ANALYSIS OF CRYPTOCURRENCIES USING HASHTAG CRYPTOCURRENCIES (#cryptocurrencies) between 2020-02-19 02:58:09 UTC and 2020-02-21 04:50:06 UTC

Presented to:

EKRAAL Innovation Hub

Nairobi, Kenya

Presented by:

Clinton Moshe Mwachia

Data Scientist

February 21st, 2020

Table of Contents

EXECUTIVE SUMMARY

OVERVIEW

This report is aimed to help the people who would love to get into cryptocurrency but have fears about it. The data used in this analysis is from twitter and the tool used for analysis is RStudio.

It covers data collection, data cleaning, visual analysis and sentimental analysis of the various users of cryptocurrency. The sentiments are analyzed from the users of cryptocurrencies and are willing to share their experience with cryptocurrencies by tweeting.

The methodology used in this analysis is both quantitative and qualitative.

THE PROBLEM

The local finance system is centralized hence being controlled by some individuals, companies and the government and hence the users have no control over the money system.

The best alternative is to use cryptocurrencies but with all the negatives that comes with it, people fear using it. Since most people do not have enough knowledge and information for them to be able to make decisions, this justifies this analysis and clears all the doubts.

THE SOLUTION

To solve this problem, I used a sample of tweets of cryptocurrency users to do sentiment analysis of the users to get their experience and facts about cryptocurrencies.

All the data collection, data analysis, data visualization, data cleaning and sentimental analysis are discussed below.

1.0 INTRODUCTION

This report is an overview of the analysis of 2000 tweets based on the hashtag cryptocurrencies, #cryptocurrencies, that was done on February 21, 2020 for tweets tweeted between 2020-02-19 02:58:09 UTC and 2020-02-21 04:50:06 UTC.

Over the years, there has been a great advancement in the field of finance and security, with technology being the key motivator. The normal money system has been described as being manipulated by the government, insecure, limited to country boundaries and limited transactions.

Due to this, quite a lot of people adopted cryptocurrencies since it's described as being highly secured, decentralized, very fast and transparent but this has not been the case for every cryptocurrency user.

Despite Cryptocurrency being advantaged as compared to the local currencies, many people still have fears of using it as a medium of exchange due to its high volatility, irreversible payment and inadequate knowledge about it.

This analysis is aimed to use data science techniques in order to help people make decisions about cryptocurrencies. It focuses on using sentimental analysis to study the opinion of cryptocurrency users in 2020.

The objective of this analysis is to: -

- Explore the satisfaction of cryptocurrencies to the users in 2020.
- Analyze from which region do most users come from based on their tweets.

To achieve the above objectives, data science plays a major role to help In this problem. This study will help people in decision making on cryptocurrencies in 2020.

2.0 METHODOLOGY

This report presents the findings obtained using both qualitative and quantitative analysis that was done on the tweets. Qualitative analysis done included sentimental analysis for understanding the feelings, opinions and perceptions of the cryptocurrency's users.

Quantitative analysis done included drawing tables, calculating percentages and proportions and graphing.

2.1 DATA COLLECTION

The dataset used was obtained from twitter. With the help of the twitter API, I scraped the data using RStudio and the rtweet package from the twitter website. The data has a total of 90 column variables and 1753 observations.

I downloaded the data into RStudio using the search_tweet () function with the query #cryptocurrency. Out of the 90 columns, I had an interest with only three columns variables namely: -

- Created_at: the time the tweet was sent
- Screen_name: user name of the user
- Text: tweet text

2.2 DATA ANALYSIS

I used the R software with RStudio IDE for my data analysis. After loading the data into RStudio, I did quantitative analytics such as calculating the number of tweets, viewing the first 5 records of the tweets, checking the start and stop time of the tweets, getting the number of users tweeting about cryptocurrency, their location and user names.

I then plotted the users as seen in figure 1.0 and created a new data frame which consists of my desired column names.

The new data set has no NA values. I then cleaned the data by removing links, stopwords and filtered it to remove retweets. After filtering the retweets, I tokenized the data to obtain tokens of the tweets in a single column with the output of column header being words, refer to table 1.0.

I then plotted the word counts to obtain the most frequent words in cryptocurrencies, refer to figure 2.0

I then did the word paired analysis to know which of the words are used together most frequently, I presented the results in a table, refer to table 2.0.

I used bing lexicon for sentiment analysis. It categorizes words into either negative or positive categories, table 3.0. I used an inner joint to join the lexicons to my data and they outputted a table, table 4, and then plotted the data as seen in figure 3.0. the table by proportions for the sentiments is represented by table 5.

3.0 RESULTS

From the analysis, the results were reported in tables, graphs and figures. Using figure 1 below, we can note that most of the people who use cryptocurrencies and tweet about it come from the United states, London, New York and the majority have not stated their location.

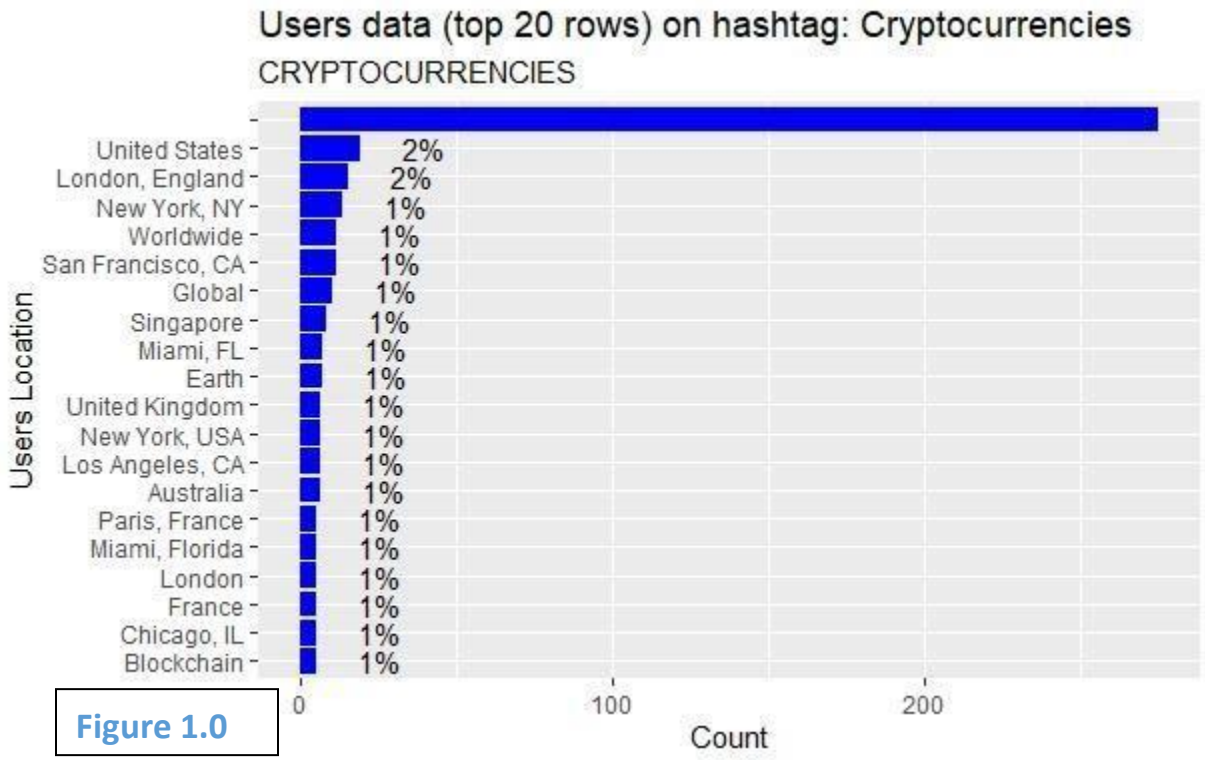

Figure 1.0

After tokenization, removing stop words and filtering out retweets, we remained with a data frame of length of 3886 words. The table below shows the top 10 tokenized words.

Table 1.0: TOKENIZED WORDS

Id	Word
1	Testing
2	Cryptotab
3	Browser
4	Day
5	2
6	Easy
7	Bitcoin
8	Mining

9	Cryptotab
10	Browser

Cryptocurrencies, bitcoin, blockchain, crypto and btc are in the top list of the most used words in cryptocurrencies. The plot below summarizes the word count of the words.

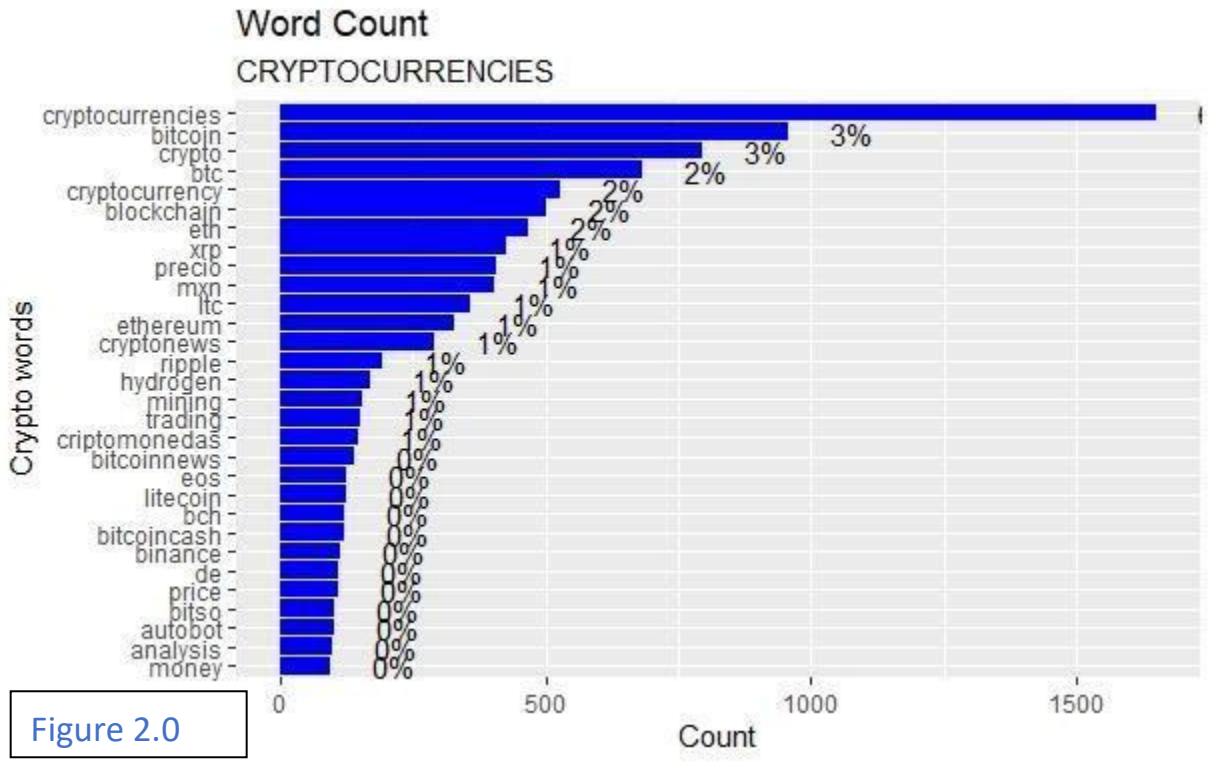

Figure 2.0

The table below summarizes the paired words in cryptocurrency

TABLE 2.0: PAIRED WORDS			
Id	**Word 1**	**Word 2**	**count**
1	Bitcoin	cryptocurrency	29
2	Cryptocurrency	crypto	23
3	Cryptonews	cryptocurrencies	22
4	Cryptocurrencies	cryptonews	21
5	Blockchain	bitcoin	19
6	Crypto	cryptocurrencies	18
7	Bitcoin	cryptocurrencies	16
8	Bitcoinnews	coinbase	16
9	Blockchaintechnology	bitcoinnews	16
10	Coinbase	eth	109

table 3.0 below summarizes the bing sentiment used in the sentimental analysis. It has only two categories, negative and positive.

Table 3.0

Id	word	sentiment
1	2-faces	negative
2	abnormal	negative
3	abolish	negative

About 65% of the words used in tweeting about cryptocurrencies are positive and 35% are negative. Some of the positive words include top, protection, bullish and the negative words include fraud and scam. Table 4.0 below gives a summary of the sentiment of the words used in cryptocurrencies tweets.

Table 4.0

Word	sentiment	count (n)
Fraud	negative	14
Top	positive	8
Boost	positive	6
protection	positive	6
Amazing	positive	5
Support	positive	4
intelligence	positive	4
Scams	negative	4
Bullish	positive	3

Most of the people who have been tweeting on February between 19th and 21st are victims of fraud incidences in cryptocurrencies.

the sentiments above can be presented in a plot as shown in figure 3 below.

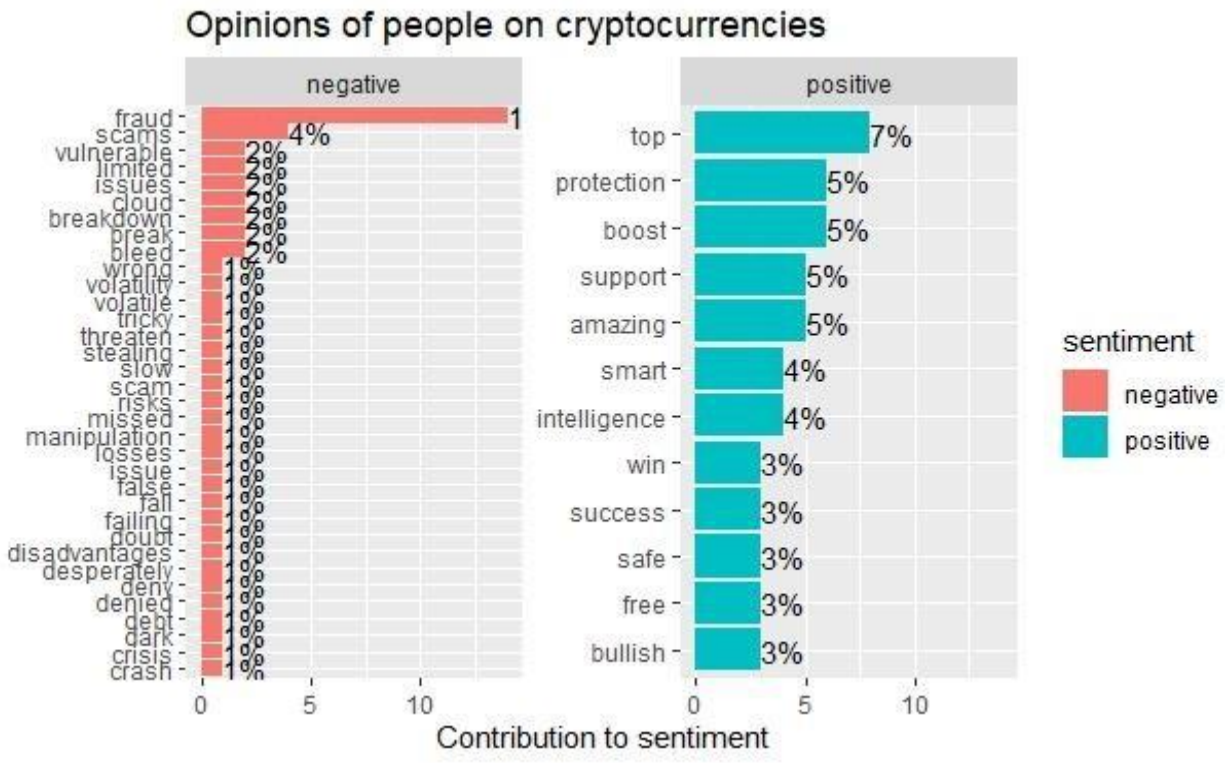

Figure 3.0

Table 5.0

Negative	Positive
35%	65%

4.0 DISCUSSION AND CONCLUSION

From the results of the analysis, the rate of frauds in 2020 is leading because most of the users are victims of fraud.

Based on the statistics, averagely speaking, cryptocurrencies still take the lead with an amazing 65% positive feedback from the users. Individually, there are more positive than negative feedback from users of cryptocurrencies.

For instance, using cryptocurrencies comes with a lot of benefits such as its highly secure, good support, it's a smart way to trade and you are also considered intelligent when you use it, as it is summarized in figure 3.0.

With all the benefits, we also have to consider the fact that it has quite a lot of demerits such as fraud, scams, limited and its volatility (refer to figure 3.0). In a normal economic system, everything comes in with its merits and demerits.

In conclusion, cryptocurrencies are a good medium of exchange since the positive responses outweighs the negative responses. The people intending to get into cryptocurrencies should make sure they meet certified cryptocurrency experts to avoid being frauded or scammed.