

Exploring the Correlation between Twitter Sentiment and Price Movement

Author: Isaac Lee

Date: 26/02/2022



[Source Code](#)

Exploring the Correlation between Twitter Sentiment and Price Movement

Introduction

Gathering the Data

Price Data

Twitter Data

Calculating Sentiment

Exploring Price Movement Vs. Twitter Data

Sentiment Compound Scores Vs. Price Movement

Tweet Count Per Day Vs. Price Movement

Summary

Introduction

The main aim of this report is to answer the following questions:

Questions

1. Is there a correlation between Twitter sentiment regarding a cryptocurrency and its price movement?
2. Does the *usefulness* of Twitter data depend on the cryptocurrency in question?

We will look at the past five years for a time frame and examine three different coin pairs:

- **USDT_BTC** (Bitcoin)
- **USDT_ETH** (Ethereum)
- **USDT_XRP** (Ripple)

First we gather the necessary data using various APIs, analyze the results using Python and then interpret our findings.

Gathering the Data

Price Data

In order to get pricing data we send a request to the Poloniex API ¹. We ask for thirty minute candle data, from the past year. Note that this API does not support hourly candles, so instead we aggregate our thirty minute candles into hourly candles. Also we are limited to how far back we can go, so instead of asking for the past five years all in one request, we use a for loop and loop over each year, sending five separate requests, one for each year. We convert the response data into a Pandas DataFrame object. The first few rows are given by:

	weightedAverage	close	high	low	open	volume
2017-01-01 01:00:00	0.006486	0.006498	0.006498	0.006481	0.006487	15.481315
2017-01-01 02:00:00	0.006498	0.006498	0.006498	0.006498	0.006498	0.000000
2017-01-01 03:00:00	0.006560	0.006563	0.006563	0.006551	0.006551	18.741457

(Note: the index is a Pandas time index object).

Twitter Data

The Twitter data was slightly harder to obtain, since Twitter now charges for users to access historical data on tweets. Also the official Twitter Python API has very harsh limits on how far back we can get tweets from and the number of requests allowed. So instead we turn to a really nice Tweet scraping python package, called "TWINT" ² which stands for:

- (TW)itter,
- (IN)elligence,
- (T)ool.

It is much nicer than the official Twitter API with the following benefits:

- Can fetch almost **all** tweets (Twitter API limits to last 3200 tweets only);
- Fast initial setup;
- Can be used anonymously and without Twitter sign up;
- **No rate limitations.**

However the issue with TWINT is that we want to go back five years and currently the from and until parameters that you can pass to your TWINT search seem to not work, so instead we use a little known Twitter feature that allows you to filter your search directly with a query, i.e if we input the following into the twitter search bar then we can filter tweets by date, language and minimum number of retweets. We can also filter *out* links and replies which are not useful to us:

```
{search_query} min_retweets: n lang:en until:{end_date} since:{start_date} -
filter:links -filter:replies
```

So then we incorporate this method of filtering by day with a for loop which loops over each day in the year and sends a separate request to get tweets from that day which mention the cryptocurrency pair we're interested in.

Note that we set the minimum number of retweets to a sufficiently large number so that we can only scrape *important* tweets.

Once we have all the tweets for each day for each year we convert to a Pandas DataFrame and save as a CSV.

The first entry of our tweet DataFrame looks something like this (with the actual tweet highlighted):

	txt	name	username	likes	retweets	replies
2021-12-29	The main reason why #XRP will become the most ...	Blood	bloodorcrypto	1741.0	303.0	111.0

(We also perform some data cleaning, removing all non-alphanumeric characters from each tweet).

Calculating Sentiment

In the interest of time we use a pre-trained sentiment analysis model from the Natural Language Tool Kit Python library, which contains many great tools for natural language processing and modeling.

We use the **Vader** model from NLTK, which stands for:

- (V)alence;
- (A)ware;
- (D)ictionary;
- for s(E)ntiment;
- (R)easoning.

Without going into too much detail of how Vader works³, the main idea is that it uses a dictionary with a large number of common phrases/words that were rated by humans on their polarity and intensity. Also several grammatical heuristics such as contrasting conjunctions and intensifying adverbs are used.

A compound score is calculated, which is the sum of the scores of the sentiment features found in the text, normalized to fall between 0 and 1.

Also we get the percentage of (**neu**)tral, (**pos**)itive and (**neg**)ative sentiment features found in the text as **neu**, **pos** and **neg** outputs.

Exploring Price Movement Vs. Twitter Data

So now that we have all our data and sentiments calculated we can finally try and answer the question: is Twitter data correlated with price movement? Also does this depend on the cryptocurrency of choice?

Sentiment Compound Scores Vs. Price Movement

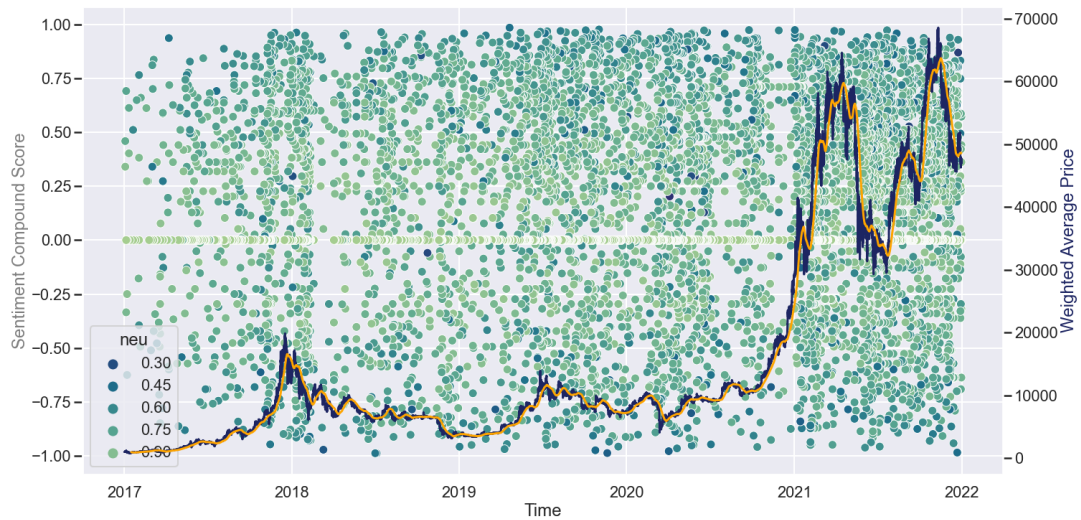
We begin by looking at the overall compound scores for each tweet for the past five years.

Each point on the plot represents a tweet. For each tweet we can see its compound sentiment score, with a positive score meaning a positive sentiment and a negative score a negative one. Each point is also given a *hue* which is a different shade from dark blue to green corresponding to how neutral the tweet was, so if a tweet is very light green then although it may have a positive or negative compound score, it has a largely neutral sentiment, meaning it may not be relevant.

Firstly we look at **USDT_BTC**:

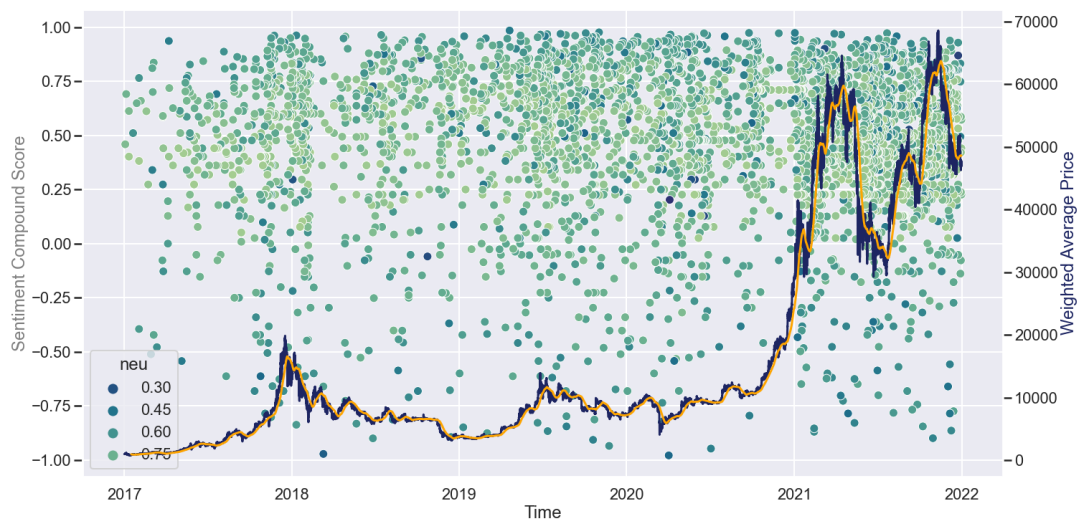
All tweets:

USDT_BTC Weighted Average Price Vs. Twitter Sentiment

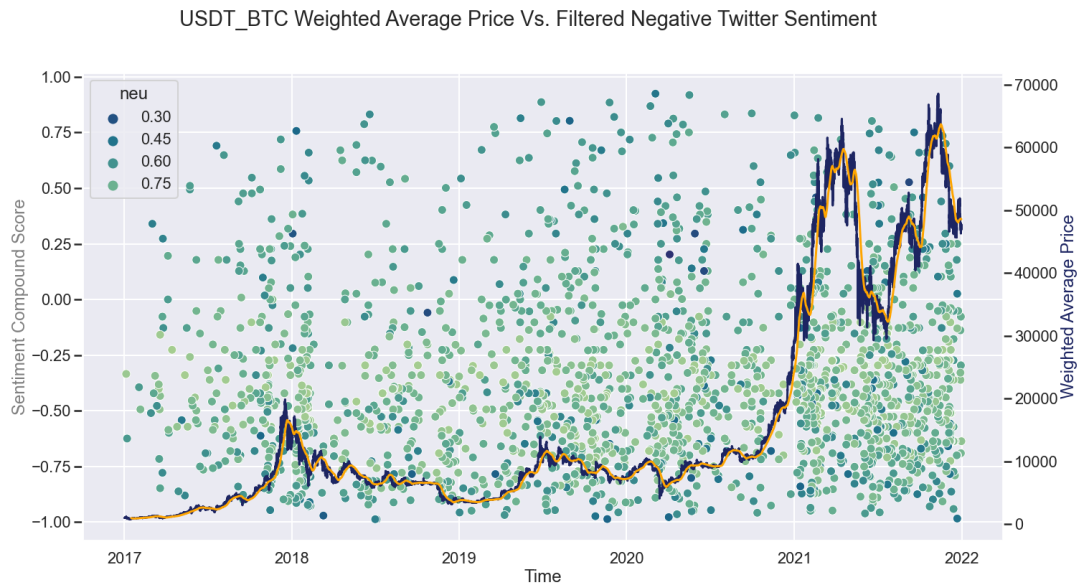


Only tweets with positive compound sentiment scores:

USDT_BTC Weighted Average Price Vs. Filtered Positive Twitter Sentiment



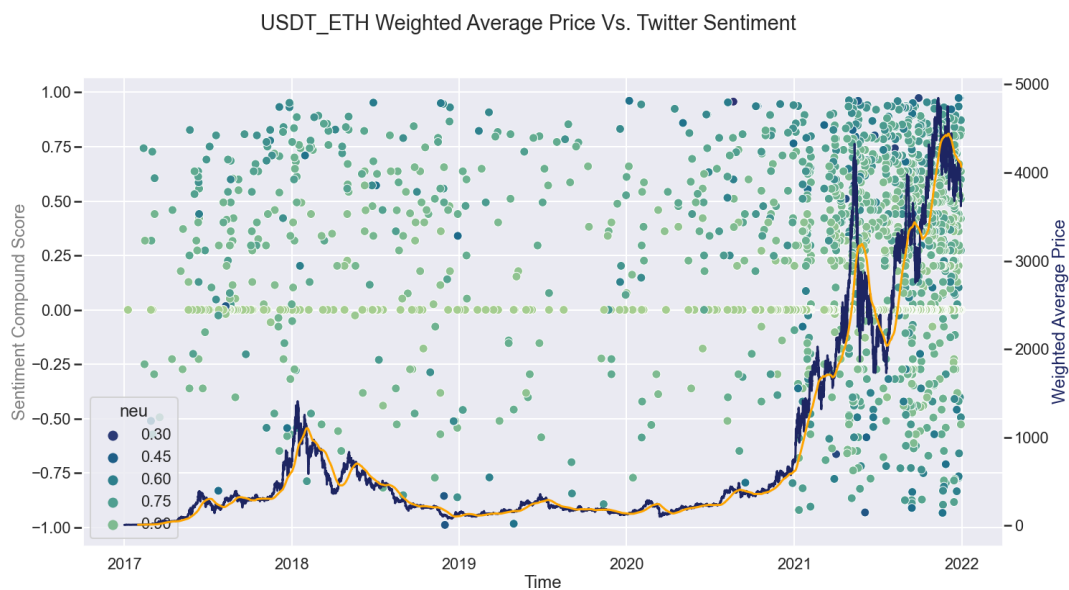
Only tweets with negative compound sentiment scores:



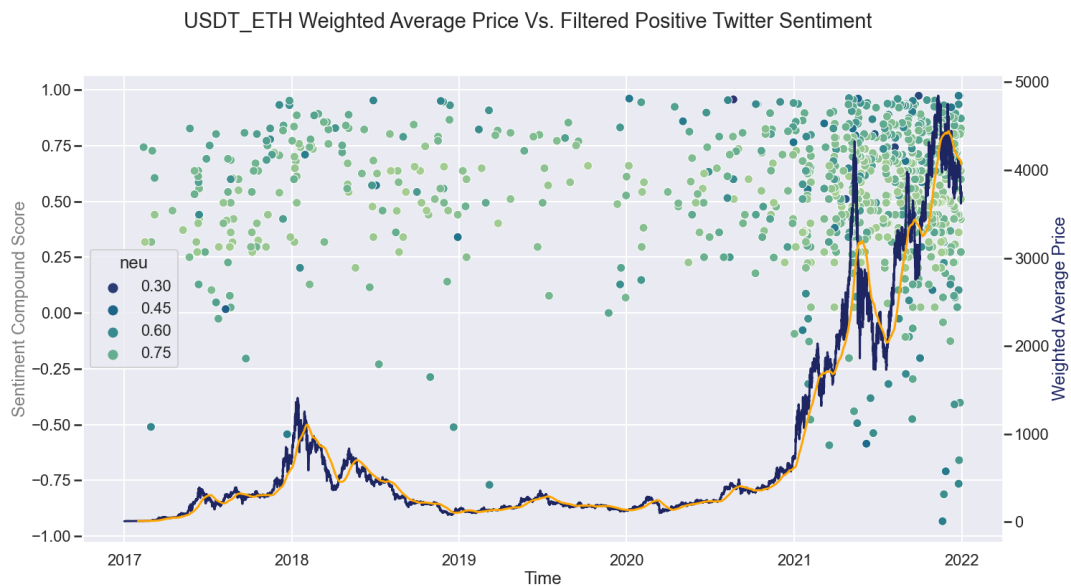
As can be observed, there is a **massive** amount of tweets regarding Bitcoin, so it may perhaps be slightly harder to spot any correlations, with so much noise. There is no observable trend between sentiment and price movement, *however* it is evident that the density of tweets seems to change, which perhaps suggests that the frequency, instead of the sentiment could be linked with price movement, an idea which we will explore later.

Maybe bitcoin is now too popular and well known for the tweet sentiment to be useful. We then move onto looking at ethereum, more specifically **USDT_ETH**:

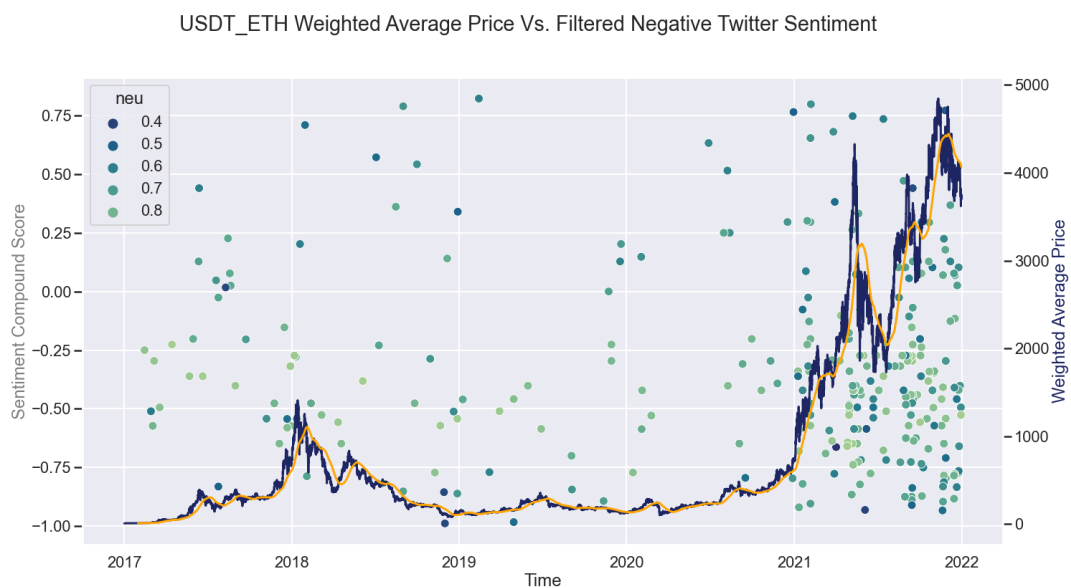
All tweets:



Only tweets with positive compound sentiment scores:



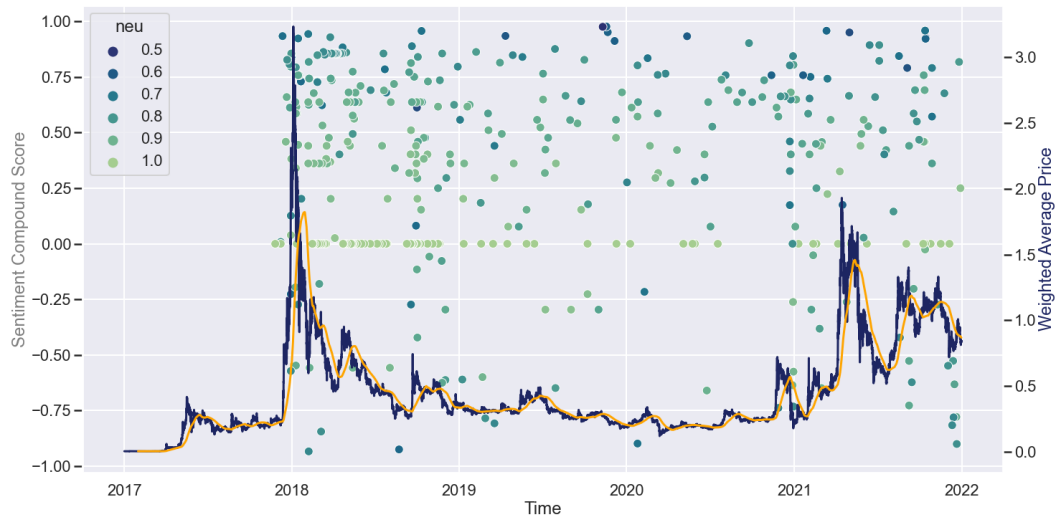
Only tweets with negative compound sentiment scores:



Again, little to no correlation between the sentiment and the price movement although definite changes in density of tweets. Finally we will look at a less popular cryptocurrency, Ripple, more specifically **USDT_XRP**:

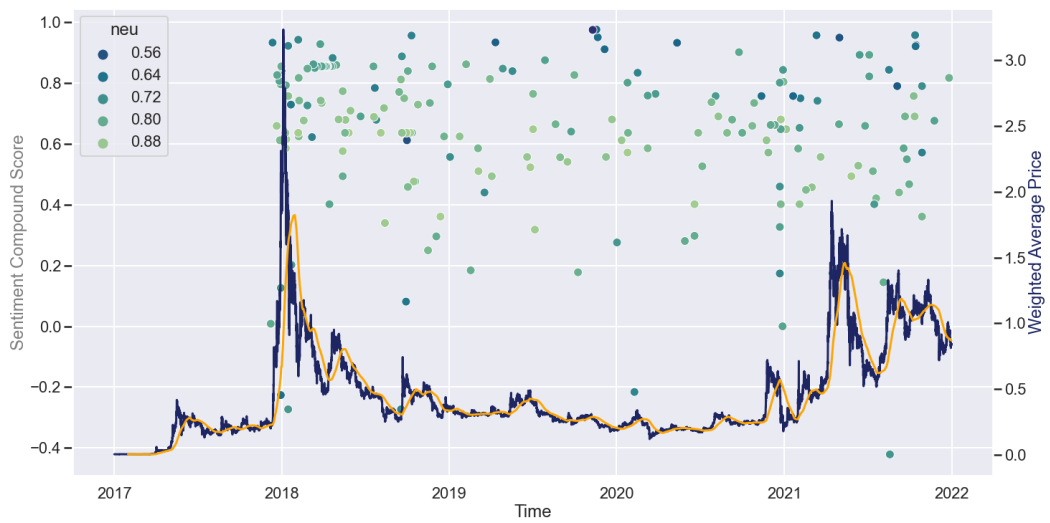
All tweets:

USDT_XRP Weighted Average Price Vs. Twitter Sentiment

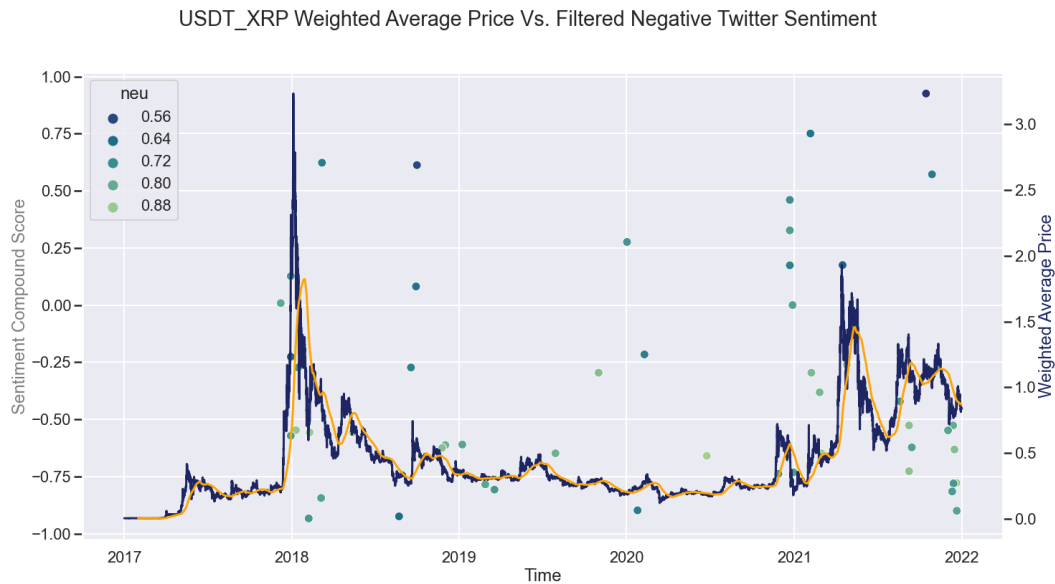


Only tweets with positive compound sentiment scores:

USDT_XRP Weighted Average Price Vs. Filtered Positive Twitter Sentiment



Only tweets with negative compound sentiment scores:



As expected, we have less data for Ripple, since it is a less well known coin. Interestingly there seems to be more of a correlation with Ripple price movement and sentiment than the other coins studied, with an increase in less neutral positive sentiment tweets when large price increases are observed. Although there is really not enough data to make any definite conclusions.

Conclusions

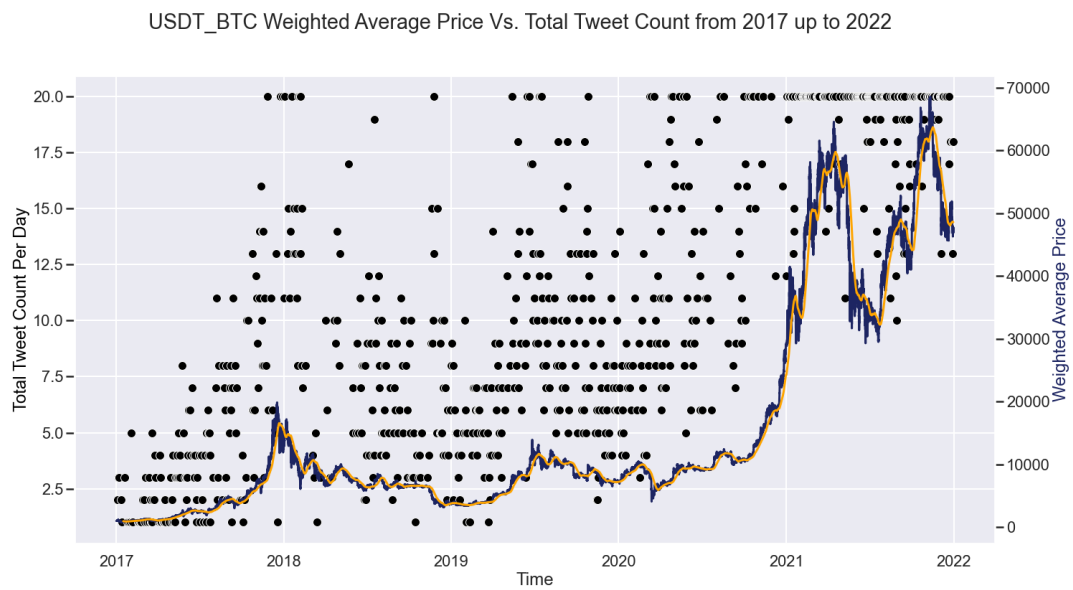
- Bitcoin is too saturated/popular for twitter sentiment to be useful.
- Ethereum also sees little correlation between twitter sentiment and price movement.
- Ripple and perhaps other less popular coins could have correlations with price movement and sentiment however there is a trade off between noise and not enough data.
- This could also be down to the fact that we are using a pre-trained model, which perhaps is not very accurate for this task and maybe most tweets are too neutral in sentiment.
- However there definitely seems to be a correlation between tweet density and price movement...

Tweet Count Per Day Vs. Price Movement

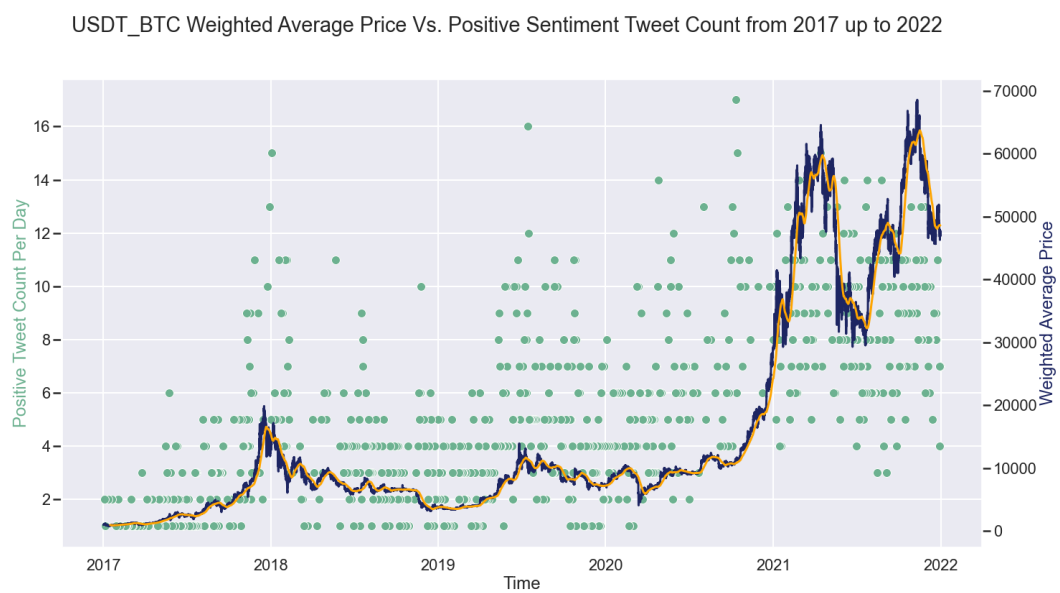
Does the number of relevant tweets per day (which we also refer to as the tweet density) have any correlation with price movement? For each cryptocurrency we will examine the price movement Vs. the total number of tweets per day, the total number of tweets per day with positive sentiment and the total number of tweets per day with negative sentiment.

As before, we start with the **USDT_BTC**:

Total tweet count:

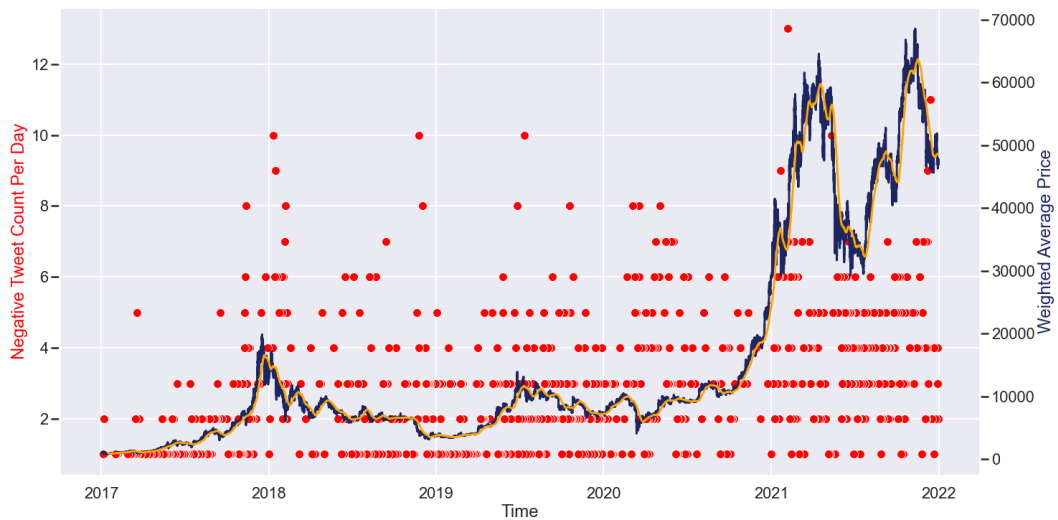


Positive sentiment tweet count:



Negative sentiment tweet count:

USDT_BTC Weighted Average Price Vs. Negative Sentiment Tweet Count from 2017 up to 2022



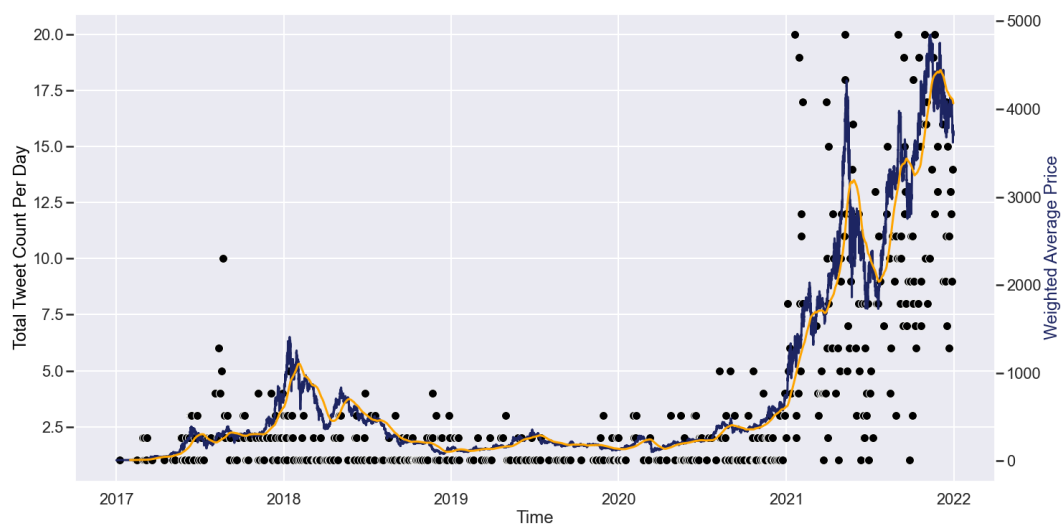
So we can see that there is definite correlation between the tweet count per day and the price movement, with an increase in number of tweets for big price increases. Note that our data is capped at maximum twenty tweets per day, so some days would actually have more tweets than plotted.

When comparing positive Vs. negative sentiment tweet densities we can see that there are always more positive tweets than negative, which perhaps suggests that posters have a bias to favour Bitcoin, since if they're posting about it they likely have an interest. Also interestingly it seems that in the big price dips, even though there are still more positive tweets (probably from the bias) the actual number of positive tweets decreases relative to the previous high.

Moving on, we look at tweet counts for **USDT_ETH**:

Total tweet count:

USDT_ETH Weighted Average Price Vs. Total Tweet Count from 2017 up to 2022



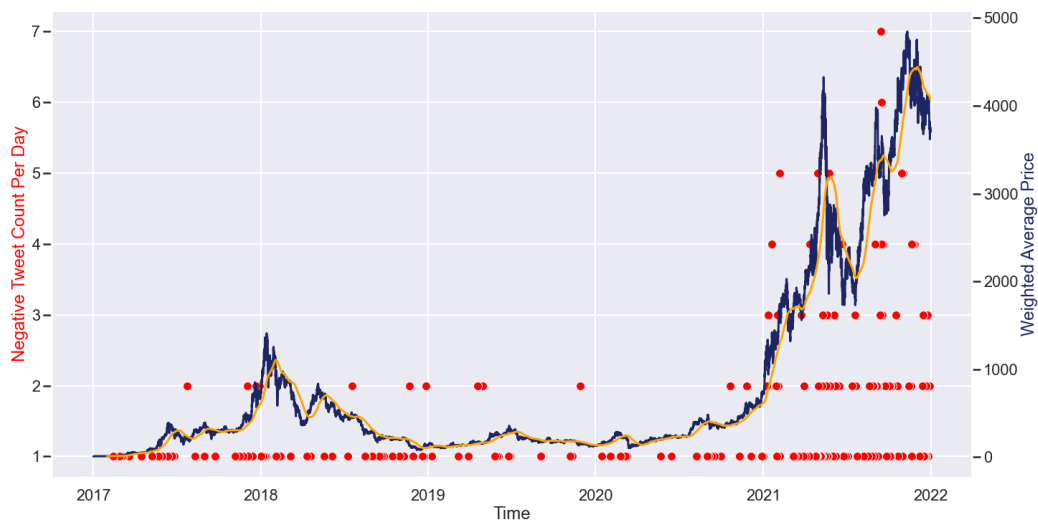
Positive sentiment tweet count:

USDT_ETH Weighted Average Price Vs. Positive Sentiment Tweet Count from 2017 up to 2022



Negative sentiment tweet count:

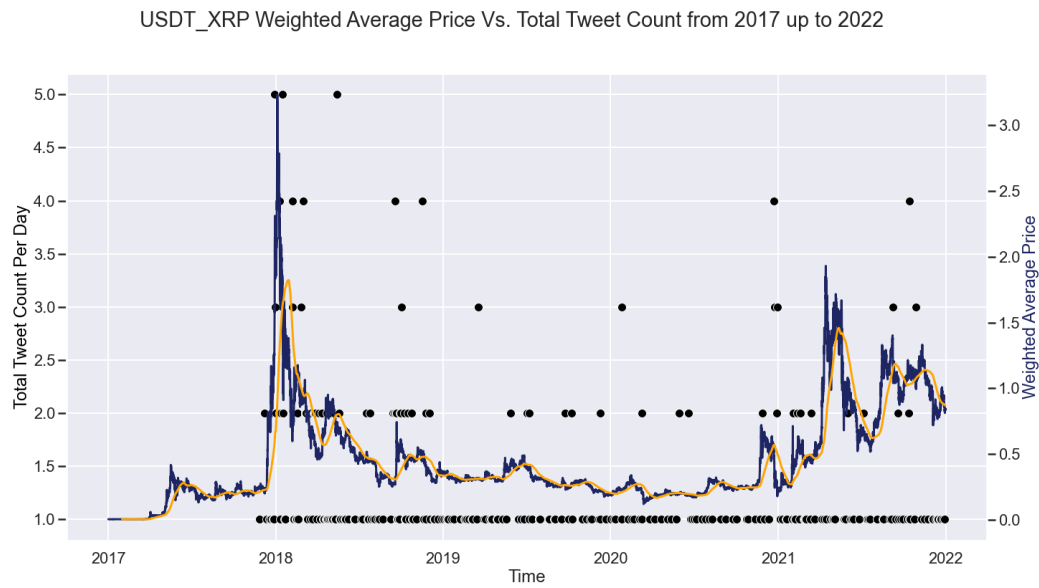
USDT_ETH Weighted Average Price Vs. Negative Sentiment Tweet Count from 2017 up to 2022



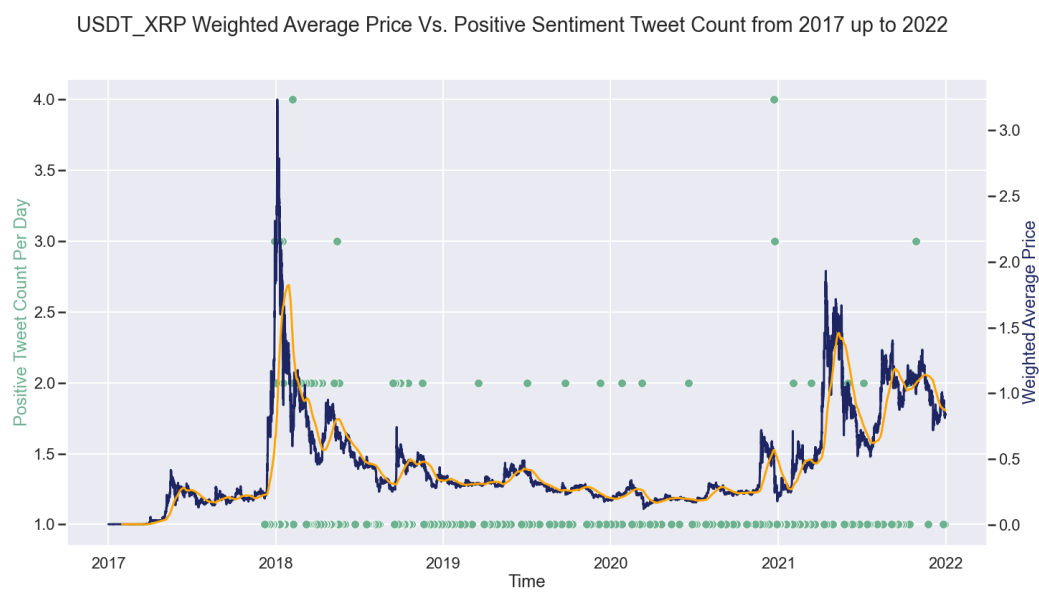
Ethereum seems to be even more correlated than Bitcoin! With a definite increase in tweet density just before big price increases, suggesting that tweet count could perhaps be a good predictor for Ethereum. Perhaps this is because whilst Ethereum is still well known, it is far less well known outside of the cryptocurrency/financial world.

Finally we look at the tweet counts for **USDT_XRP**:

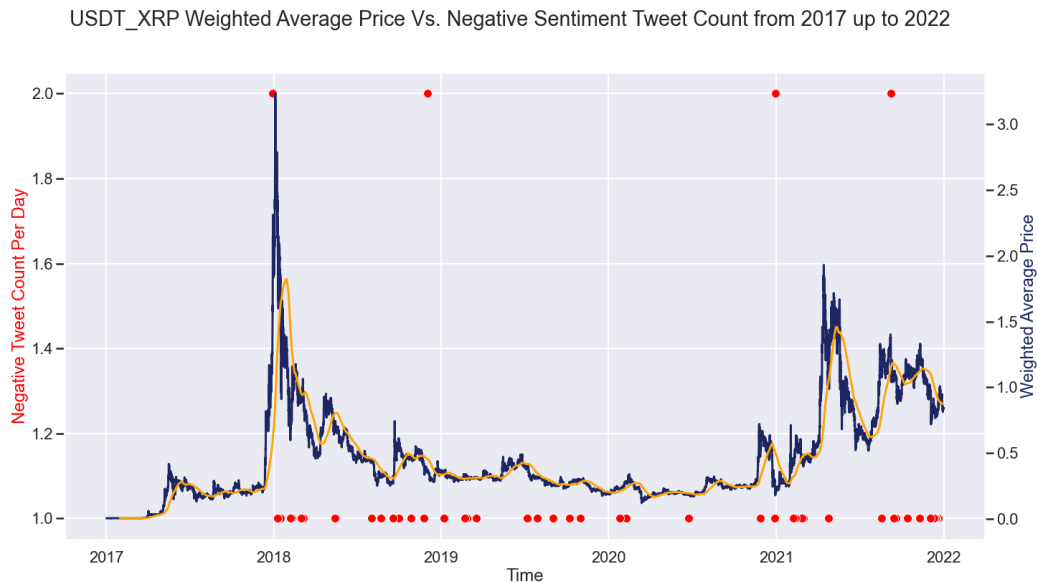
Total tweet count:



Positive sentiment tweet count:



Negative sentiment tweet count:



Again, a definite correlation, however as before there is not really enough data to make any definite conclusions.

Conclusions

- Definite correlation between Bitcoin price movement and the density of tweets.
- Even stronger correlation for Ethereum.
- Not enough data for Ripple.
- (When there is enough data) increases in tweet count are often followed by increases in price, decreases in *positive* tweet counts are followed by price dips and the count of *negative* tweets seems to have little use.

Summary

Now that we have finished our analysis we can answer the initial questions we asked:

Answers

Q1. *Is there a correlation between Twitter sentiment regarding a cryptocurrency and its price movement?*

A1. There seems to be little to no correlation between the actual sentiment of the tweets, however this could be down to the model that we used, and if we had more time, we could perhaps build our own model which is more suited for financial tweets.

However, on further examination we found that the "tweet density" *does* often have a definite correlation with price movement.

Q2. *Does the usefulness of Twitter data depend on the cryptocurrency in question?*

A2. We have found this to definitely be the case, with Bitcoin being perhaps now too popular and mainstream for its tweets to be of use. On the flip side Ripple seems to have correlations however it is hard to get enough data for a less well known cryptocurrency. Ethereum seems to be the sweet spot, with strong correlations **and** plenty of relevant tweets.

So it seems that when using Twitter data, the coin in question needs to be in a "sweet spot" of popularity, where it is not over saturated but still popular enough such that there is enough twitter interest.

1. <https://docs.poloniex.com/#returnchartdata> ↩

2. <https://pypi.org/project/twint/> ↩

3. https://www.nltk.org/_modules/nltk/sentiment/vader.html ↩