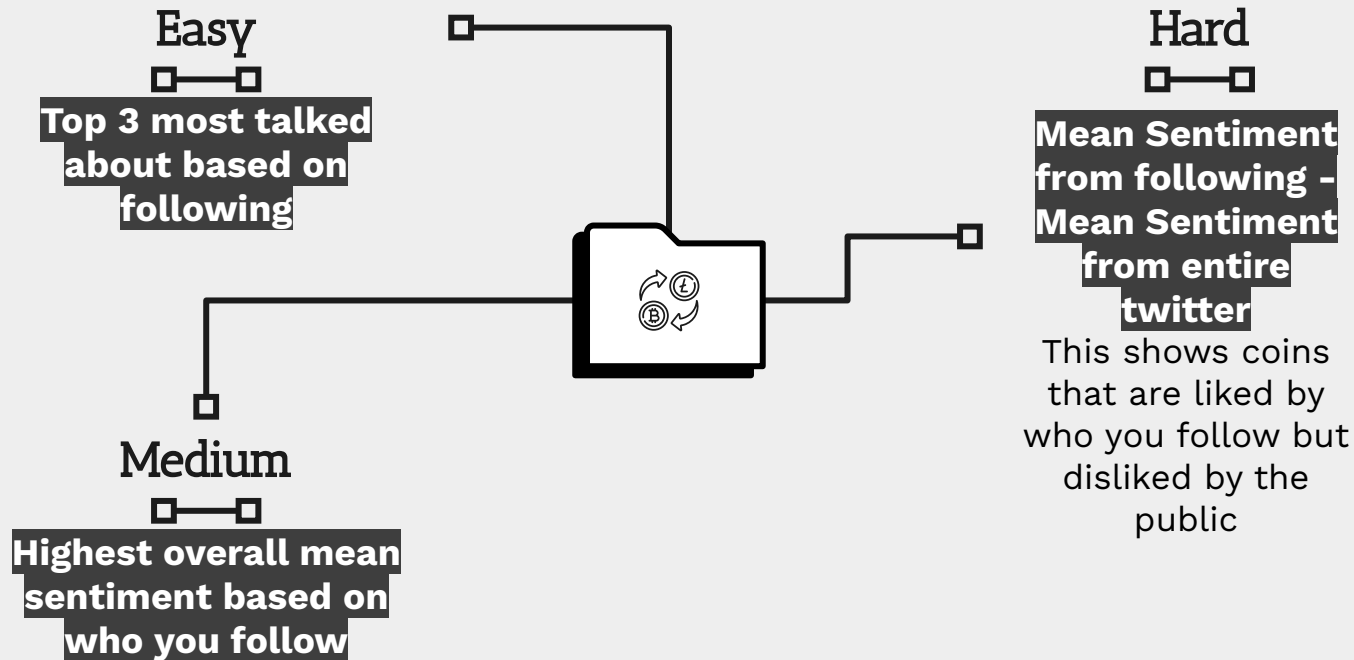
$ALGO/Algorand

* Sentiment 3 day rolling average

$PEPE/ Pepe

$FINU
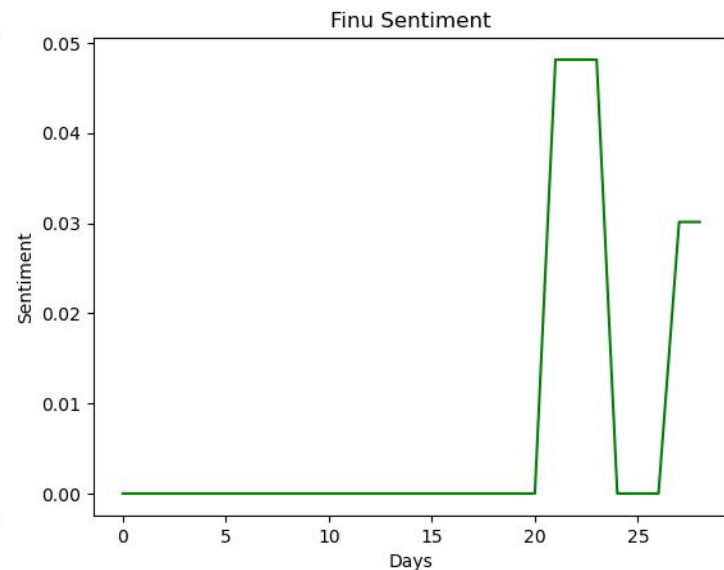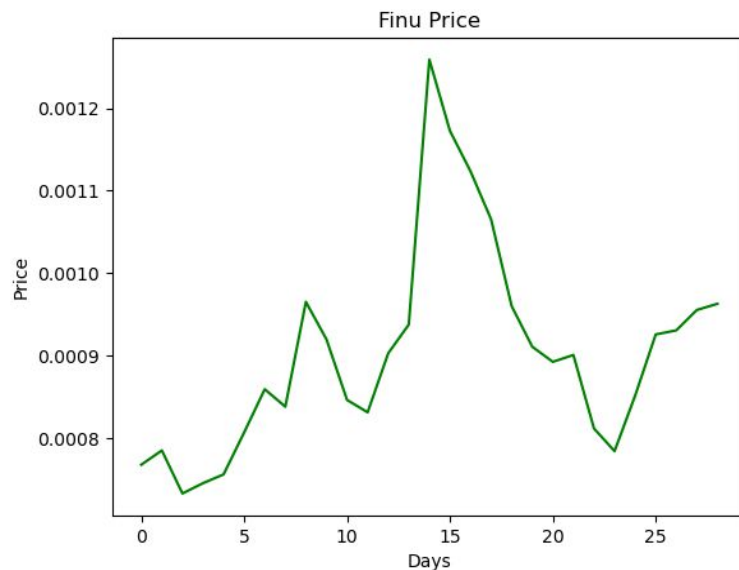
What's Next?

Modeling
Run XGboost Classifier to predict if go up/down

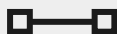### Forecasting
Forecast the coins we recommended

Trends
Find keywords which will help identify coins

Recommender
Combine price with sentiment

Sentiment Rank
Specific user's tweet history vs price instead of follower #
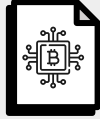
Sell Indicator
Sell if sentiment is too high?
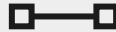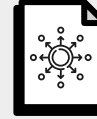
What's after What's Next?

More Data

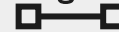Google Trends, Blockchain Monitor

Even More Data

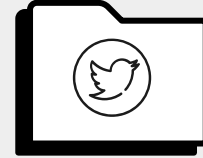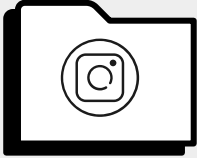Volume, BTC/ETH Dominance, Fear/Greed Index

Fund Allocator / Risk Management

How much should be invested given the risk?

Automation

Prefect ... Automate this process

Thanks!

Do you have any questions?

knicknguyen@gmail.com