**Text Analytics and Natural Language Processing – DAT-5317**

**Et bilde som inneholder tekst

Automatisk generert beskrivelse**

**A3: Business Insights Report**

**Bitcoin @ Twitter: Marketing “BitCover” ETF**

**By**

**Marius Heje Maehle**

**05. December 2021**

**Bitcoin @ Twitter: Marketing BitCover ETF**

**Business case:** A new Bitcoin Exchanged Traded Fund (ETF) called “BitCover” want to market their fund on Twitter. Which words and sentiments should they focus on in their marketing on Twitter to acquire customers?

**Bitcoin and Twitter**

Twitter as a Social Media platform

As twitter has been one of the most used social media platforms the last years (Most popular social media networks, 2021) it can provide a lot of data and value to consumer. As twitter provides a lot of text it can be used to find valuable insights, but also there can be a lot of noise in the data. All the data that is retrieved and used in the analysis is from Twitter’s database.

Cryptocurrencies and Bitcoin

Since the pandemic there has been a lot of new investors coming into the market due to a post-pandemic boom (Reinicke, 2021). I will not go into depth why this is, but it is important to acknowledge as the stock market and cryptocurrencies has been booming since the crash in the spring 2020 (Reinicke 2021).

As seen on the chart below the total cryptocurrency market cap has overall increased a lot towards almost a market cap now around 2,5 trillion dollars (Global Cryptocurrency Charts, 2021). The cryptocurrency has a history of being volatile, but the market size has increased exponentially since 2020 (Global Cryptocurrency Charts, 2021).

Chart, line chart

Description automatically generated

*Chart 1: Total cryptocurrency Market Cap. Retreived from: Global Cryptocurrency Charts, 2021*

BitCover ETF

As there are more investors in the market, there is a lot of hype and interest around cryptocurrencies which has created a new wave of crypto traders as seen on the Global Cryptocurrency Chart (2021). A bitcoin ETF mimics the price of the digital currency to allow investor to buy into the ETF without trading bitcoin itself (Reiff, 2021). This can make it easier for investors to buy and own bitcoin on the market. To separate the signal from the noise in all the data on twitter these days can be difficult, and BitCover wants to se be signal in the data. As BitCover is a newly started ETF, they want to see how they should approach their words association and word sentiments in their marketing to acquire and get more customers.

The business problem statement is as mentioned: *“Which words and sentiments should they focus on in their marketing on Twitter to acquire customers?”*

**Analysis of the Twitter data**

The data analysis will focus on popular words, sentiments, positive and negative words, and word connections to give some valuable output from the data to provide useful business decision for the twitter marketing.

Sentiment analysis

As sentiment analysis provide a good way to understand the attitude and opinions expressed in the text it can give valuable insight on which word has positive sentiments regarding bitcoin (Silge & Robinson, 2017). NRC is the most important sentiment library as it offers a lot of different flavors of sentiments and gives the best structure to provide valuable business insights. As seen on the table below the NRC method is giving a positive sentiment, while the Bing and Afinn is both giving negative sentiments values.

Chart

Description automatically generated*Chart 2: Sentiment methods*

Most common positive and negative words

When looking at the most positive and negative words in the data it is important so which can be leveraged in a positive way in the marketing, and which words that should be avoided.

*Positive words*: As we clearly see that the positive words are strongly connected to feelings with **super**, **support**, **strong**, **love** and **happy**. These words should be used on how the marketing should be framed. The positive words such as **free**, **super**, **superb**, **bullish**, and **top** can be used more in the way to describe the investment strategy and the meaning of bitcoin. As bitcoin is seen by some as a hedge against other investments, I’m assuming that **gold** is used in the text to demonstrate the investment strategy of investing in bitcoin, but this is only an assumption.

*Negative words*: All the most common negative words are probably used when the people on twitter use bitcoin as a negative investment and currency. **Resistance** is maybe the only negative word which can be seen as positive in a way that bitcoin is seen as an investment. Still, it’s good to know that the word can be seen as negative and should therefore be avoided in usage in the BitCover marketing.

Chart, funnel chart

Description automatically generated

*Chart 2.1: Most common positive and negative words*

N-grams

By using the n-grams we can see which words tend to appear after others and how the relationship between the words is (Silge & Robinson, 2017). The arrow on the bigrams shows us the arrow to end before touching the node on which word follows which. We clearly see that bitcoin is connected to many words on the bigrams, but it is important see which word that can be useful to promote BitCover’s ETF and filter out the noise. Words such as **crypto**, **cryptocurrencies**, **blockchain**, **price**, and **pushes** can be useful words to know about. Most of them makes sense without the bigrams, but it’s important to highlight the words people connect in pairs in connection to Bitcoin. Connections such as bitcoin 🡪 **btc** 🡪 **transactions** highlight the important of using bitcoin as a transaction. **Bitcoin 🡪 crypto 🡪 gamify and bitcoin 🡪 nft 🡪 gamefi** on how people connect bitcoin to gamification and NFTs, and how this can be leveraged long-term. Lastly, **bitcoin 🡪 usd 🡪 prices** on how bitcoin is compared to the United States Dollar and price range.

A picture containing text, indoor, counter

Description automatically generated

*Chart 3: Common bigrams in the bitcoin twitter data*

Pitfalls

It’s important to mention that the analysis should be carefully used because correlation does not always imply causation. The sample data from twitter gives a clue on what words which is popular, which ones are used together and which gives the different sentiments. If we want to investigate it further we should take more sample data, to find more connections and which ones that have a strong correlation between each other.

**Business decision**

Based on the facts from the background, sentiment analysis and n-grams we can see that there are some words and sentiments that should be focused on in the marketing of BitCover’s ETF. Keep it simple, and straight forward with playing on the sentiments of **support**, **strong**, **love** and **happy** to trigger the feelings of the buyer. In the more concrete wording in relation to bitcoin strategy use words such as **super**, **bullish**, and **top**. To provide more depth in relation to the word we found in the n-grams by connecting it to **crypto** and **blockchain**. More concrete BitCover can market their ETF that it is easy to focus on **price** in relation to the **USD**, and the easier way of doing **transaction** with bitcoin and BitCover’s ETF.

As these insights are straight forward and not so in depth, BitCover should further look deeper into the words and correlation between the words to get a better understanding which words helps create customers buying their Bitcoin ETF. By monitoring the change in words used over time can also be useful to change with the tides, such as the use of **NFTs** and **gamify**.

**Appendix:**

Global Cryptocurrency Charts (05.12.2021). *CoinMarketCap*. Retrieved from: <https://coinmarketcap.com/charts/>

Most popular social media networks (updated for 2021) – digital marketing’s most powerful tool. (26.08.2021). *Revive.digital.* Retrieved from: <https://revive.digital/blog/most-popular-social-media/>

Reiff, N. (19.11.2021). Bitcoin ETFs Explained. *Investorpedia*. Retrieved from: <https://www.investopedia.com/investing/bitcoin-etfs-explained/>

Reinicke, C. (02.08.2021). New investors are jumping into the market during the post-pandemic boom. *CNBC*. Retrieved from: <https://www.cnbc.com/2021/08/02/new-investors-are-jumping-into-the-market-in-the-post-pandemic-boom-.html>

Silge, J. & Robinson, D. (2017). Text Mining with R: A Tidy Approach. *O’Reilly Media*

**R Code and R output:**

######################################################

######A3: Business Insights Report - Marius Heje Mæhle #######

###############Querying Twitter for bitcoins################

#install the necessary packages

#install.packages("twitteR")

#install.packages("tm")

#install.packages("NLP")

#install.packages("rtweet")

library(rtweet)

library(twitteR)

library(tm)

library(NLP)

#consumerKey, consumerSecret, access token and access secret from my twitter

consumer\_key <- 'osmot3QLGT5ithOP2A3CgHGHh'

consumer\_secret <- '2hP8tl70oJHeBLW0AZOaNfbZQj4hc51HeBFfkO310IV6CJleO6'

access\_token <- '1204793374638837761-HXz9ePnCBLE93FJFk2mdet0rxQj2wc'

access\_secret <- 'DpTIPGkY0H9PGHFVW3GSSrZYkOdC3oV3pBuPNyNVoxVBz'

name\_of\_app <- "Incerto Coding"

twitter\_token <- create\_token(

app = name\_of\_app,

consumer\_key = consumer\_key,

consumer\_secret = consumer\_secret,

access\_token = access\_token,

access\_secret = access\_secret)

rt <- search\_tweets("#Bitcoin", n = 10000, lang = "en", include\_rts = FALSE)

#I receive a Warning messages:

#1:In class(object) <- "environment" :Setting class(x) to "environment" sets attribute to NULL; result will no longer be an S4 object

#put dataset into a dataframe

library(dplyr)

mydf <- data.frame(line=1:9037, text=rt$text)

print(mydf)

token\_list <- mydf %>%

unnest\_tokens(word, text) %>%

count(word, sort=TRUE)

print(token\_list)

Text

Description automatically generated

#remove stopwords

library(stringr)

data(stop\_words)

frequencies\_tokens\_stop <- mydf %>%

unnest\_tokens(word, text) %>%

anti\_join(stop\_words) %>%

count(word, sort=TRUE)

print(frequencies\_tokens\_stop)

Text

Description automatically generated

### Sentiment analysis######

library(tidytext)

library(dplyr)

library(stringr)

library(tidyr)

library(tidytuesdayR)

nrcsurprise <- get\_sentiments("nrc") %>%

filter(sentiment == "surprise")

frequencies\_tokens\_stop %>%

inner\_join(nrcsurprise) %>%

count(word, sort=T)

nrcsurprise <- get\_sentiments("nrc") %>%

filter(sentiment == "surprise")

frequencies\_tokens\_stop %>%

inner\_join(nrcsurprise) %>%

count(word, sort=T)

afinn <- frequencies\_tokens\_stop %>%

inner\_join(get\_sentiments("afinn"))%>%

summarise(sentiment=sum(value)) %>%

mutate(method="AFINN")

bing\_and\_nrc <- bind\_rows(

frequencies\_tokens\_stop%>%

inner\_join(get\_sentiments("bing"))%>%

mutate(method = "Bing et al."),

frequencies\_tokens\_stop %>%

inner\_join(get\_sentiments("nrc") %>%

filter(sentiment %in% c("positive", "negative"))) %>%

mutate(method = "NRC")) %>%

count(method, sentiment) %>%

spread(sentiment, n, fill=0) %>%

mutate(sentiment = positive-negative)

library(ggplot2)

bind\_rows(afinn, bing\_and\_nrc) %>%

ggplot(aes(method, sentiment, fill=method))+

geom\_col(show.legend=FALSE)+

facet\_wrap(~method, ncol =1, scales= "free\_y")

Chart

Description automatically generated

####### Most common positive and negative words #########

bing\_counts <- frequencies\_tokens\_stop %>%

inner\_join(get\_sentiments("bing")) %>%

arrange(desc(n))

bing\_counts

Text

Description automatically generated

bing\_counts %>%

group\_by(sentiment) %>%

top\_n(10, n) %>%

ungroup() %>%

mutate(word=reorder(word, n)) %>%

ggplot(aes(word, n, fill=sentiment)) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~sentiment, scales = "free\_y")+

labs(y="Contribution to sentiment", x=NULL)+

coord\_flip()

Chart

Description automatically generated

###### N-grams and tokenizing ###############

library(dplyr)

library(tidytext)

library(tidyr)

library(tidytuesdayR)

rt\_bigrams\_new <- rt %>%

unnest\_tokens(bigram, text, token = "ngrams", n=2)

rt\_bigrams\_new #We want to see the bigrams (words that appear together, "pairs")

rt\_bigrams\_new %>%

count(bigram, sort = TRUE) #this has many stop words, need to remove them

Text

Description automatically generated

#remove stop words from the bigram data, use the separate function:

library(tidyr)

bigrams\_separated <- rt\_bigrams\_new %>%

separate(bigram, c("word1", "word2"), sep = " ")

#remove stopwords

bigrams\_filtered <- bigrams\_separated %>%

filter(!word1 %in% stop\_words$word) %>%

filter(!word2 %in% stop\_words$word)

#creating the new bigram, "no-stop-words":

bigram\_counts <- bigrams\_filtered %>%

count(word1, word2, sort = TRUE)

#want to see the new bigrams

bigram\_counts

Text

Description automatically generated

###### VISUALISING THE BIGRAM NETWORK ##########

#install.packages("igraph")

library(igraph)

bigram\_graph <- bigram\_counts %>%

filter(n>100) %>% #check the size, and edit proport

graph\_from\_data\_frame()

bigram\_graph

Timeline

Description automatically generated with medium confidence

#install.packages("ggraph")

library(ggraph)

ggraph(bigram\_graph, layout = "fr") +

geom\_edge\_link()+

geom\_node\_point()+

geom\_node\_text(aes(label=name), vjust =1, hjust=1)

Diagram

Description automatically generated

#polish the ngram to look better

a <- grid::arrow(type = "closed", length = unit(.15, "inches"))

ggraph(bigram\_graph, layout = "fr") + geom\_edge\_link(aes(edge\_alpha = n), show.legend = FALSE,

arrow = a, end\_cap = circle(.07, 'inches')) +

geom\_node\_point(color = "gold", size = 5) +

geom\_node\_text(aes(label = name), vjust = 1, hjust = 1) +

theme\_void()

A picture containing text, indoor, counter

Description automatically generated