#Bitcoin Twitter Sentiment Project: Part 1

Looking at the correlation of bitcoin price and twitter tweets containing #bitcoin

Jonny Morsicato

November 13, 2020

Introduction

The purpose if this research project is to start exploring if there is a potential for using twitter data containing #bitcoin as a variable for indication of positive or negative price runs for financial prediction. Bitcoin is a highly volatile decentralized currency belonging to a new emerging asset class. The high swings in volume and price fluctuations provides excellent potential for short term investors to gain an excellent return on investment. While the long term increasing adoption has been proving to institutional investment that there is potential for bitcoin and other cryptocurrency to act as a hedge against economic downturn. As a disclaimer this is just a thought experiment I am using to broaden my data science skills and this project has been done before.

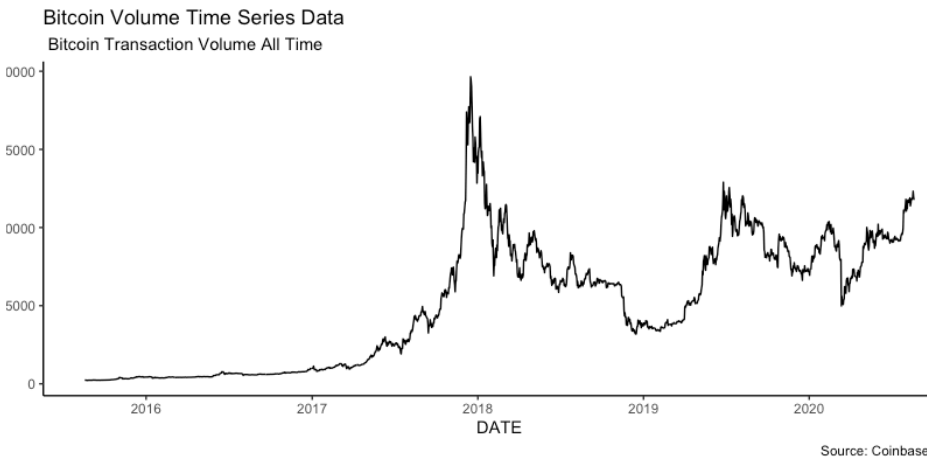

Figure 1: Bitcoin price in USD since the individual or group Satoshi Nakamoto started it in 2009

What is the USD

To first understand what exactly the US dollar is, it is important to develop a historical context for why it superseded over other historical options in the past. In early governments coin currency, coins made directly from a precious metal, were predominant forms of wealth transfer for goods. Coin currency, in many cases, succumbed to seigniorage. Where a government would shave metal from coins in order to earn revenue without their citizens knowing. Government coins also were very cumbersome and not very fluid for large purchases, so bills of sale were

created as a solution. These were essentially a signed agreement from a reputable merchant, the drawee, that was worth X amount of money. A bill of sale could then be passed onto other merchants, all of which understood that when the bill matured, the original drawee would honor what the bill was worth. Of course, often a bill was too large for most transactions so there was severe limitations as to the usefulness of the bill in an economy long term.

While bills of sale were useful they were often too large for day to day use, so merchants came up with bank notes as an intermediary between coin currency and large bills of sale. A bank note essentially was created as a small denomination currency that could be traded right out for coin currency, this allowed the merchant to use large amounts of coin currency to purchase bills of sale with interest accrued on them. Of course in order to avoid paying interest themselves, the paper currency created was said to be openly convertible back to coin currency at any time. So these bank notes were essentially totally convertible, although no merchant would actually be able to back their claim if all bank notes were taken in at once and exchanged for coin currency. Very similar to how a current bank would not be able to surrender all of its stored paper currency without goin bankrupt.

Naturally the government wanted profit from whatever transactions were occurring, so they created representative currency. Essentially the government issued money that could be taken into the federal reserve bank and exchanged directly for precious commodity metals. Since 1971 the US dollar has not been exchangeable for any physical commodity. The current US dollar currency system is considered a fiat currency.

What gives the US dollar Value?

The use dollar is give value based on a simple economic principle, supply and demand. Value = supply/demand. When the US exports a good or service that requires payment in USD, demand is driven. There is obviously a large array of factors that play into what drives demand, but a huge factor is the strength of the US military. Currently the US dollar is valued well against the rest of the world because no military can stand up and implement their currency as the main store of value for foreign transactions. Currently the US accounts for 38 percent of spending focused on military worldwide and so as long as we maintain firepower superiority, the US dollar will maintain its value.

BTC as a currency

As an individual it is essentially impossible to pay for a good without using a third party intermediary or through direct trade. Bitcoin was built to present a possible solution to this issue, although as with many things, the first may not always be the best. Bitcoin does provide a peer-to-peer payment system and a store of value that has slowly been displaying its staying power in the world economy. For this project I pulled the idea of what gives value to the USD on an economic level and decided to compare it to the price of BTC. The final hope of this is that there could be a leading correlation of either tweet volume or positive sentiment and the increase of bitcoin price. Potentially proving use of twitter sentiment and volume for use in trading crypto currency.

Discussion

Procedure

- Obtain time series twitter data centered on #bitcoin
- Obtain time series BTC volume and price data for same time
- Clean twitter data. Remove websites, numerics, emoji, stop words, punct

- Create daily tweet volume data.frame
- Find top 15 common words in tweets
- Categorize twitter data into positive/ negative
- Look for data correlation between btc price and tweet volume
- If correlation make procedure for LSTM project with longer time series

Cleaning twitter data

1	2018-07-11 19:35:15.736769	b'@Miss_rinola But you\xe2\x80\x99ve heard about …	But you ve heard about BITCOIN
2	2018-07-11 19:35:15.744769	b'Duplicate skilled traders automatically with Bitcoin! …	Duplicate skilled traders automatically with Bitcoin …
3	2018-07-11 19:35:15.867339	b'Project Manager - Technical - FinTech - Central Lon…	Project Manager Technical FinTech Central Lond…
4	2018-07-11 19:35:16.021448	b'Coinbase App Downloads Drop, Crypto Hype Fading…	Coinbase App Downloads Drop Crypto Hype Fading …
5	2018-07-11 19:35:16.687237	b'If you don\xe2\x80\x99t understand Bitcoin, you d…	If you don t understand Bitcoin you don t underst…
6	2018-07-11 19:35:17.021966	b'August 10th is an important day. #Bitcoin #ETF $BT…	August th is an important day Bitcoin ETF BTC
7	2018-07-11 19:35:17.885198	b'With No One Price Law for Bitcoin, Japan\xe2\x80\x…	With No One Price Law for Bitcoin Japan s FSA n
8	2018-07-11 19:35:18.274546	b'Crypto Theft Rising and May Reach $1.5 Billion This…	Crypto Theft Rising and May Reach Billion This Y…

Column 3 original twitter data, Column 4 cleaned twitter data

Cleaning the twitter data for the initial sentiment analysis proved to be a little difficult. The first hurdle I ran into was to remove all of the emojis from the data. When I got the data that was pulled from the twitter API all of the emojis were converted into long UTF-8 strings. Once I figured out the pattern and better understood the UTF-8 values I was able to build a regex expression and remove the values using a gsub function. I also used regex to remove things like punctuation, hashtags, retweets, and website links.

The next hurdle was to remove stop words from the twitter data. Stop words are words that don't really add much meaning to a sentence. These words can include filler speech such as; in, the, you, we, etc. Stop words can also be variation of the same word like; love, loved, loving, lover, etc. These type of words take up computation time and are necessary to be removed to keep efficiency up. I also

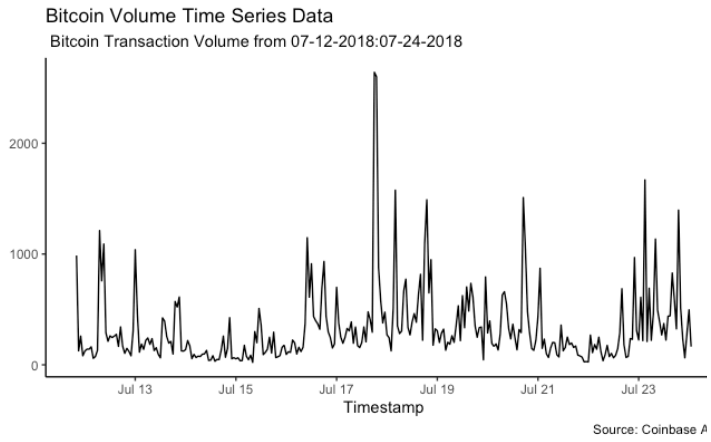

Figure 2

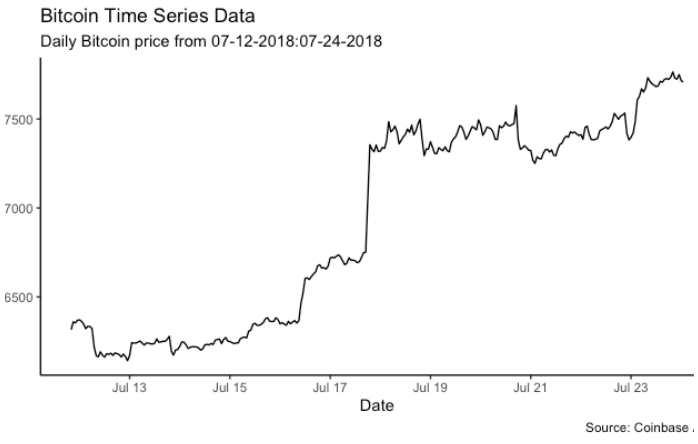

Figure 3

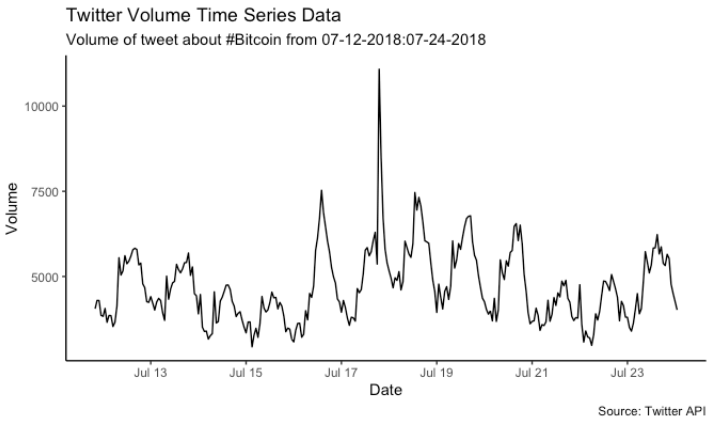

Figure 4

don't need to know that one of the top 10 words used about #bitcoin is, "the".

Comparing Time Series Data

After cleaning the twitter data I wanted to make a general comparison about the time series data from 7-12-2018 to 7-24-2018. Looking at the graphs between the BTC volume and the BTC price in USD, I was curious about if the twitter volume would show the same increase in volume as seen so clearly on 7-18-2018. Using a count function and converting the data into a POSIXct data type, I was able to build a data frame containing the volume of each tweet per day. I continued to plot the volume data in the same manner as I had plotted the BTC volume and price data, revealing the same uptick in volume on 7-18-2018. This shows at least the beginning of proof of concept for more exploration.

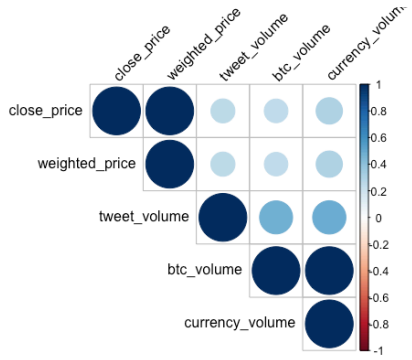

Figure 4: Correlation heat-map

Twitter Volume Price Correlation

After inspecting the time series data of the BTC Volume, BTC USD price, and the Twitter Volume. I used a basic correlation function to see if the relationship between the the variables was causal. By looking at the heat map correlation plot, figure 5, there is a correlation coefficient of r = 0.6 or higher between tweet volume and the etc volume. There is only a correlation of around r = 0.2 between BTC tweet volume and BTC price. This does still imply that there is a positive correlation between that could have potential for financial modeling, but not as strong as expected when comparing to the time series data visually.

Twitter Sentiment Graphs

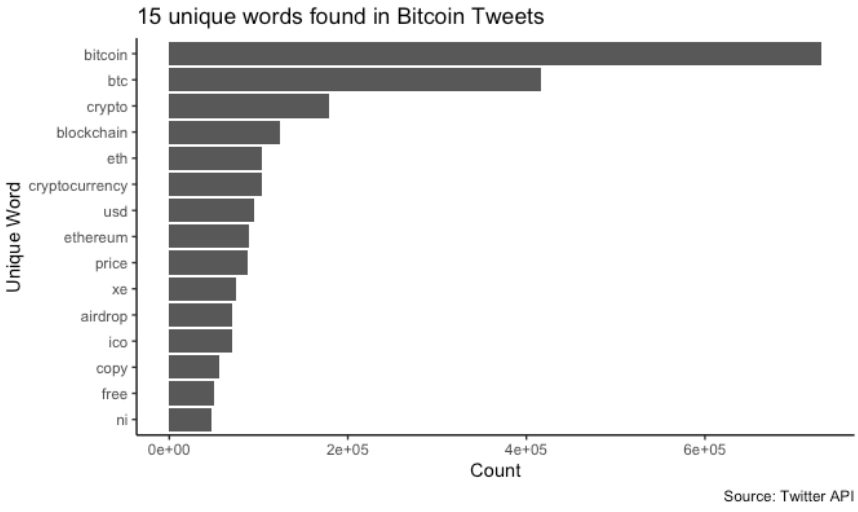

Figure 5: Top 15 most frequently used in twitter data

After looking at possibly correlation between the BTC data and twitter data, some basic sentiment analysis was done. In figure 5, the top 15 most common words in the twitter data was found. The “ni” word at

the bottom revealed an issue that must have shown up somewhere while cleaning the twitter data.

Continuing on, the top 20 positive and negative words were found. This can be seen in Figure 6. This also revealed that there was clearly more positive sentiment centered on Bitcoin. There is also a bit of ambiguity in the words used to define

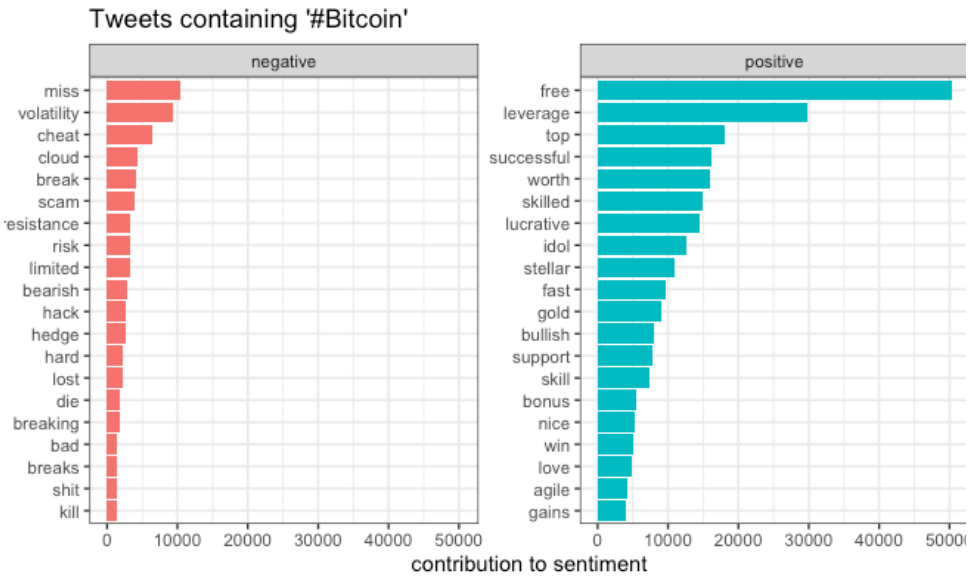

Figure 6: Top 20 positive and negative sentiment words used in twitter data

positive and negative out of context. For example the phrase, “This week has killed it on bitcoin gains”, Is positive, but the sentiment analysis from the tidyverse package will categorize the words as negative. For a better sentiment analysis a LSTM model could be used that worked to understand the sentiment of an entire sentence.

Sources:

https://bitcoin.org/bitcoin.pdf

https://files.stlouisfed.org/files/htdocs/publications/review/91/09/Currency_Sep_Oct1991.pdf