Lorenzo Gordon

IST-652 Data Scripting

Project Report

**Bitcoin Tweets Report**

Bitcoin and cryptocurrency have been major buzz words in the world over the last few years. As a form of payment that is outside the control of any one person, group, or entity it has been touted as a potential currency of the future and for the people. Twitter is a one of the leaders in social media platforms allowing millions of people to connect via 140 text blobs. In 2021 Bitcoin news and thoughts were all over twitter and I wanted to investigate the tweets relevant to bitcoin. The goal of this research is to see if we are view the overall sentiment of bitcoin in 2021 on sentiment. Also I would like to look if the price action matches the sentiment.

To do this analysis I originally attempted to scrape twitter via snscrape but ran into numerous issue due to my pc processing power. I was able to find a dataset via Kaggle ([Link](https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets)) that contains tweets with #bitcoin or #btc. The number of records in this dataset is 4,132,335 records with 13 columns. Also in addition to the Kaggle dataset I downloaded bitcoin price data from yahoo finance. To get the Data in the correct format for my analysis there were a few steps that I needed to complete. I took the following steps to get the data into the right condition to start my analysis:

1. Read in the CSV file into a pandas dataframe
2. Convert date into a datetime field
   * Ran into an error where there were “Btc” or “ETH” located in the date columns had to remove these rows. (27 Rows)
3. Checked for any tweets with an NA text
4. Filter the data using datetime for the year 2021
5. Dropped columns (source, user\_created, user\_favourites,user\_name(not @ name))
6. Took a sample of 5% of the records to start the sentiment analysis (about 100,000 observations

After my data was in the correct form I ran a sentiment analysis on each tweet. I started my creating an empty list called TweetRating, then assign the SentimentIntensityAnalyzer to a variable. Next I made a for loop that would go from 0- # of tweets in the dataframe and get the compound score. Finally it will append these scores into the tweetrating list. This gives me access to the compound score and the tweet text and can further join that information with different dataframes. The reasons why I choose compound score is because it considers the neutral, negative, and positive polarity scores and normalizes it in a -1 to 1 scale. 1 represents extreme positive, -1 represents extreme negative.

When going into this analysis the goals are to the overall sentiment throughout the 2021 on a month-to-month basis, what types of users are tweeting about bitcoin, and Price activity on bitcoin. The first graphic that I looked at was the average follower amount based on sentiment direction.

Chart, bar chart

Description automatically generatedText

Description automatically generated

The 0 in this graph represent a neutral tweet. When looking at the follower account the users with the highest number of followers seem to have more neutral sentiments on bitcoin while positive sentiments are last in average follower amount. Meaning that overall positive sentiments were coming from accounts with a smaller followings and neutral tweets came from a bigger following.

Looking further into the types of users and the sentiment of their tweets I investigated verified users.

Chart, bar chart

Description automatically generatedTable

Description automatically generated

In general, most of bitcoin related tweets came from unverified accounts, and the highest amount of tweet activity came from unverified users with positive sentiments. It is interesting seeing that the largest amount of bitcoin related tweets were positive tweets from unverified accounts. Also looking back at the previous followers count graph positive sentiment tweets on average had a smaller following. I would be interested to see what type of accounts these are. If they are real accounts that just have a small following or small group pages that promote bitcoin.

Next I looked at the volume of tweets by month in 2021. Looking at the overall bitcoin related tweets by month we can we that there is a big spike that happed towards the second half of the year. Also, there is a relative spike from march to April

Chart, bar chart, histogram

Description automatically generated

In addition to looking at the tweet volume by month I investigated the overall sentiment by month.

Chart, bar chart

Description automatically generatedTable

Description automatically generated

Looking at the overall bitcoin related tweets by month we can we that there is a big spike that happed towards the second half of the year. Also, there is a relative spike from march to April after the fall. In general these sentiment score seems to follow the tweet volume that we saw in tweets by month. The second half of the year is where bitcoin tweets picked up in popularity and positive sentiment.

Finally I investigated the bitcoin highs by month to see if we are getting new price highs when positive sentiment was increasing. We can see that bitcoin price similarly to the sentiment started to raise in the second half of the year (Jul🡪Dec). Bitcoin hit a year high in oct of $66,930

Chart, line chart

Description automatically generated

Some key takeaways that I found in this analysis are:

1. The overall sentiment in 2021 for bitcoin was positive and this can be seen in price highs throughout 2021. To be more specific the second half of 2021 is where we see the highest level of consistent positivity.
2. Positive tweets tend to come from accounts with smaller followings relative to neutral and negative tweets. In addition to this the largest volume user type tweeting about bitcoin were positive unverified account. (Also the most verified tweets were positive)
   1. In the further I would love to look into the specific accounts and tweets that made up these positive tweets for both verified and unverified