

MIP  
POLITECNICO DI MILANO GRADUATE  
SCHOOL OF BUSINESS

WEB AND SOCIAL MEDIA ANALYSIS

---

# The Price of Bitcoin is Driven By the People's Sentiment

---

*Authors:*

PIZZINI Alessandro  
SALA Riccardo  
CECCARELLI Gianpaolo  
RIVERA Juan  
KOSSMANN Leonhard

*Assignment for Evaluation:*

WEB AND SOCIAL MEDIA ANALYTICS

May 30, 2021



## **Contents**

- 1. General Definition of the Business and Context**
- 2. Definition of the Main Targets/Personas**
- 3. Definition of Thinking Hats Outcomes (6 Hats)**
- 4. Definition of the Selected Idea**
- 5. Definition of the Customer Journeys**
- 6. Definition of the Data Analysis Workflows**
- 7. Description of the Implemented Prototype**
- 8. Mock-up**
- 9. Description of Analysis/Results**

## 1. General Definition of the Business and Context

A cryptocurrency is a digital or virtual currency designed for use as a medium of exchange. It uses cryptography to verify and secure transactions and to control the creation of new cryptocurrency units.

Cryptocurrencies are going back a long way, up until the year of 1983 where David Chaum conceived an anonymous cryptographic electronic money called “ecash”. Today other currencies are more known and have dominated the media and especially Social Media over the last decade. The most renown is surely Bitcoin, representing the fast cash Dream that many Crypto Investors have, and therefore we will speak more in detail about it.



Figure 1 - Bitcoin a fast cash dream

Bitcoin started with a value of 0.1\$/Bitcoin back in 2010 and grew exponentially over the years, reaching a Peak of nearly 17.000 in late 2017 and its all-time High of 64.784 Dollar on the 14 of April 2021.

Considering this immense value growth within the Bitcoin market in such a short amount of time, underlines the massive demand that exists for Cryptocurrencies. The number of “Blockchain wallet” users worldwide grew since 2011 constantly and exploded from time to time, leading to the peaks we can see in the timeline. Today, we count more than 73 Million owners of Blockchain wallets, mandatory to be able to buy or sell Bitcoin. This creates a clear picture of *possible costumers that are interested in applications helping them to manage or buy their Bitcoins*.

Seeing just the growth and trends over the last years in Bitcoin, may suggest a great *investment opportunity*. Although this surely can be the case, we should not forget the

*biggest risks* of the Cryptocurrency market; the *immense volatility* and the non-existing regulation methods to ease the market.

The *growth of Social Media Platforms* such as Facebook and Twitter play a major role for the value development of Bitcoin. The success and foundation of the movement lays within the IT community and was pushed to huge dimensions by these Platforms. In contrary today we see that exactly those Platforms are one of the major disruption factors. Personalities such as *Elon Musk* with Millions of followers on Social Media have created a certain cult around his persona, and the possibility to influence the entire Crypto market with a single tweet.

As an example of this we would like to refer to the [Tweet of Elon Musk on May 12<sup>th</sup> 2021](#), stating that Tesla will suspend the purchase of vehicles using Bitcoin due to the great amount of fossil fuels used to create them. This eventually led to a Drop of Bitcoin by 12% and even other Crypto currencies such as; Ethereum losing 14% and Dogecoin shedding nearly 20%. On the other hand, Tweets from Elon Musk suggested his sympathy with different Cryptocurrencies and his intentions to use them, eventually leading to the acceptance of Bitcoin for the purchase of a Tesla and therefore driving up the market.



Figure 2 - Tesla suspend Bitcoin payment

In this fast-changing market, we want to create a solution that helps Cryptocurrency investors to have a better overview what is trending on social Media and where the share price for certain cryptocurrencies is heading. The Idea is to create an Interface where our Software tracks in real time Tweets or News and formulates an Impact diagnosis of the current situation and the trend. Eventually this will be expanded with a trading bot reacting to the costumers interested resolving out of our analysis, combined with constant updates of the Market related news.

## 2. Definition of the main targets/personas

The main targets would consist of two parts, the *Analysis of the current market* and resulting out of these insights the *Prediction and Anticipation of the Market reaction*. These two targets help us to generate value for our Customers by giving them the time to react to those possible market drops or increases. This will eventually lead us to our main goal, to help our Costumers to better manage their Investments in Cryptocurrencies but also generate more worth by being able to early adopt to market trends.

Potential personas which would be interested in our Idea would mainly be two Groups:

- **The Heavy Investor (HODLERS)**, people that have been investing in multiple cryptocurrencies over a long period of time and managing a large amount of assets within their Wallet.
- **The Greenhorn Investor (NEWBIES)**, a big group out of all Investors are the new people joining the movement, drawn by Social Media and News about the potential of Cryptocurrencies. These Investors are mainly focussed on short term value adds and are not that informed nor constantly checking the market.

These two groups combine a good combination for our Costumer pool, potentially covering the big part of the newly added Costumers and the more Elite part of Heavy Investors that would be using our service to enrich their own Investment plans. Having such valuable insights about these two groups of Investors can be a major help for our own forecasting Analysis. The Heavy Investors probably register certain trends and movement quite early in the Market and start to buy or sell. The Big money nevertheless comes only with the big amount of Greenhorn Investors joining in on these trends.



Figure 3 - Bitcoin Investor



### 3. Definition of Thinking Hats Outcomes (6 Hats)



#### Facts

**Customer:** our focus lies in two customer groups consisting of new and inexperienced investors and on the other hand the more mature and experienced heavy investors

**Data Collection points Brainstorming,** defining certain players on social media which have a great outreach and are very trusted. Example would be Elon Musk, but further sources should be found and evaluated

Customer gain thru certain social media platforms, possible **cooperation with big crypto traders** such as Coinbase



#### Intuition

Creating a **higher security** for our clients' investments

Using the **freely accessible data** that has real impact on the market



#### Benefits

Using the knowledge of **how Tweets**, from certain personalities, **can impact the market** to predict crypto fluctuation

**Better prediction and understanding** which news and Tweets of which people influence the crypto market

Deeper understanding about what customers build **their buy or sell decisions on**

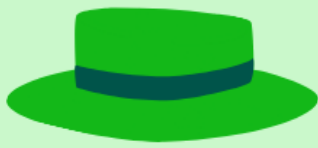
Conceptual thinking about the final recommendation to our customers, by understanding the **evaluation of the market and the bounce back**



#### Risks

**Evolution of the market**, will customers react in the future as excessive as now to Tweets from Elon Musk etc

Potentially becoming a **major impact factor our self**, when suggesting all the clients to sell



### Ideas

**Web scraping** and **social media API** to gather real time data and content from social media such as Twitter

**Sentiment analysis** to evaluate the sentiment around a specific cryptocurrency to anticipate market moves



### Process

Our goal is to use a variety of analytic tools to **predict shifts on the Cryptocurrency market**

The objective is to build a functional tool which is **easy to use, fast to reach and reports in real time**

Possibility of accessing the service by subscription **via internet browser**

Figure 4 - 6 Thinking Hats

## 4. Definition of the Selected Idea

Our idea is founded on the *recognition of the impact Tweets can have on the cryptocurrency market and therefore the price*. Out of this problematic evolves one of the biggest uncertainty factors, the risk of losing the worth of your investment in a very short period. By using sentimental analysis and different analytic tools, we try to analyse all the data on social media platforms, related to crypto and Tweets about it, and thereby anticipate a market reaction.

Our product will be based on a *subscription model*, gathering monthly income and different levels of subscription. Here is an example of 2 subscription levels:

- The **light version** Includes the option to receive newsletters and real time analysis of of daily tweets about Bitcoin only. You will also have access to one specific user's most relevant tweets from the last week regarding Bitcoin.
- The **premium version** will enable users the same options as in the light version but furthermore link the information gathered from the sentimental analysis, on an unlimited number of users, and link it to a trading bot. This trading bot will allow users to set certain thresholds, both considering prices and evaluation stages from our website, where the user can personally decide when an automatic buy or sell option should be performed by the bot.

All versions, even those that do not give access to a trading bot, will give advice on whether to sell or buy depending on the type of investor you are.

Our product will be accessible both in format of an app or a browser version. We would also consider the version specification based on what user type our costumer is and access this via a test after the registration. Based on further knowledge about the user we will be able to offer the most specific and personalised solution for every user group.

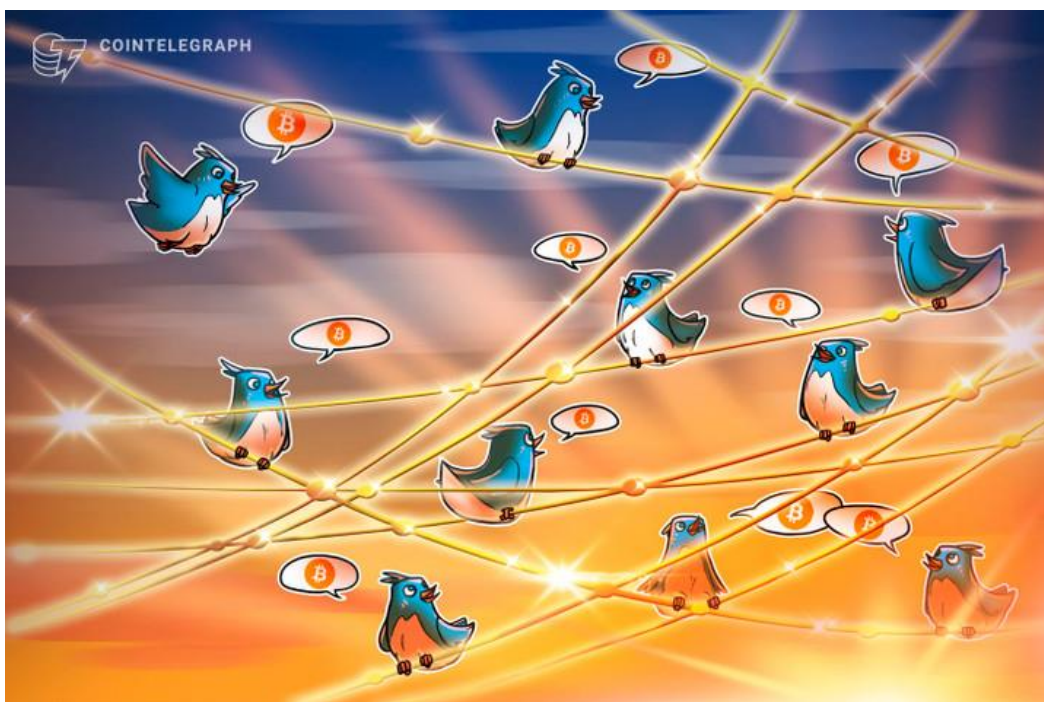


Figure 5 - Tweets have a Great Impact on Cypto's Price



## 5. Definition of the Costumer Journeys

As mentioned in the Definition of our Idea, we mainly want to focus on two main products that we can offer our clients. Since we thought about the difference in needs for each client, we agreed that both costumers want the same information from our product but might be differently focussed on details. Therefore, the costumer journey will differentiate a little bit, depending on what sort of client you are and what version you have.

The basic Costumer journey starts the same for each version, consisting of the *Registration* or *Log-in*. After this point the Versions will differentiate in the number of actions that can be performed and the functionalities coming with a full subscription.

### **The light Version:**

After the log in, the user will be able to access a Home area, which consists of the Cryptocurrencies he chose. In this Version a maximal amount of 3 different Versions is allowed, for further analysis the users need to subscribe. Behind every Cryptocurrency that the users chose is a small tendency arrow which indicates by colour and direction the trend resulting from our analysis. For further information gathering, the user can click on the Arrows and see a detailed analysis to which are attached the most significant tweets related to the trend of the currency.

Besides the three Currencies displayed on the menu, we will also have a news alert section, which will notify the user about any big news related to his chosen currencies. The news is being analysed and chosen given factors such as relevance of the author, retweets or comments mentioning on of the users Currencies. The user will be able to get our take on the impact of this tweet and therefore be able to make an early buy or sell action, since we deliver this information in real time and base it on the sentimental analysis of the tweets and comments.

Lastly the user has a profile section where he can connect to his personal accounts on different crypto platforms such as Coinbase and easily access his wallet. The profile section also offers the opportunity to join groups of likeminded investors and have a big forum where Q&A can be performed by the users. Having the opportunity of the users to interact creates a bigger connection to the platform and could be used to solve issues within the community rather than initialising corporate solutions.

### **The Full Version:**

The Full Version differentiates on the devices that you are accessing our product from. With the normal app, we will display the full version like the Light Version, only differentiating in the additional amount of information and functions that can be used.

Further applications to the light Version would be an explore options, where users can discover news and analysis that we performed with our opinion on hot topics and new cryptos which are increasing in demand. On this Option we try to provide the User the opportunity to reach out for further investments and base his decision on our analysis of the market. Furthermore, news involving cryptocurrencies that the user has not yet in his portfolio, would be referred to him. For example, the case of Elon Musk tweeting about Dogecoin would raise a news alert to our full version costumer and make him aware about the new post and currency. The full version would be unlimited in regards of currencies that will be tracked for the costumer. This would then be displayed on the Menu, each one ranked by the significance for our user, based on the amount of the investment and/or the chosen rank of the user.

The Browser version would also have an option where the user will be able to see Reports about different analysis performed by our system, considering different topics. Here the user could see more in depth for example the timeline of a post of a personality on twitter about a certain currency. We would then show a graph that would exactly show the stock evaluation of a certain currency over the time and display timeframes or thresholds where we would suggest our users to buy or sell.

One other major Part of the Full version is the Trading Bot which will function for the setting the user set. The user can select the Bot button and select from a range of different filters and constraints/thresholds how he wants to set the parameters of the functionality. On this specific page, the user will also have a personal contact from our company to cover the case of possible mistakes or problems. We will also provide a Video and a small tutorial that every user must see and a test he must pass to use and calibrate the Bot to his needs.

The community option which is also included in the light version will slightly differentiate for the full version users, as it would be possible to interact with premium Investors and/or personalities which would also join the App. Here the user could send a request to personally interact with a famous user or with one of our Analyst and talk about investment strategies and/or risk management. We would like to give the Full version user the possibility to have all the comfort of other social media sites and combine it with an easy-to-use software.

In order to create a more unique and special experience, we offer the personalisation of certain options. The user would be able to specify the news alert section based on a variety of options, consisting of different people or currencies, that would be analysed and send directly to the user. In addition, the user would be able to personalise the analysis by choosing specific users on twitter and or specific news channel. Thereby we would analyse the areas which can be specified by the user and provide a more user-oriented analysis that is more personal.

## 6. Definition of the Data Analysis Workflow

After the very important endorsement by Tesla CEO Elon Musk, *Bitcoin-related social media activity is at an all-time high*. In recent months, Tweets related to BTC (Bitcoin) are in the *hundreds of thousands every day*.

For these reasons, we decided to extract and analyse the Tweets related to a particular currency and the most relevant Tweets about cryptocurrencies written by influential people such as Elon Musk.



Figure 6 - Elon Musk Talking About DogeCoin

We have mainly done two separate analyses to provide different insights. The main difference lies in the time period to which the data refer, we will talk about a *real time analysis* of the latest news and an analysis on a *more extended historical period* in order to verify our initial assumptions.

In our web application we want to provide the user with the ability to know the general sentiment of users regarding a specific cryptocurrency each time they access their home page. This data will reflect the current market situation and how that specific cryptocurrency is perceived by twitter users.

To do this, we mainly performed a sentiment analysis on tweets filtering them by cryptocurrency and by user.

**Short-term analysis** - daily analysis by search word and weekly analysis on a specific account

In detail, we will implement a strategy for the extraction of the tweets of *today* with a specific hashtag (related to a specific cryptocurrency and selected by the user) and automation for a real-time extraction and analysis.

- I. **Get tweets** using the **Tweepy library** (tweets mining).
- II. Definition of **the average sentiment score**
- III. Definition of the **trending tweets**
- IV. **Word counts** and **Word clouds**
- V. Provide the final user with easily understandable **graphs and insights**

**Long-term analysis** - monthly analysis using specific words related to a certain cryptocurrency (e.g. Dogecoin and Bitcoin)

In detail, we will implement a strategy for the extraction of the tweets of previous months and of current month for all the cryptocurrencies relevant and of interest to our users.

- I. **Scrape tweets from Twitter** in different time periods and for different currencies using **snscraper** (tweets mining).
- II. **Comparing the positive/negative sentiment scores for each month**
- III. Definition of the **polarity** of the tweets over the time
- IV. **Comparing the sentiment (polarity) with the prices** over the time
- V. **Word counts** and **Word clouds**
- VI. Provide the final user with easily understandable **graphs and insights**

Both data analysis workflows (short-term and long-term) are iterative and a further step, the **Cron job**, must be added. The cron job is a way to schedule some repetitive actions. Here the cron job is used to download the tweets of the previous day every day.

## 7. Description of the Implemented Prototype

### 7.1. Short-Term Analysis

For a better understanding we suggest following this paragraph by opening the reference notebook → [01\\_Bitcoin\\_twitter\\_analysis](#)

**Data Acquisition (tweets mining):** Twitter has an API that allows, after a registration as a developer, to extract tweets with some limitations but in an easy way.

The extraction can be done based on the topic of interest (in our case #bitcoin) or querying the API based on specific different endpoints (e.g. username, twitter id etc.). The twitter has some limitations for the data extraction: The number of tweets is limited and the time period of extraction stops at the last 7 days.

The extraction was done based on the hashtag #Bitcoin and the user *Elon Musk*. We managed to extract the last 1000 tweets with the following characteristics: #Favourite, #Retweets, the timestamp, the text, etc.

**Data Preparation:** The data obtained from the extraction were very dirty and needs an accurate cleaning in order to perform the corresponding analysis. In particular, the text of the tweet may have several elements that could produce noise during the sentiment analysis; we have then created functions to clean and to perform lemmatization and stemming.

**Sentiment Analysis (TextBlob):** The most important analysis on which our proof of concept is mainly focused is a sentiment analysis of the tweets.

Sentiment analysis is one of the most used machine learning techniques in social media because it allows the extraction of two numerical components that represent the sentiment of the text under examination.

We want to associate a sentiment composed of polarities, positivity and negativity, and its subjectivity to each tweet in order to provide our client with an updated average sentiment for a given cryptocurrency.

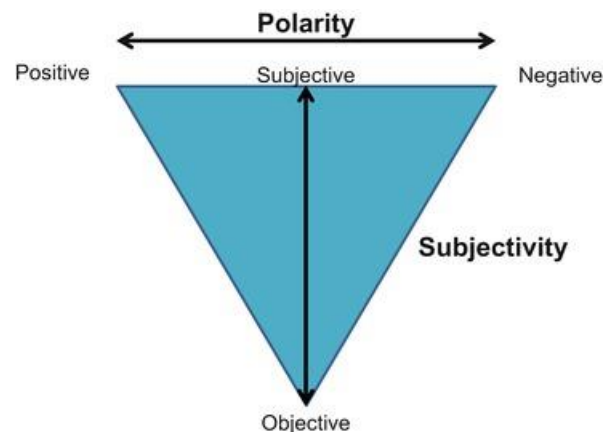


Figure 7 - Polarity and Subjectivity

To implement our analysis, we have used TextBlob, one of the most basic and used libraries in this field. To our dataframe we have added three different columns: Polarity, Subjectivity and overall Sentiment

At this point, grouping the information we tried to extract the insights that such as the total number of tweets for each sentiment category, the average value of polarity and as shown below which are the sentiment regarding the most popular tweets in term of likes and retweet.



**Visualization: Word cloud and Wordcount:** To conclude our first analysis and to have a better idea of the insights obtained we have created word clouds that allow us to quickly see which are the most influential words within the tweets considered. The insights that come out are very curious and reflect the sentiments previously calculated. For example, for the positive tweets within the most influential words we have found: 'Future, airdropdet (a famous twitter page specialized in cryptos), successful, hopefully and project'.



Figure 8 - Bitcoin Word Cloud

After this daily analysis we performed an in-depth analysis of the past week *Elon Musk* tweets, by identifying only those related to Bitcoin and extrapolating only the most relevant (*trending*) ones. A more detailed analysis of the impact Elon Musk's tweets have on the price of Bitcoin will be made later.



Figure 9 - Tweet from Elon Musk (24 May)

## 7.2 Long-Term Analysis

For a better understanding we suggest following this paragraph by opening the reference notebooks → `02_Twitter_scraping_snsscrafer` ; `03_Crypto_analysis_months`

The workflow followed in this analysis corresponds for the most part to that set out above. The main difference lies in the acquisition of the data and the ultimate goal of our research. The techniques used such as sentiment analysis and NLP process remain the same.

The period of time considered for this analysis is the key of our research because we want to find a correlation between an average sentiment of a given long period with the price of the cryptocurrency considered.

**Data Acquisition:** We extracted 300,000 tweets for four different months (January, February, April and May) regarding Dogecoin and January for Bitcoin.

We focused on dogecoin because we know of a particular episode that has shocked the world of twitter and that we want to investigate to verify our hypothesis.

As said before the Twitter API presents some restriction so we have implemented the use of a web scraper called 'SNS: Social Network Scraper' that didn't require credentials and has no limit of extraction. It was a hard task because there were few documentations about, to manage it we have cloned the source code by a GitHub folder relying on python language.

All you have to do is setting some parameters such as the number of tweets requested and the date of interest, the user/and the hashtag.



*Figure 10 - Meme About Dogecoin and Elon Musk*

**Data Preparation:** After having eliminated many superfluous columns, we proceeded in the same way as before, using the same functions to make the content of tweets analysable (Lemmatization and Stemming).

**Data Analysis:** The analyses performed on this data are the same as those done previously, but with a different purpose. After extracting then the polarity and subjectivity of each of these tweets we went to investigate this sentiment from a temporal point of view.

The datasets were grouped in order to make the visualization clearer by having an average value of sentiment for each hour of the chosen period and then go to build time-line graphs.

This study is useful to be able to compare the perceptions of users in different months about a specific cryptocurrency and then investigate the relative price fluctuation.

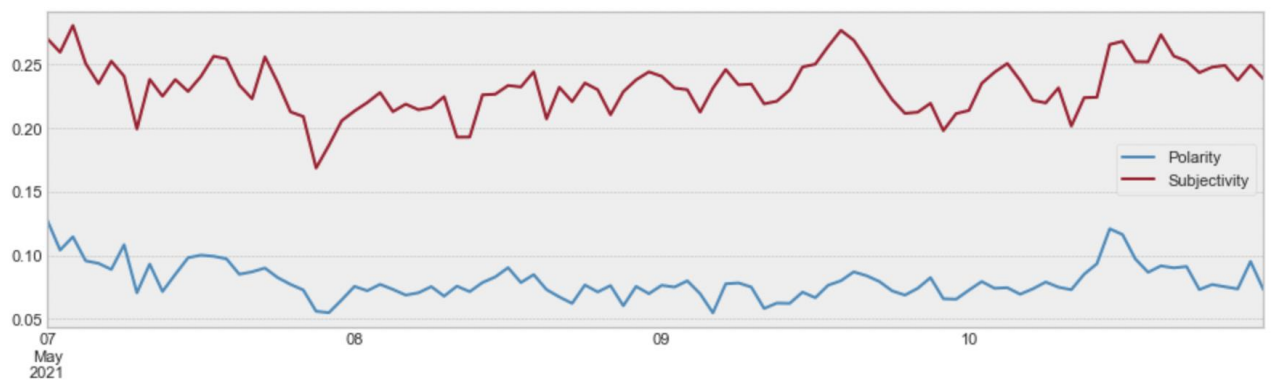
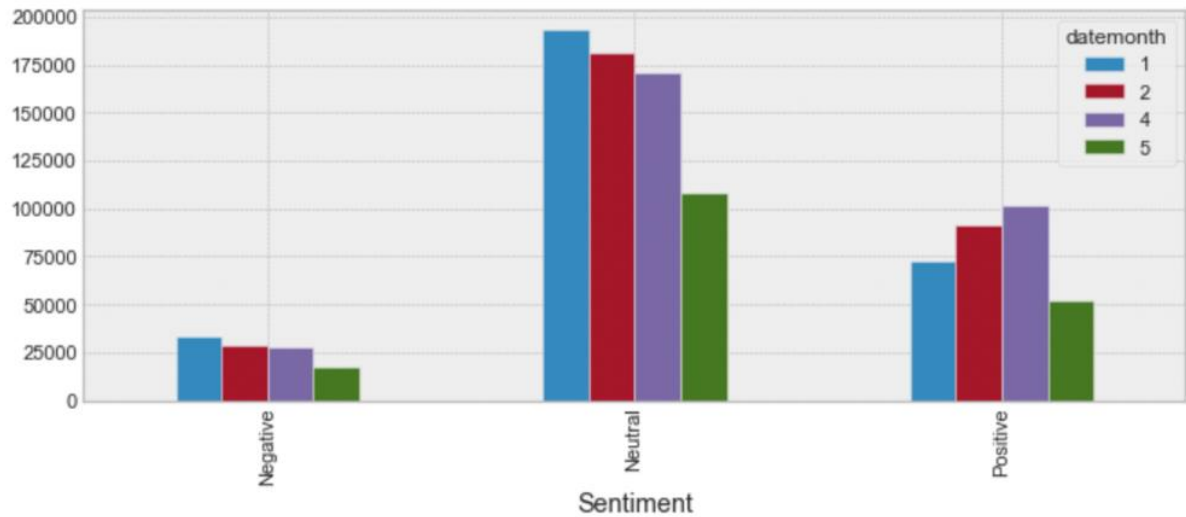
In this regard we have chosen as period of analysis the one related to May and more precisely from 7 to 10 May, because after Elon Musk's speech on Saturday Night Live, Twitter went crazy and consequently so did the price of Dogecoin.



*Figure 11 - Elon Musk at Saturday Night Live*

Our goal is to research this drastic drop in price of dogecoin because of his statement and to go and study how the twitter world reacted to the event and how sentiment about it changed. To do this we have graphically represented the results obtained from the sentiment analysis for that period and we have done a web search on the relative price trend

**Data visualization:** In order to produce meaningful insights, the visualization of this analysis is fundamental. Having a longer period of time, we were able to plot our data along time lines to highlight how the world of cryptocurrency and therefore also sentiments about them vary greatly. Below we report the very variable trend of polarity and relative subjectivity regarding the period of 'Elon Musk event' and the bar plots that compare the different sentiments in our different periods.



As you might imagine, a simple word cloud shows that, apart from dogecoin and it's like, the most tweeted word refers to Elon Musk and his now famous motto: "to the moon".



## 8. Mock-up

For a better understanding of our PoC and what the final result might look like after a few adjustments and implementations, we developed a mock-up of a web dashboard.

This dashboard shows just one of the many functionalities that could be implemented with our analysis, specifically it shows the influential tweets **(1)** in a specific day, according to the selected button **(2)**. A simple daily sentiment analysis and the daily price of the cryptocurrency are then shown **(3)** with the price trend index and the daily sentiment score **(4)**. In addition, a word cloud with the most recurring daily words and a bar plot containing the number of occurrences of individual words, and respectively how many times they have appeared in positive and negative tweets is shown **(5)**.

On the basis of these analyses, the system advises the investor what he should do, distinguishing whether he is a long-term investor (HODLER) or a day/swing trader **(6)**.



Figure 15 - Mock-up



## 9. Description of Analysis and Results

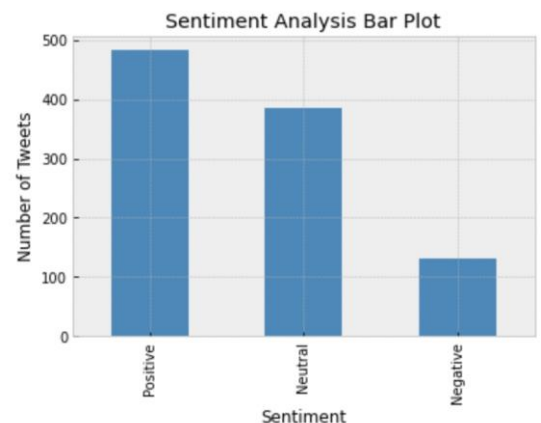
We are aware of the magnitude of our assumptions, the results presented below reflect them and try to be as realistic as possible also considering the fact that we conduct some web researches that confirm the link between social media and cryptocurrencies. It is a hot topic that is discussed a lot in this period. Our analysis to be the most affordable as possible should be implemented more in the future with additional analyses and additional data. However, our interpretation of the results tries to be as realistic as possible considering the many limitations found.

The analyses illustrated previously provides numerous insights, here we bring back the three more interesting and that they can have a development application to the inside of our web application.

### Overall sentiment nowadays regarding bitcoin

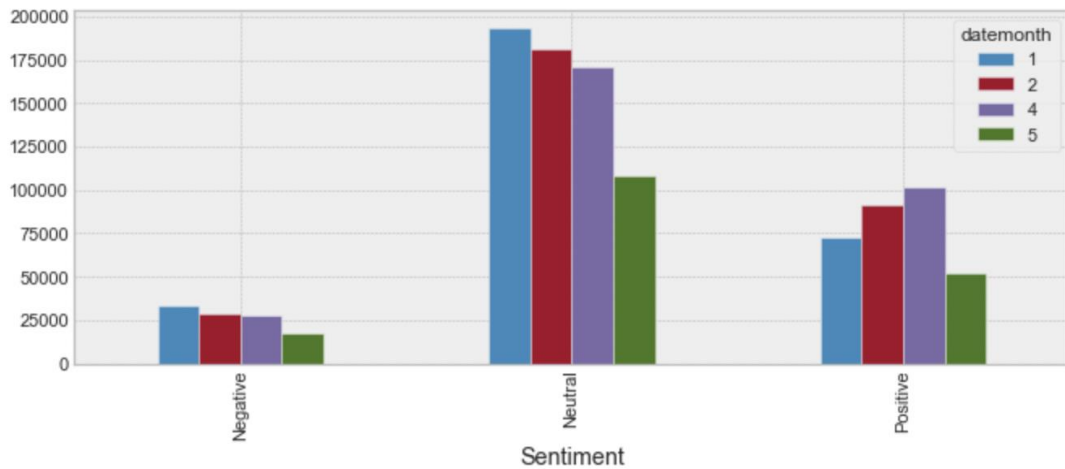
From the sentiment analysis we can provide to the user the overall sentiment of the twitter world regarding bitcoin. We made this analysis the 27 of May, this is our results. we can see that the general sentiment is positive and that this positivity reflects a quite good increase of its price

BITCOIN	
POLARITY	0.35
SUBJECTIVITY	0.16



### Which was the best period to invest on Dogecoin - Price and Sentiment

Having seen that from a sentiment point of view April was the month in which the highest number of positive tweets were recorded compared to the other periods considered in the analysis.



We can expect this to produce a growth in the price of dogecoin in that period. The evidence of this is shown in the graph below where we can see that the price of dogecoin increases in that period.



### Elon Musk Interview at Saturday Night Show on 8 May 2021

This result maybe is the most significant of our project. It fully represents how media and social media exposure reflects the actual market performance of the cryptocurrency price, in this case dogecoin.

During the show, Elon Musk made a sarcastic joke against Dogecoin causing a drastic drop in the price of the coin and making the whole “*twitterverse*” talk about the incident. Our results show that

on twitter the most influential word of those days is just Elon Musk and the graphs below represent the great volatility of the period in terms of social media sentiment and market price.

in the graph of the price it is obviously clearer the drastic descent connected to the event, tweets as a source of information do not represent 100% truth, but we can see a certain similarity in particular when both sentiment and price decline

