Universidad del País Vasco
Euskal Herriko Unibertsitatea

EKONOMIA ETA ENPRESA FAKULTATEA
FACULTAD DE ECONOMÍA Y EMPRESA

DEGREE: ECONOMICS

Academic Year
2020/2021

TWITTER SENTIMENT ANALYSIS AND STOCK MARKET MOVEMENTS: U.S. ELECTIONS 2020

Author: Gaizka Lete Segura

Advisor: Miren Arantzazu Gorostiaga Alonso

Bilbao, 25th June 2021

Abstract

The aim of this work is to analyze if there is a relationship between general twitter political sentiment and the movements of the stock market, more precisely, the American indexes S&P 500, DJIA and Nasdaq Composite. To do so, a short window event study made of the U.S. presidential elections 2020 was made. Between October 27th and November 8th, being the elections on the 3rd, more than 400.000 tweets were collected using the hashtag “Election Day”. To get the sentiment of these tweets, the sentiment analysis tool “VADER” was used. After this, a collection of correlation and regression analyses was made. Regarding the results, we cannot make concrete conclusions, as they were not statistically significantly different from zero. However, some trends are met and show that sentiment analysis used on Twitter could be used in future projects.

Keywords: Sentiment analysis, stock-market, U.S elections, VADER, Twitter.

Resumen

El objetivo de este trabajo de fin de grado es analizar si existe relación entre el sentimiento político general de Twitter y los movimientos del mercado de valores, más concretamente los índices americanos S&P 500, DJIA y Nasdaq Composite. Para ello, se ha realizado un estudio de un evento de corta duración como lo son las elecciones presidenciales de Estados Unidos del año 2020. Se han recogido más de 400.000 tweets utilizando el haghtag “Election Day” pertenecientes al periodo comprendido entre el 27 de octubre y el 8 de noviembre, siendo el 3 de noviembre el día de las elecciones. Para extraer el sentimiento de los tweets se empleó la herramienta “VADER”. Posteriormente, se llevaron a cabo diferentes análisis de correlación y de regresión. No podemos extraer muchas conclusiones de los resultados, ya que no pueden ser determinados como estadísticamente significativos distintos de cero. Sin embargo, algunas tendencias sí se cumplen y muestran que las herramientas de análisis de sentimiento utilizadas en Twitter pueden ser utilizadas en futuros proyectos.

Palabras clave: Análisis de sentimiento, mercado de valores, elecciones de Estados Unidos, VADER, Twitter.

Laburpena

Twitter-eko sentimendu politiko orokorraren eta balore-merkatuko mugimenduen, hain zuzen ere S&P 500, DJIA eta Nasdaq Composite indize amerikarren artean, erlaziorik dagoen ikertzea da gradu amaierako lan honen helburua. Horretarako, 2020-ko Estatu Batuetako presidentziarako hauteskundeekin denbora laburreko ekitaldi baten azterketa egin da. Urriaren 27tik azaroaren 8ra, hauteskundeak 3an izanda, 400.000 txio baino gehiago batu ziren “ElectionDay” hashtag-a erabiliz. Txioen sentimendua lortzeko, “VADER” sentimendu analisi tresna erabili zen. Honen ostean, hainbat korrelazio eta erregresio analisi burutu ziren. Emaitzak erreparatuz, ezin ditugu ondorio zehatzak atera hauek ez direlako estatistikoki esanguratsuak. Hala ere, joera batzuk bai betetzen dira. Gainera, Twitter-en erabilitako sentimendu analisiko tresnak etorkizuneko proiektuetan erabili daitezkeela erakusten du.

Gako-hitzak: Sentimendu analisisa, balore-merkatua, Estatu Batuetako hauteskundeak, VADER, Twitter.

Table of Contents

Index of Figures

Index of tables

1.Introduction

Twitter was founded in March 2006. It is a social media platform that allows users to publish short microblogs, known as tweets, limited to 280 characters. Those tweets can be sent directly to users or posted publicly to followers and the rest of users of the platform. The last time Twitter disclosed its monthly active users was in the first quarter of 2019 when it averaged 330 million monthly active users (Tankovska 2021). It is also the 11th most visited website with more than 3,4 billion monthly visits, as of December 2020. (Clement, 2021)

Another feature of Twitter is the use of hashtags, key words preceded by "#" that allow tweets to be collected in threads. Thanks to this, users can follow specific events and topics and start interactions. Then Twitter makes a "trending topic" list tracking the most mentioned hashtags and words, letting users follow the most popular affairs at any time. Users can also "retweet" tweets, which, as its own name says, re-posts the original tweet on the account of the user retweeting. This allows original tweets to go all along users and eventually, some tweets will go viral when retweeted by thousands of users.

Twitter is just one example of all the sites on the Internet where people have access to information, opinion and discussion among people all around the world. This new way of communication, which includes social media, blogs, and so on, started growing really fast in early 2000, because of the democratization of the Internet, and thanks to the inception of smartphones. Along with this, for the first time in history a huge volume of people's opinions began to be recorded.

This big amount of recorded opinion motivated the birth of what we call sentiment analysis which belongs to the field of Natural Language Processing (NLP). The NLP field of study mixes, on the one hand, techniques of computer and information science, mathematics, electrical and electronic engineering, artificial intelligence and robotics, and on the other hand, linguistics and psychology. The aim of this area of research is to improve our knowledge on how humans use and understand language in order to develop techniques to allow computers to understand and analyze language data, written or in speech (Grosz, 1982).

The applications of this field are wide, for example, the autocorrect and autocomplete mechanisms used in the google search bar, voice assistants, web language translators, chatbots used in customer service, targeted advertising, among others.

As we have seen, the uses of NLP and the fields that benefit from this are wide. Nevertheless, what interests us in this work is sentiment analysis. Sentiment analysis is the part of the NLP field that "analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" (Liu, 2012).

The goal of this work is to analyze if there is a relationship between general political sentiment around the 2020 U.S. elections and the movements of the stock market. To do so, we are going to follow the analysis study in Nisar and Yeung (2018) to the UK elections 2016. To assess the general political sentiment, we are going to analyze data recorded on Twitter and then assign a sentiment with a sentiment analysis tool called "VADER" Then we are going to make a collection of correlation and regression analysis to test our whether the stock market and sentiment captured by tweets are related.

In the next section a brief review of the literature will be made in order to put our work in context. The previously mentioned sentiment analysis and the correlation and regression analysis will be explained in section 3. The different data collection methods will be described in section 4. In section 5 we are going to explain the hypotheses to be tested. The results of our analysis will be presented in section 6 and to end we will make a conclusion regarding the results of this work.

2. Literature review

Since the beginning, sentiment analysis drew the attention of academics because, on the one hand, was an interesting and innovative field where a lot of different research could be made, and secondly, because it has applications in so many different areas, some of them could even be commercialized.

In this section, we will discuss some articles that, although they have different motivations, aims and the techniques they use are different from what we are going to make, are interesting in order to learn how sentiment analysis, and data extracted from Twitter can be applied in a field like finance.

Bollen, Mao, and Zeng (2011) studied if the collective mood taken from Twitter feeds are correlated with the value of the Dow Jones Industrial Average. They found an accuracy of 87,6% in the prediction of the daily movements of the DJIA index when adding mood variables.

In Rao and Srivastava (2012) the authors try to identify if there exists a relationship between Twitter sentiment and the short term performance of different companies and indexes. They take more than 4 million tweets over a time of 14 months, found strong correlation between Twitter sentiment and the stock market movements.

Other research focused on tweets that contain financial information. For instance, Mao et al. (2012), in order to get the tweets that mention every company on the S&P 500, they search for tweets that contain the company's symbol prefixed by a dollar sign, that is how people usually mention the companies on Twitter. Applying this method allows them to make analysis at three different levels, stock market, industries and individual stocks. In this research they find that the daily number of tweets is correlated with the stock market indicators, so in their opinion including Twitter data in the models can be useful to predict whether the S&P 500 closing price will go up or down.

Another surprising result was found by Souza et al. (2015) who investigated if there was relationship between twitter sentiment and volume, and stock movements and volatility while they also made a comparison with traditional newswires. They actually found that for the companies in the retail sector Twitter's sentiment presented a stronger relationship with stock market returns than traditional newswires.

Burggraf, Fendel, and Huynh (2020) investigated the influence of political news on stock market movements. They analyzed tweets from Donald Trump's personal account and found that tweets related to the US-China trade war negatively affected stocks in the S&P 500 index and caused an increase in volatility measured by the Board Options Exchange Market Volatility Index, VIX.

3. Methodology

This section will be divided in two, firstly, we are going to explain more about sentiment analysis and the tool we are using, “VADER”. Then, the statistical analysis used to test our hypotheses will be explained.

3.1. Sentiment analysis and “VADER”

As we have said, sentiment analysis includes different computational methods which analyze text written by people and assign them a sentiment in order to identify people’s opinion in relation to products, services, politics and so on.

One of the most important tools for sentiment analysis are the “sentiment lexicons”. Sentiment lexicons are collections of words that are labeled taking into account their semantic meaning. In some lexicons, words are categorized in a binary way, for example, classifying words as positive or negative, or whether they indicate happiness or anger. One of the most important issues of this kind of lexicons is that they do not capture the intensity of the words. For example, “the match was good” and “the match was wonderful” would have the same positivity score, although the latter has more positive intensity than the former.

Other lexicons go a bit deeper and classify words by giving them a score of sentiment “intensity”. These kind of lexicons assign a score to each word going from, for example, -4 to 4, completely negative to completely positive respectively, with a neutral point, that in our example would be 0. These ranges help researchers to measure the intensity of people’s opinions.

Another way of increasing the accuracy of the sentiment associated to words, regardless of the lexicon used, consists in taking into account the context-awareness of each word. Words that may have more than one meaning depending on the context are classified taking its context into consideration.

Usually, lexicons have issues in social media context because they do not capture the way in which text is written on social media, using for instance emojis. In addition, some lexicons also fail on grasping general sentiment intensifiers. For example, “The match was good” and “the match was really good” do not express the same positivity.

(Hutto and Gilbert 2014) proposed “VADER” in order to solve these issues and make a sentiment analysis tool that works well on social media but that can also be used in different domains. For doing so, firstly, they took some pre-existing and well established lexicons, and added to them a list of emoticons, then sentiment related acronyms and initialisms such as LOL or WTF and finally, they added some slang which expresses sentiment like “nah” or “meh”. With these addings they tried to capture better the way emotions are expressed on social media. Then, they used the approach of giving a score to each word depending on the sentiment they express. The scale goes from -4 (completely negative) to 4 (completely positive), being 0 neutral.

Finally, they implemented five heuristics that humans use to express sentiment intensity in texts. The first one is that exclamation marks increase the intensity of the sentiment without changing the sentiment itself. Secondly, the use of capitalization, specially using all-caps in a relevant word around other ones that are not written in all caps. This increases the intensity but does not change the direction of the sentiment. Thirdly, they implemented degree intensifiers in order to solve the issue previously mentioned. Fourthly, when the word “but” appears in a phrase it changes the sentiment of the phrase, So to capture this “VADER” mixes the sentiments of the two parts of the sentence with the latter half having more importance on the overall score. Finally, “VADER” examines the three words preceding a word that expresses a sentiment, to capture better when a negation changes the sentiment of a text.

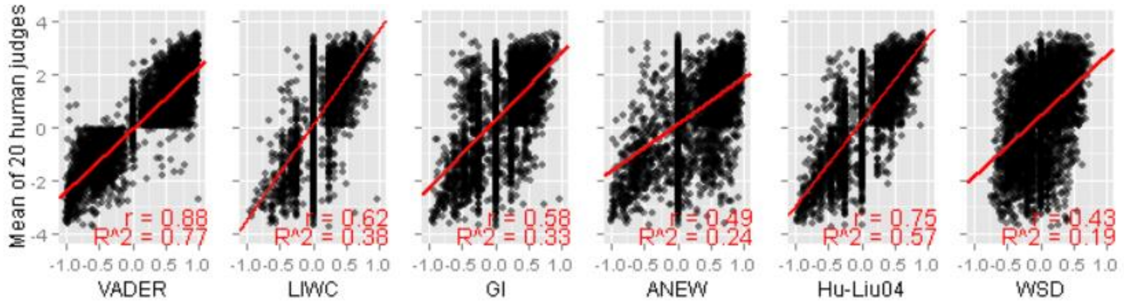
Figure 1: Example of the results obtained by applying the five heuristics in “VADER”

	text	word_scores	compound	pos	neu	neg	but_count
1	The match was good	{0, 0, 0, 1.9}	0.440	0.492	0.508	0.000	0
2	The match was good!!!	{0, 0, 0, 1.9}	0.583	0.557	0.443	0.000	0
3	The match was GOOD	{0, 0, 0, 2.633}	0.562	0.548	0.452	0.000	0
4	The match was really good	{0, 0, 0, 0, 2.193}	0.493	0.444	0.556	0.000	0
5	The match was good :)	{0, 0, 0, 1.9, 2}	0.710	0.663	0.337	0.000	0
6	The match was not good	{0, 0, 0, 0, -1.406}	-0.341	0.000	0.624	0.376	0
7	The match was good but we did not pass a good time	{0, 0, 0.95, 0, 0, 0, 0, 0, -2.109, 0}	-0.287	0.129	0.664	0.206	1

Figure 1 presents the application of the five heuristics in real examples. Vader returns a vector containing the score of each word, an overall score for the text named as compound, going from -1 (totally negative) to 1 (totally positive), the weighted percentage of negative, positive and neutral words, and the frequency of the word “but” for each tweet. Regarding the compound score of each phrase, we see that this value changes depending on the meaning of the phrase.

In (Hutto and Gilbert 2014) it is shown that using a sample of 4000 tweets “VADER” has human precision on computing sentiment. Although this could be true, having a look at the sentiment scores of tweets in our data set it could be said that like many other sentiment analysis tools Vader has difficulties computing negative sentiment to those that are neutral.

Figure 2: Accuracy comparison between VADER and other sentiment analysis tools.



Source: Hutto and Gilbert (2014)

Figure 2 helps us visualize VADER´s classification precision and compare it with other sentiment analysis tool. In the Y axis we see the value given by 20 human judges and on the X axis the value given by each tool to a data set of 4,000 tweets. In general, paying attention to the vertical line on the 0.0 value, we see that the majority of the tools classify as neutral, tweets that human judges consider non neutral. If we divide each plot into four quadrants, being lower left the true negatives and upper right the true positives we see that although “VADER” does it better compared to the rest of the sentiment analysis tools, it is not perfect either.

3.2. Statistical analysis

In order to test our first pair of hypotheses ($H0_1$ and Ha_1), two different multiple regression models will be run which will include for each day t the volume of tweets "$Vtweets_t$", the volume of positive tweets "$Vpos_t$" and the volume of negative tweets "$Vneg_t$" as the independent variables, and volume of the daily trades in each index and the daily closing price of each index, as the dependent variables. The aim of a multiple regression model is to explain the relationship that each independent variable has with the dependent variable, taking into account the other independent variables.

Multiple regression models:

$$Volume_t = \alpha + \beta_1 Vtweets_t + \beta_2 Vpos_t + \beta_3 Vneg_t + \varepsilon_t$$
$$Close_t = \alpha + \beta_1 Vtweets_t + \beta_2 Vpos_t + \beta_3 Vneg_t + \varepsilon_t$$

Alpha, α, is the intercept of the model. The value of the intercept represents the value of the dependent variable when the independent variables are 0.

We denote as, β, the coefficient of each variable. The value of the coefficient indicates the change in the dependent variable when the variable that goes with that coefficient increases in 1 unit.

ε is the error of the model. It represents the difference of the observed outcome and the outcome we would predict based on our model.

To get similar and consistent scales for all the variables in the regression exercise, we standardize our variables, "Vtweets", "Vpos" and "Vneg", along with the volume and close variables of each index, with the mean and standard deviations.

To test our second pair or hypotheses ($H0_2$ and Ha_2), we are going to analyse the relationship between the average daily mood "MOOD" and, as the dependent variable, the daily change in price for each index. And secondly, test the relationship between "MOOD", as the independent variable, and the daily closing price of each index, as the dependent variable.

4. Data Collection

In this section the methods used to obtain our data collection will be explained. In order to do so, we will divide it into two subsections, one regarding the twitter data collection, and the other one regarding the index data collection.

4.1. Twitter data collection

Common twitter scraping methods use twitter API (Application Programming Interface) to access the data, but they have limitations regarding the total number of tweets that can be requested and in regard to the date in which those tweets were published. In order to solve those issues, "Twint" was used for this research. Twint[1] is a twitter scraping

[1] https://github.com/twintproject/twint/wiki

tool written in Python that without using Twitter's API allows the downloading of tweets, letting you scrape from users, hashtags, trends, and specific topics. It only has limitations when using it to scrap from a user's timeline, nevertheless we did not use it with this intention.

A total of 423,829 tweets was collected in the time window between October 27th 2020 and November 8th 2020. Those dates were taken to measure the mood prior, during and after the Election Day, November 3rd.

Figure 3: Caption Twint output

	date	time	tweet	hashtags
1	2020-10-27	23:59:49	@RepLizCheney @RichLowry You haven't "improved...	['gop', 'resist', 'maga', 'kag', 'trumpmeltdown', 'bide...
2	2020-10-27	23:59:49	Mercury is going direct on #ElectionDay. Hmm...	['electionday']
3	2020-10-27	23:59:46	#ElectionDay #VoterIntimidation	['electionday', 'voterintimidation']
4	2020-10-27	23:59:32	Just found out production is on hiatus next Tuesday,...	['electionday']
5	2020-10-27	23:58:30	#AmericaOrTrump #freedom #vote #VOTE #Electi...	['americaortrump', 'freedom', 'vote', 'vote', 'election...
6	2020-10-27	23:58:20	I imagine Don dreams of a time when he can return ...	['vote', 'electionday', 'election2020', 'votebluetosav...
7	2020-10-27	23:57:53	At this point I really wonder how many write in's Th...	['vote', 'voteearly', 'votethanos', 'election2020', 'ele...
8	2020-10-27	23:57:29	Thankfully the #wind is weakening and the fire dan...	['wind', 'california', 'october', 'weather', 'forecast', 'e...
9	2020-10-27	23:57:24	Thankfully the #wind is weakening and the fire dan...	['wind', 'california', 'october', 'weather', 'forecast', 'e...
10	2020-10-27	23:56:53	CBS hires 24-hour security for #LesleyStahl after sh...	['lesleystahl', 'trump', 'usa', 'trumpsupporters', 'hate...
11	2020-10-27	23:56:49	#Trump tells large crowd he's trying to get"your hu...	['trump', 'covid', 'misogyny', 'trumphascovid', 'trum...
12	2020-10-27	23:56:37	We have 7 days to go!!!!!! Hey, drop a comment a...	['electionday', 'vote']
13	2020-10-27	23:56:33	@JoeBiden As #ElectionDay approaches, I discusse...	['electionday', 'covid19', 'election2020']
14	2020-10-27	23:56:00	The deficit was virtually identical to a poll from the ...	['electionday', 'elections2020']
15	2020-10-27	23:55:14	It's crazy that the 42nd (Clinton), 43rd (G.W. Bush), ...	['trump', 'biden', 'electionday', 'election2020']
16	2020-10-27	23:54:48	Cada voto cuenta (pancarta en el caserón de la Soci...	['election2020', 'electionday']
17	2020-10-27	23:54:47	I hope I'm wrong but Trump is going to steal the ele...	['election2020', 'electionday', 'leadership', 'trump', '...
18	2020-10-27	23:54:46	#ElectionDay #Elections2020	['electionday', 'elections2020']
19	2020-10-27	23:54:33	Anyone else noticed the black gloves .@realDonaldT...	['trump', 'donaldtrump', 'trumprallymichigan', 'mag...
20	2020-10-27	23:53:37	Wisconsin folx...drop your ballot off. Do NOT mail it. ...	['wisconsin', 'vote', 'election2020', 'electionday']
21	2020-10-27	23:53:29	Hey guys, make sure you go and #vote ASAP! We h...	['vote', 'electionday']
22	2020-10-27	23:53:01	We have to vote for Trump in the 2020 election . #K...	['keepamericagreat', 'makeamericagreatagainagain'...
23	2020-10-27	23:52:51	#Election2020 #ElectionDay #VOTE https://t.co/p...	['election2020', 'electionday', 'vote']
24	2020-10-27	23:52:29	The deal between Trump & the Evangelicals is ...	['electionday']
25	2020-10-27	23:52:15	@SteveKornacki A week to go. Have you been eatin...	['electionday']
26	2020-10-27	23:51:37	HEADS UP, MAIL VOTERS! The Hillsborough County ...	['electionday', 'vote', 'voteearly']
27	2020-10-27	23:51:26	Holy... #COVID19 #ElectionDay	['covid19', 'electionday']
28	2020-10-27	23:51:02	Election Day is in 1 WEEK! Make your voice heard b...	['votingearly', 'votingbymail', 'electionday', 'election...
29	2020-10-27	23:50:48	Well this is not good. Vote early, folks. #ElectionDay...	['electionday']
30	2020-10-27	23:50:12	📦One week until #ElectionDay ! #GOTV #VOTE in ...	['electionday', 'gotv', 'vote', 'election2020', 'justice',...
31	2020-10-27	23:50:12	@Ordinary1World #Election2020 #ElectionDay #de...	['election2020', 'electionday', 'debates']
32	2020-10-27	23:50:09	@MollyBeck Can you say swing state #ElectionDay	['electionday']
33	2020-10-27	23:50:08	One week left before #ElectionDay and early voting...	['electionday', 'vote', 'voteearly', 'aisdvotes']
34	2020-10-27	23:49:59	Only 7 days left before the biggest US #ElectionDay...	['electionday', 'presidentialelection', 'american', 'vot...
35	2020-10-27	23:49:18	With just one week until #ElectionDay, over 2.75 mi...	['electionday']

Showing 1 to 36 of 423,829 entries, 4 total columns

Figure 3 shows a caption of how R the output returned by Twint, that include the date when each tweet was posted and its time, the tweet itself and the hashtags that each tweet contains.

Our method for collecting relevant tweets is adapted from Nisar and Yeung (2018), they follow previous literature to search in the trending topic list for hashtags that fit their analysis of UK elections 2016. Like them we use Trendogate[2], a web page that allows us to search the trending topics by country and date. Some of the main hashtags of the U.S. elections were: "#Trump2020", "#Biden", "#ElectionDay" or "#VOTE". For our analysis only "#ElectionDay" was taken; firstly, because it was the most used hashtag in the majority of the dates we were analyzing; secondly, because it is general enough to avoid sentiment bias towards specific parties. That way principal aim of measuring the general mood of the public in relation to the election will not be disturbed. Finally, because it gathers enough data to carry out our analysis.

Some of the limitations of using methods like these are that tweets that do not contain any hashtag will be ignored, or that hashtags that do not reach enough number of tweets to enter the trending topic list will also be ignored. Both of these issues might make us lose relevant data.

In addition to that, we have used "cld2" Google's Compact Language Detector 2[3] to filter our data set and get only tweets written in English. 83,7% of tweets from our data set are written in English and will be passed to Vader.

Next, we consider three tweets (Figure 4, Figure 5 and Figure 6) from our data set and present the score assigned by VADER and how they are classified as positive, negative or neutral regarding the following threshold values:

1. Positive sentiment: compound score ≥ 0.05.

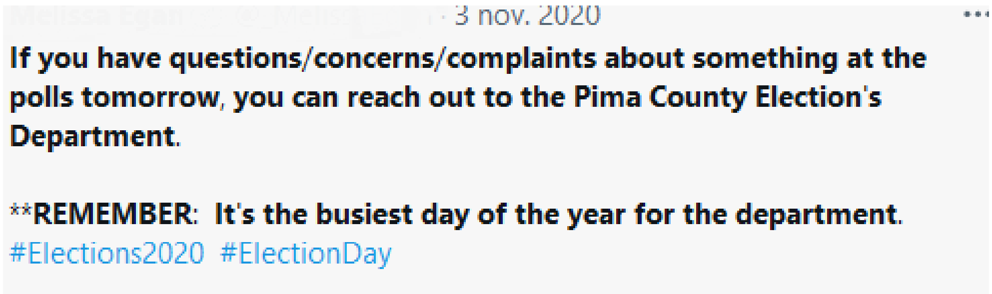
[2] This web page is not currently available.
[3] https://github.com/ropensci/cld2

Figure 4: Example of a tweet from our dataset with a positive compound value of 0.995.

Source: Twitter.

2. Neutral sentiment: -0.05 < compound < 0.05.

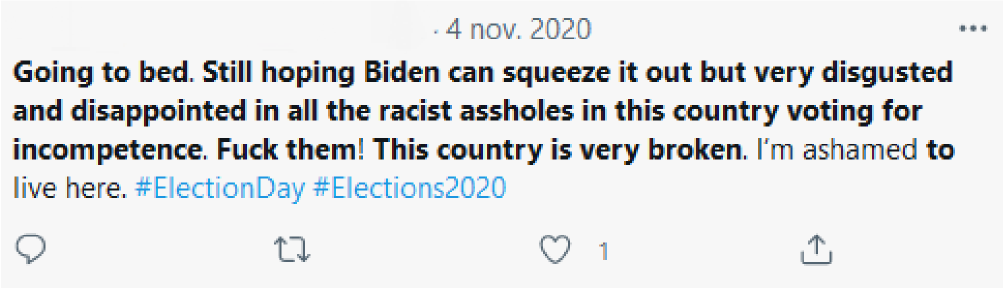
Figure 5: Example of a tweet from our dataset with a neutral compound value of 0.026.



Source: Twitter.

3. Negative sentiment: compound score ≤ -0.05.

Figure 6: Example of a tweet from our dataset with a neutral compound value of -0.992.



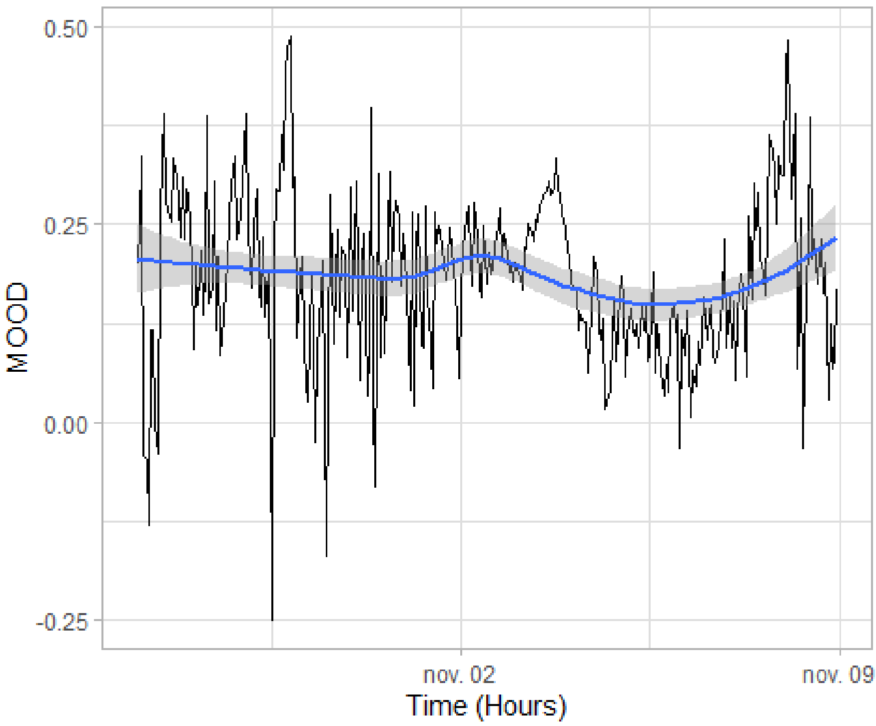
Source: Twitter.

Applying this threshold, we have divided our tweets into "positive", "neutral" and "negative", and, following (Nisar and Yeung 2018) we have calculated the independent variable "MOOD". Then we have given neutral tweets a value of 0, assuming that these tweets do not express any sentiment. They explain that this is not strictly true because neutral content can affect on the overall sentiment of the crowd, nevertheless the "MOOD" on a given day t is defined as follows:

$$MOOD_t = \frac{Vpos_t - Vneg_t}{Vtotal_t}$$

The value of "MOOD" moves between -1, completely negative, and 1, completely positive, addressing a sentiment to each day taking into account the computed sentiment of every tweet in a given day.

Figure 7: Hourly evolution of the variable "MOOD"



Source: Made with our data on R, "VADER", "tidytext" and "lubridate" packages

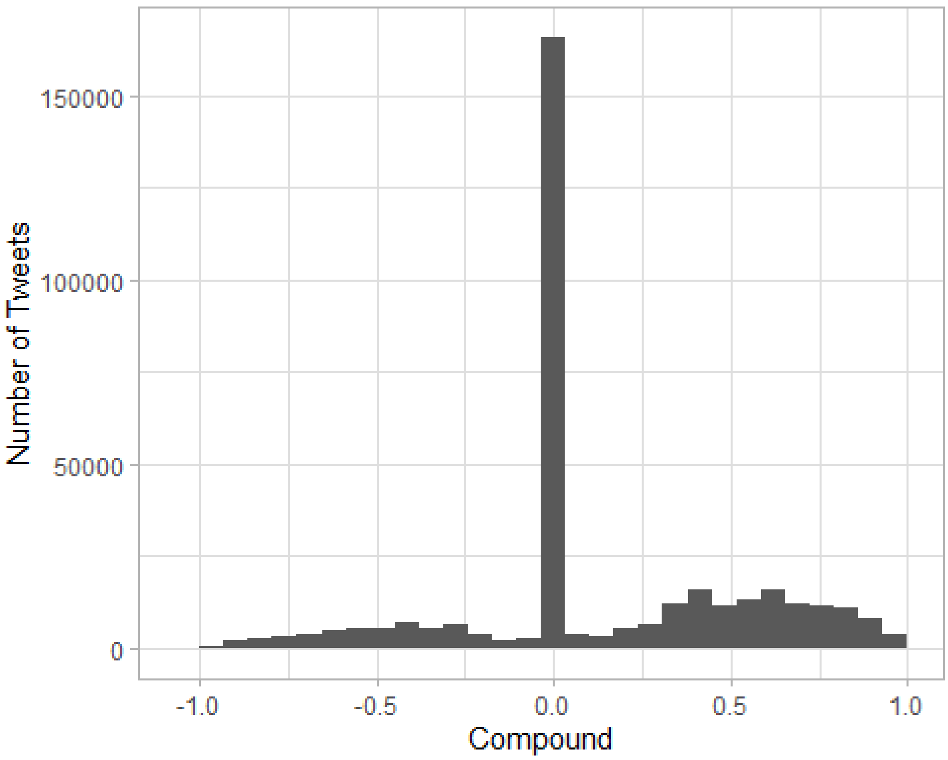
Figure 7 displays the hourly evolution of the variable "MOOD". We can see big mood fluctuations, that go from -0.25 to almost +0.5. Note that although we have hours when the mood is negative its absolute value never reaches the maximum value reached by the positive mood, that is, mood positivity seems to be stronger than mood negativity.

Table 1: Daily number of total, positive and negative tweets written in English.

	date	Vtweets	Vpos	Vneg
1	2020-10-27	4123	1885	795
2	2020-10-28	3234	1377	693
3	2020-10-29	2970	1388	563
4	2020-10-30	4299	1793	1132
5	2020-10-31	4342	1863	974
6	2020-11-01	6809	2724	1503
7	2020-11-02	23666	9684	4578
8	2020-11-03	171006	70286	23825
9	2020-11-04	99405	30557	16096
10	2020-11-05	9960	2890	1823
11	2020-11-06	11952	3336	2044
12	2020-11-07	9297	3852	1101
13	2020-11-08	3661	1459	456

Table 1 shows the daily distribution of total, positive, negative tweets written in English, with a total of 354,724 tweets. We see that the dates around the 3rd of November, the Election Day, have a higher volume of tweets, being the volume peak on the Election Day itself. Taking into account the threshold mentioned earlier we find out that from our data set 15,67% of the tweets are negative, 46,81% are neutral and 37,52% are positive.

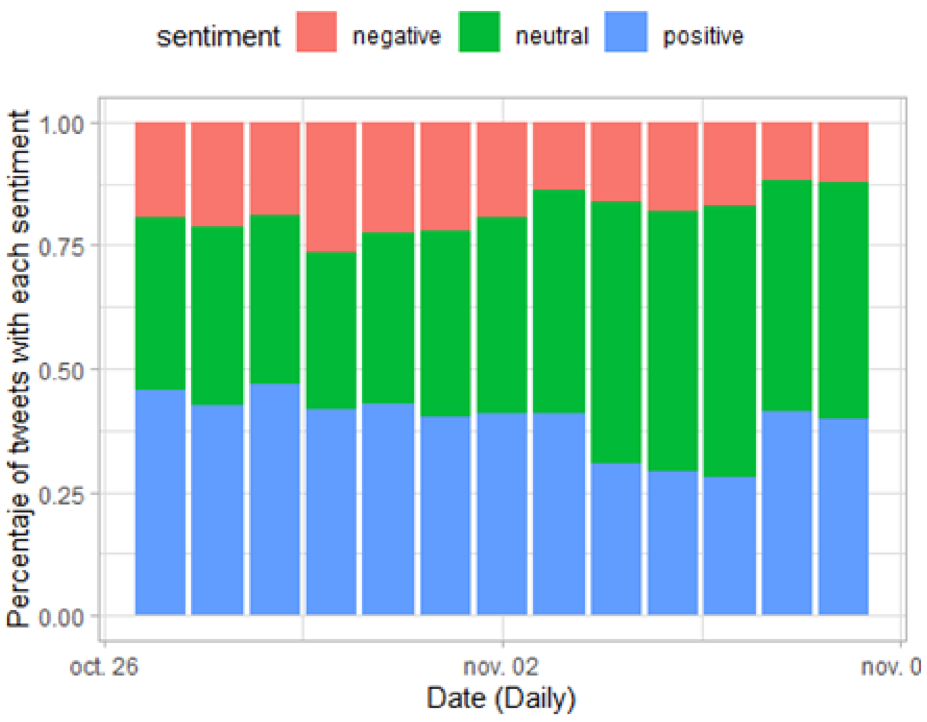
Figure 8: Distribution of all the tweets in English regarding its compound score



Source: made on R, "tidytext" and base packages.

Figure 8 shows the sentiment score distribution of all the tweets in English in our data set returned by "VADER". As we can see more than 150,000 tweets pile up on a compound score of 0.0. This could be, firstly, because of the bias that VADER has towards neutral sentiment, as it evaluates tweets that indeep express a feeling as neutral tweets, as we have seen in Figure 2; secondly, because those tweets could include characters that VADER does not understand, and finally, maybe because as they are political discussion tweets, a proper sentiment cannot be addressed. Indeed, we have find difficulties in trying to evaluate by ourselves the sentiment of some of the tweets, as the feeling that they express is not clear.

Figure 9: Evolution of the proportion of positive, negative and neutral tweets



Source: made with our data on R, "VADER", "tidytext", and base packages.

In Figure 9 we see the distribution of negative, neutral and positive tweets for each day. The proportion of positive tweets prior to the election (November 3rd) is higher than the days immediately after the election. In those days the proportion of neutral tweets widens, maybe this is because the general sentiment is more negative but VADER cannot compute it correctly, for example, Figure 10 shows a tweet from the list of tweets evaluated as neutral by VADER:

Figure 10: Example of a tweet from our dataset with a neutral compound value of 0.026.

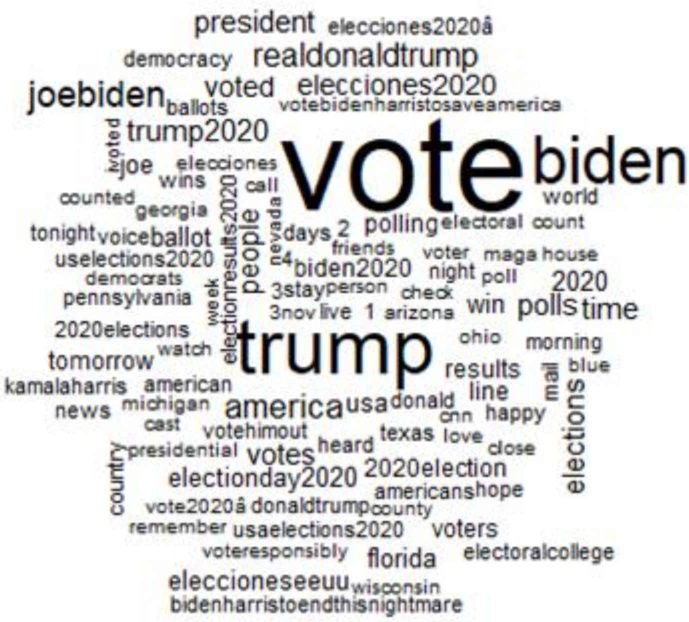
Source: Twitter.

This negative feeling towards the election day may be because of the uncertainty and negativity towards the results of the election. For example, the following tweet, Figure 11, from our data set expresses that feeling:

Figure 11: Example of a tweet from our dataset with a neutral compound value of 0.048.

Finally, the proportion of positive tweets increases again with a decrease in the proportion of negative tweets.

Figure 12: Wordcloud of the hundred most used words in our data set.



Source: Made with our data on R, “tidytext”, “wordcloud” and base packages

In order to show some uses of NLP apart from the sentiment analysis previously done, we have, for example, this wordcloud, in Figure 12. In general, a wordcloud is useful to see with a glimpse of an eye the general sentiment of whatever we want to analyze, a product, a film or even politics as it is our case.

The cloud contains the hundred most used words in our data set, the bigger the size the more repeated the word is, and vice versa. We are not surprised by the results, the most used word in our data set is “vote”, this expresses well the general sentiment around this election where people highlighted the importance of going to vote. Indeed, lots of the tweets that were evaluated as neutral by VADER express this feeling of unease. Just a few examples of this are showed in Figure 13, Figure 14, Figure 15 and Figure 16:

Figure 13: Example of a tweet from our dataset with a neutral compound value of 0.049.

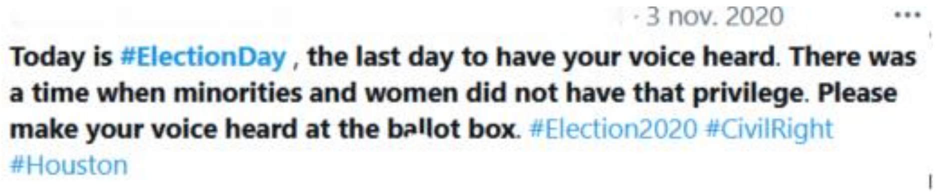
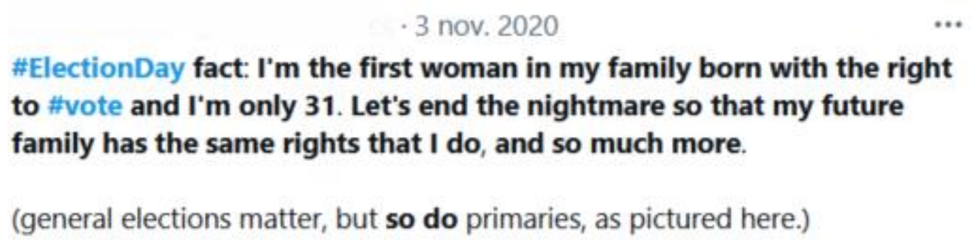
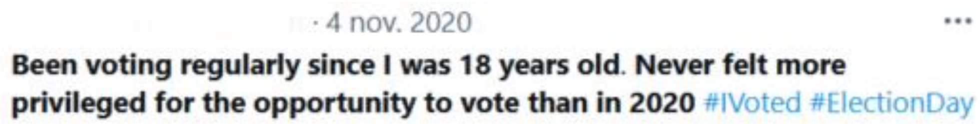


Figure 14: Example of a tweet from our dataset with a neutral compound value of 0.047.



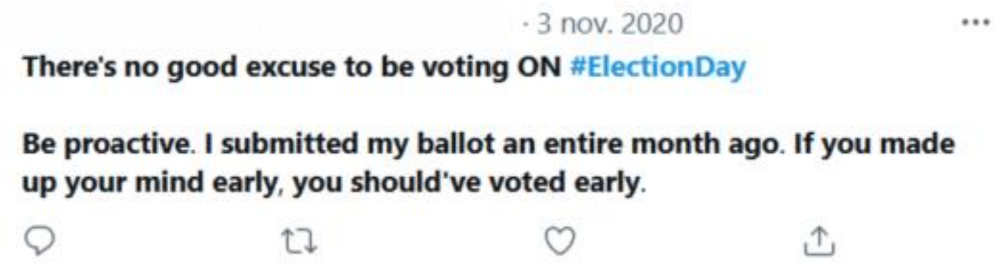
Source: Twitter.

Figure 15: Example of a tweet from our dataset with a neutral compound value of 0.046.



Source: Twitter.

Figure 16: Example of a tweet from our dataset with a neutral compound value of 0.044.

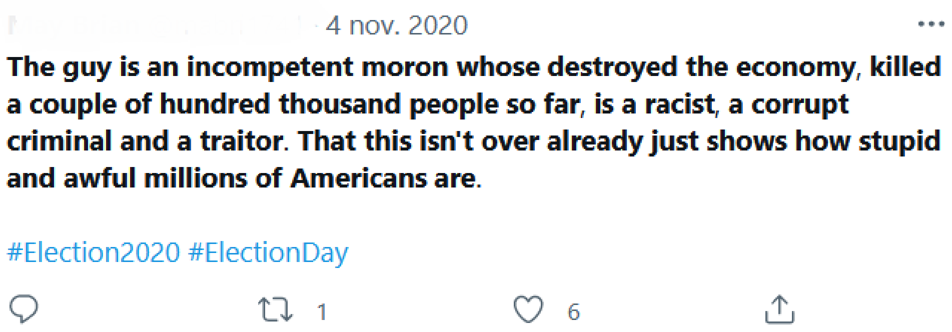


Source: Twitter.

Regarding the size of the names of the two candidates, both are almost equal, showing that they were mentioned a similar number of times. Paying attention to the small sized words, we can see some interesting mentions such as the names of some important states that could be important to determine the winner of the presidency, or expressions such us “votehimout” or “maga”, make America great again.

“votehimout”, it refers to vote Donald Trump out, who was governing the U.S. at the time of the elections. There are plenty of tweets that show the citizens’ concern about him being a bad president and a problem for the U.S society, as we can see with the following examples in Figure 17, Figure 18 and Figure 19:

Figure 17: Example of a tweet from our dataset with a compound value of -0.981.


· 4 nov. 2020

The guy is an incompetent moron whose destroyed the economy, killed a couple of hundred thousand people so far, is a racist, a corrupt criminal and a traitor. That this isn't over already just shows how stupid and awful millions of Americans are.

#Election2020 #ElectionDay

1 6

Source: Twitter.

Figure 18: Example of a tweet from our dataset with a compound value of -0.05.


· 3 nov. 2020

Dear American people,

With immediate, please stop voting absolute fucking lunatics in to the White House.

Thank-you,

The Rest of the World

#ElectionDay #Elections2020 #Elections #Vote #BidenHarris2020

2

Source: Twitter.

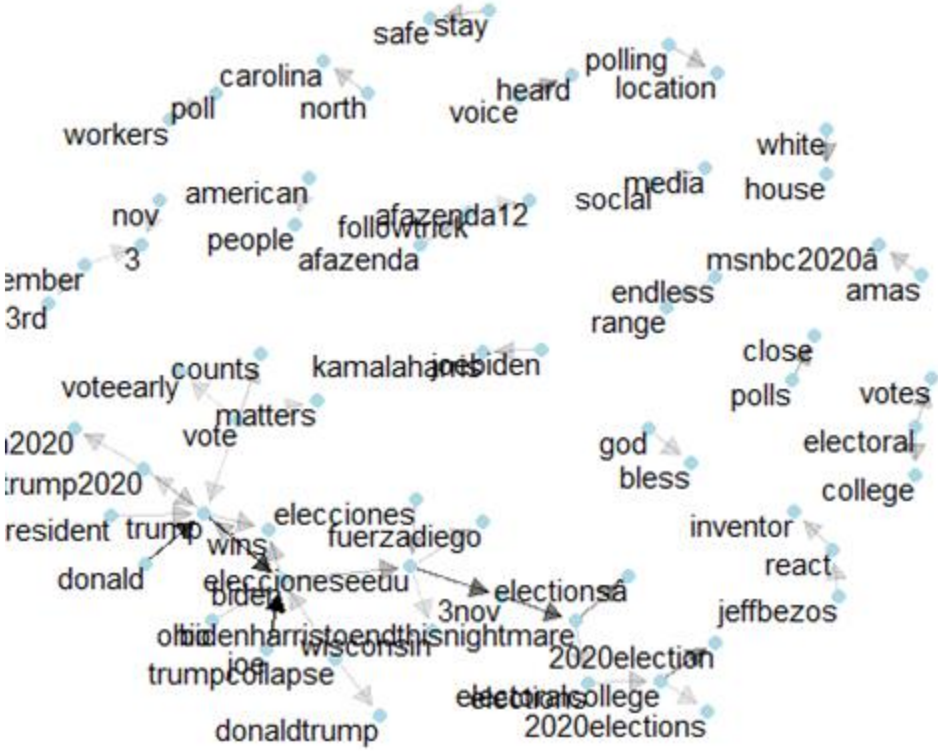
Figure 19: Example of a tweet from our dataset with a compound value of -0.05.


Obertson @Obertson · 27 oct. 2020

No Health Plan just a lie again. Trump has No plan for health care. #debates #Debate2020 #debate #ElectionInterference #ElectionDay #VOTE #VoteBlue #PresidentialDebate2020 #government #GOPHypocrisy #COVID19 #COVID__19 #COVID #COVIDIOTS #Corona #coronavirus #virus #TrumpHasCovid

Source: Twitter.

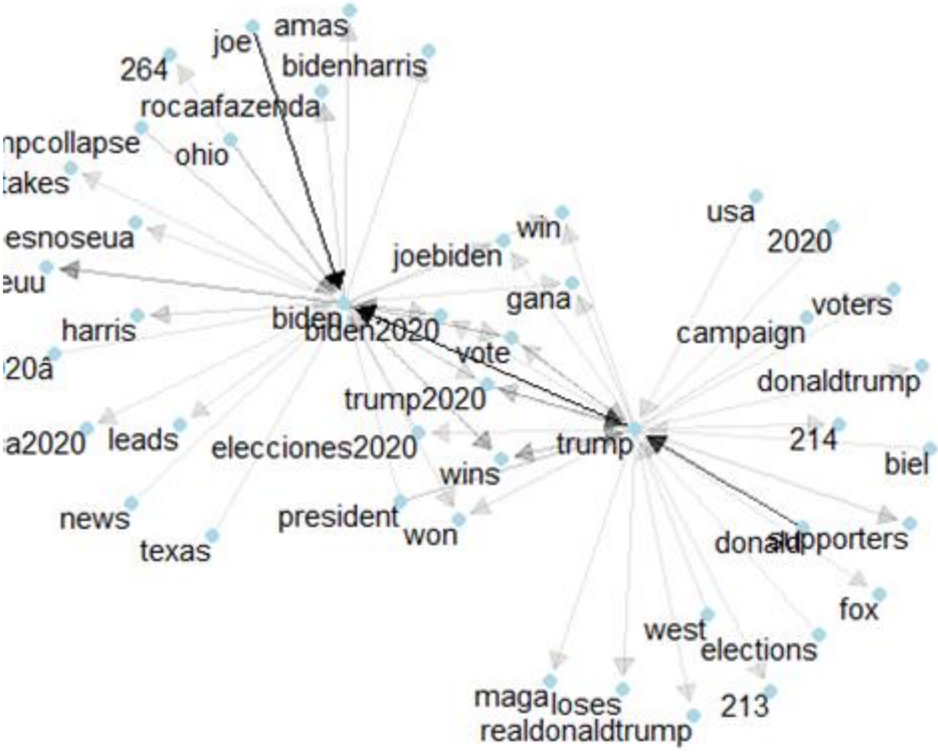
Figure 20: Graph with the most repeated bigrams in our data



Source: Made with our data on R “tidytext”, “ggraph” packages

Figure 20, shows another interesting tool of NLP: bigrams. We can divide our data set into tokens that contain every combination of two consecutive words and then represent them linked by arrows, showing the order in which, the words are used. In addition, the transparency of the arrow determines how common or rare the combination of two words is, becoming darker as the combination is more common. In our first example, we can highlight expressions such as “voice heard” or “polling location” that express the previously mentioned general sentiment of the importance of voting, the feeling of unease is clearly expressed by the bigram “stay safe”. But apart from that we do not get really relevant information.

Figure 21: Graph of the bigrams related to the election candidate's name



Source: Made with our data on R, "tidytext" and "ggraph" packages.

Figure 21 shows a graph of the bigrams of the names of both candidates. We see interesting things; words related to winning the election are in between the two names because they are related with both candidates. We see the word "gana" that shows the importance of the spanish community for both candidates. The numbers we see coming out of the names of the candidates are related to the number of delegates won by each candidate on the preliminaries, being 264 for Biden and 214 for Trump.

Interestingly, in this graph only appears the name of one state, Ohio, and it is linked to Biden. However, it was Trump who won this state. This could be because since 1864 only four candidates became presidents without winning Ohio, and it is thought that the candidate who wins the state of Ohio becomes president[4]. It was not the case.

[4]https://www.elconfidencial.com/mundo/2020-11-08/ohio-termometro-elecciones-eeuu-resultados_2818379/

4.2. Index data collection

For this analysis we have chosen to use the S&P 500, the Dow Jones Industrial Average and the Nasdaq Composite. We have chosen these three indexes because they are the most followed stock-market indexes in the U.S. and they are a good reference of the whole U.S. stock-market.

The data needed of the S&P 500[5], Nasdaq Composite[6] and DJIA[7] was obtained from yahoo! Finance. From this website we have downloaded the opening and closing prices of the three indexes as well as the volume of daily trades.

The stock market does not operate on weekends so there is no data for the 31th of October, 1st, 7th and 8th of November. As we have Twitter data for the weekends, we want to have stock-market data comparable with the one corresponding with the selected tweets. So, in order to solve this issue, we have followed a method to approximate those non existing values. Firstly, the opening price of Mondays is also used as the closing price of Sundays. The closing price of Saturdays is proxied as the mean value of the Friday and Sunday closing prices. This is a similar way of approaching this non existing values, in which, the "value on a given day is x and the next available data point is y with n days missing in between, we approximate the missing data by estimating the first day after x to be (y+x)/2" (Goel and Mittal 2012). This is because stock data usually follows a concave relationship, unless an incident of big impact occurs.

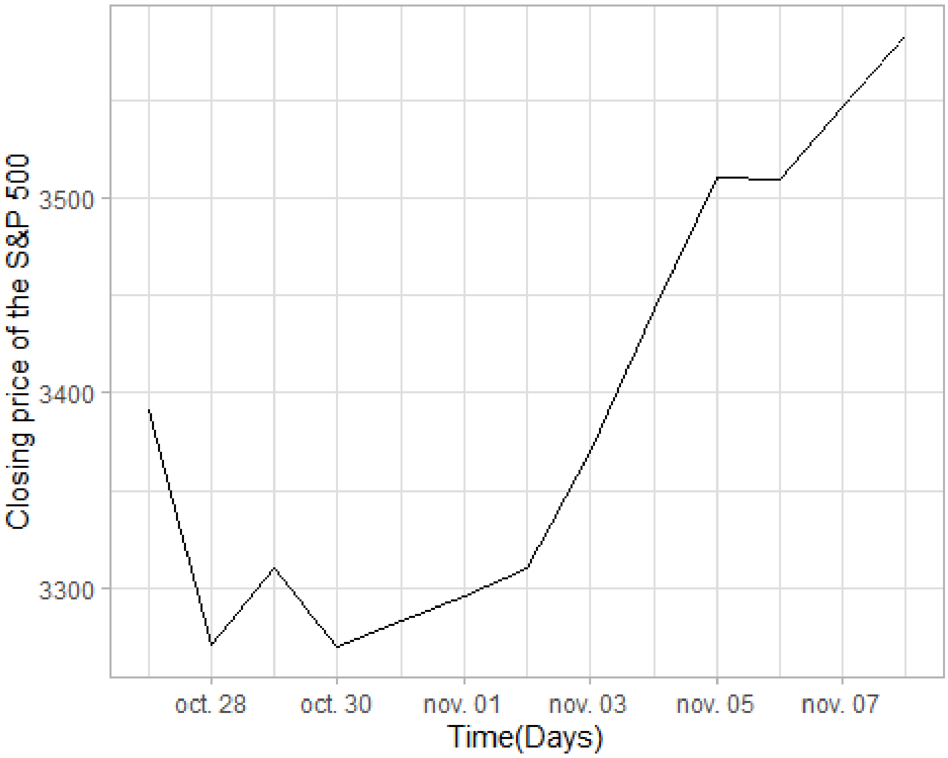
The S&P 500 Index is regarded to be the best measure of the equities from the leading industries of the U.S. economy. This is because it is an index that represents the 500 largest publicly-traded companies in the U.S. This index uses a market capitalization weighting method, giving a higher percentage of weight in the index to companies with the largest market capitalization (Kenton 2021). The 5 largest components of this index are, Apple, Microsoft, Amazon, Facebook and Alphabet (Alpert 2021).

[5] https://es.finance.yahoo.com/quote/%5EGSPC/history?period1=1592853500&period2=1624389500&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

[6] https://es.finance.yahoo.com/quote/%5EIXIC/history?period1=1603670400&period2=1605312000&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

[7] https://es.finance.yahoo.com/quote/%5EDJI/history?period1=1603670400&period2=1605312000&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

Figure 22: Daily closing price of the S&P 500

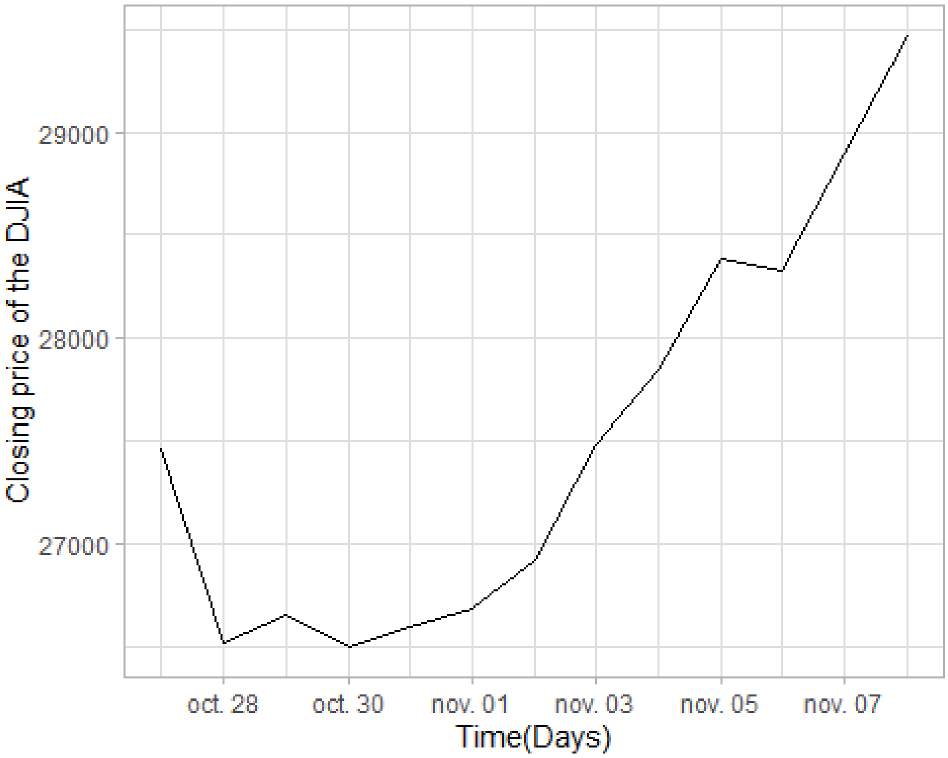


Source: Made with our data on R, “tidytext” package.

In Figure 22 we see the daily closing price of the S&P 500. As we can see on November 27th the price of the index is around 3,400 points, and it falls to around 3,300 points between October 28th and October 30th. Since then it starts to grow. The 2nd of November it grows fast until November 5th, it remains flat one day and starts growing again until November 8th when the price of the index is above 3,500 points. We will see a similar pattern in all the indexes.

The Dow Jones Industrial Average (DJIA) is the second oldest U.S. market index created in 1896 to serve as a measure of the health of the U.S. economy. This index tracks the 30 largest and publicly-traded blue-chip companies on the New York Stock Exchange and the NASDAQ. Blue-Chip companies are recognized to be well established and financially healthy companies that sell high quality and widely accepted products and services. These companies, are expected have a long record and a stable and reliable growth. The top 5 companies in the index are: 3M, American Express, Amgen, Apple and Boeing (Ganti 2020).

Figure 23: Daily closing price of the DJIA

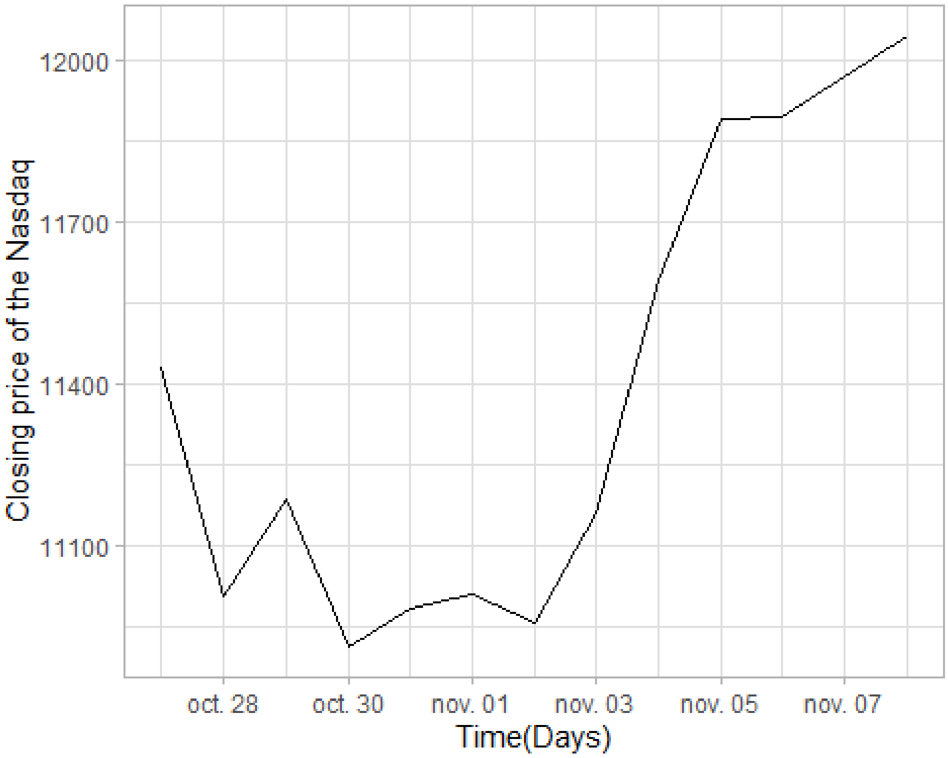


Source: Made with our data on R, “tidytext” package.

In Figure 23 we have the daily closing price of the DJIA, which moves around 26,500 and 29,500 points. As we have previously mentioned the movement is really similar the S&P 500.

The Nasdaq Composite Index is an index composed of over 2,500 common equities listed on the Nasdaq stock exchange. The Nasdaq stock exchange is a global electronic marketplace for buying and trading securities. It was the first ever electronic exchange (Hayes, 2021). Most of the world's leading technology giants are listed on it, and the index is known as a technology index because this sector represents nearly a 50% of the index’s weight (Chen, 2021). Companies from non-financial industries are listed and represented on the Nasdaq, such as retail, biotechnology, telecommunications or industrial. The weighting of the index is constructed by a capitalization method being the top 5 companies as of March 31 of 2021: Apple, Microsoft, Amazon, Tesla, Facebook “A” (Norris 2021).

Figure 24: Daily closing price of the Nasdaq Composite



Source: Made with our data on R, “tidytext” package.

Figure 24 shows the daily closing price of the Nasdaq, again, we see a similar movement to the previous two indexes but in this case it moves around 10,950 and 12,000 points.

The stock market does not operate on weekends so there is no data for the 31th of October, 1st, 7th and 8th of November. As we have Twitter data for the weekends, we want to have stock-market data comparable with the one corresponding with the selected tweets. So, in order to solve this issue, we have followed a method to approximate those non existing values. Firstly, the opening price of Mondays is also used as the closing price of Sundays. The closing price of Saturdays is proxied as the mean value of the Friday and Sunday closing prices. This is a similar way of approaching this non existing values, in which, the “value on a given day is x and the next available data point is y with n days missing in between, we approximate the missing data by estimating the first day after x to be (y+x)/2” (Goel and Mittal 2012). This is because stock data usually follows a concave relationship, unless an incident of big impact occurs.

5. Hypotheses

In this section the hypotheses that we are testing are going to be explained. We used multiple regression models we to test the relationship between political discussion sentiment and volume of tweets posted and the stock market movements. To do so we will test the following hypotheses:

First hypothesis:

> $H0_1$ (null hypothesis): There is no statistical relationship between the closing price/volume of stock index and the daily volume of positive, negative and total number of tweets.
> Ha_1 (alternative hypothesis): There is a non-zero relationship between the closing price/volume of stock index and the daily volume of positive, negative and total number of tweets.

Second hypothesis:

> $H0_2$ (null hypothesis): There is no statistical relationship between the closing price/change in the price of each index and the daily mood of Twitter.
> Ha_2 (alternative hypothesis): There is non-zero relationship between the closing price/change in the price of each index and the daily mood of Twitter.

6. Results

In this section we will go through the results of our analyses. Firstly, we will focus on the results regarding to the analysis of our first null hypothesis, which holds that there is no statistical relationship between the closing price/volume of stock index and the daily volume of positive, negative and total number of tweets. Then, we will expand on the results regarding to our second null hypothesis, which holds that there is no statistical relationship between the closing price/change in the price of each index and the daily mood of Twitter.

Table 2: Correlation results for volume and close of the three indexes, "Vtweets", "Vpos", and "Vneg"

	S&P 500		**DJIA**		**NASDAQ**	
	Volume	Close	Volume	Close	Volume	Close
Vtweets	-0.34	0.3	-0.25	0.05	-0.29	-0.7
Confidene interval	[-0.82,0.42]	[-0.53, 0.57]	[-0.78, 0.50]	[-0.52, 0.58]	[-0.80, 0.47]	[-0.60, 0.50]
Vpos	-0.38	-0.1	-0.28	0.03	-0.33	-0.11
Confidence interval	[-0.83,0.38]	[-0.55, 0.55]	[-0.80, 0.47]	[-0.53, 0.57]	[-0.82, 0.43]	[-0.62, 0.47]
Vneg	-0.33	-0.1	-0.23	0.03	-0.28	-0.08
confidence interval	[-0.81,0.43]	[-0.54, 0.56]	[-0.77, 0.51]	[-0.53, 0.57]	[-0.80, 0.47]	[-0.60, 0.49]

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Values in square brackets indicate the 95% confidence interval for each correlation.

Table 2 shows the correlation coefficients for the daily volume (first column below each index) and the daily closing price (second column below each index) of each index and the independent variables "Vtweets", "Vpos" and "Vneg". With this, we want to test if our independent variables individually are linearly correlated with the dependent variables. We do not get any of the results to be statistically significantly different from zero at a 10% level of confidence. Next, we are going to test if a multiple regression model fits our data.

Table 3: Regression results for the Volume and close of the three indexes

	S&P 500		**DJIA**		**NASDAQ**	
	Volume	Close	Volume	Close	Volume	Close
(Intercept)	-0.2642	-0.3697	-0.1738	-0,418*	-0.3086	-0.2194
std. Error	0.1265	0.2288	0.2789	0.2179	0.1928	0.2427
Vtweets	3.7094	19,4817**	-0.7952	16,8894**	6.7418	25,2738**
std. Error	3.9169	7.4735	8.6355	7.1169	5.9702	7.9264
Vpos	-1.7948	-7,917**	-0.4875	-6,6148*	-3.0946	-10,4722**
std. Error	1.5451	3.111	3.4064	7.1169	2.355	3.2995
Vneg	-2.0444	-11,6529**	1.1466	-10,3281**	-3.8303	-15,0255**
std. Error	2.5463	4.7579	5.6139	4.5309	3.8812	5.0462
R Squared	0.3175	0.4342	0.1339	0.3857	0.3244	0.5413
Adj. R Squared	-0.09208	0.2456	-0.3857	0.181	-0.08092	0.3884
F-statistic	0.7752	2.302	0.2577	1,884	0.8004	3,54*
Num. Obs.	9	13	9	13	9	13

*** p < 0.01; ** p < 0.05; * p < 0.1

In Table 3 we have the results of the regression models of the daily closing price and daily volume of each index, with the independent variables total number of daily tweets (Vtweets), the daily number of positive tweet (Vpos) and the daily number of negative tweets (Vneg). As we can see in the last row, our analysis are limited due to the small number of observations we have. However, this is an issue we cannot solve because it is the nature of our analysis, as people's opinion towards the election is limited by the duration of the event.

Firstly, we are going to look at the analysis which has the volume of the index as the dependent variable. From this analysis we get that none of the dependent variables' coefficients results are statistically significantly different from zero at a 5% confidence level. Then, we have R squared values of 0.3175, 0.1339 and 0.3244, but the adjusted R squared of these models are negative: -0.09208, -0.3857 and -0.08092 indicating that the adding of these variables does not improve our model. Finally, the null hypothesis in the ANOVA analysis is that the best approximation to the experimental points is the mean value of the dependent variable. In our case we are trying to see if the multiple linear regression model fits better than the mean of the dependent variable. If the null hypothesis is rejected our model would be a good fit for our data. Nevertheless, checking the F statistical value, 0.7752, 0.2577 and 0.8004, we see that they are too low to be statistically significantly different from zero at a 5% confidence level. So, the null hypothesis cannot be rejected, meaning that a linear model is not a good fit for the data.

(Nisar and Yeung 2018) do not find either results that are statistically significantly different from zero at a 10% level of confidence.

Secondly, we are going to look at the results of the models that have the daily closing price of each index as the dependent variables. In this case, we can see that the number of observations has increased from 9 to 13 which is still a low number. Nevertheless, the results seem to slightly improve. The coefficients' results of the independent variables for the three indexes are statistically significantly different from zero at a 5% confidence level, except "Vpos" for the DJIA which is significant at a 10% confidence level. The coefficients for the variable "Vtweets" are positive for the three indexes, indicating that a larger volume of tweets is associated with higher closing prices. Whereas the coefficients for the variables of "Vpos" and "Vneg" are negative for the three indexes, meaning that a larger volume of positive or negative tweets is associated with lower closing prices. Regarding the R squared we get 0.4342, 0.3857 and 0.5413 for the S&P 500, DJIA and Nasdaq, respectively. Then, the adjusted R squared we get 0.2456, 0.181 and 0.3884 which are still low. Finally, looking at the F-statistic we do not get results that are statistically significantly different from zero at a 5% confidence level, meaning that a linear model is not a good fit for the data. However, the F-statistic for the Nasdaq is statistically significantly different from zero at a 10% confidence level.

Again, we find the same results as (Nisar and Yeung 2018) as regard the relationship between political discussion extracted from Twitter and stock-market movements.

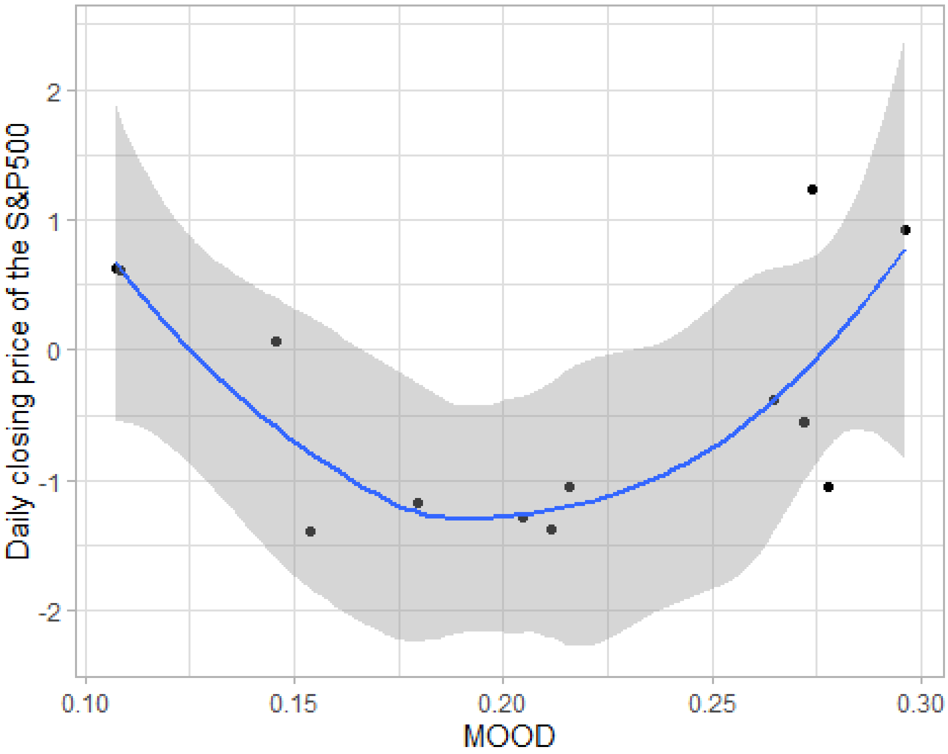
Next, we are going to talk about the results of the relationship of the variable "MOOD" and the daily closing price of each index. In Table 4 we see the correlation matrix of each variable. The results we get suggest that our variables are not linearly related. All the correlations are low and they cannot be considered statistically significantly different from zero at a 10% level of confidence.

Table 4: Correlation results of the daily closing price of the three indexes and "MOOD" and "MOOD" squared

	S&P 500 Close	**DJIA** Close	**NASDAQ** Close
Mood	0.01	0.09	-0.04
Confidence interval	[-0.55, 0.56]	[-0.49, 0.61]	[-0.58, 0.53]
Mood Squared	0.11	0.18	0.07
Confidence Interval	[-0.47, 0.62]	[-0.41, 0.67]	[-0.50, 0.60]

*** p < 0.01; ** p < 0.05; * p < 0.1. Values in square brackets indicate the 95% confidence interval for each correlation.

Figure 25: Scatterplot of the daily closing price of the S&P 500 against "MOOD"



Source: Made with our data on R, “tidytext” and “tidyverse” package.

Figure 25 shows that the daily closing price of the S&P 500 and the variable “MOOD” follow a quadratic relationship (We get the same pattern for the other two indexes as well[8]). Figure 25 suggests that we should add the square of the variable “MOOD” in the multiple regression model estimated to explore the relationship between “MOOD” and closing prices in the stock market.

[8] See Figure 29 and Figure 30 in the appendix.

Table 5: Regression results for the daily closing price of each index, "MOOD" and "MOOD" squared

	S&P 500	**DJIA**	**NASDAQ**
	Close	Close	Close
(Intercept)	7,721***	6,238**	9,832***
std. Error	2.039	2.053	2.228
Mood	-89.651***	-75.144***	-110.406***
std. Error	21.732	21.877	23.748
Mood Squared	223.284***	189.796***	273.139***
std. Error	53.623	53.979	23.748
R Squared	0.6342	0.5563	0.6852
Adj. R Squared	0.5611	0.4675	0.6223
F-statistic	8.67***	6.268**	10.88***
Num. Obs.	13	13	13

*** p < 0.01; ** p < 0.05; * p < 0.1

In Table 5 we see the results of the model for the three indexes, in it we can see that the coefficients' results of the two variables and the intercept are statistically significantly different from zero at a 1% level of confidence. Then, regarding the R Squared, we get 0.6342, 0.5563, and 0.6852, indicating that 63.42%, 55.63% and 68.52% of the changes in the daily closing price of the S&P 500, DJIA and Nasdaq, respectively, are explained by changes in the variable "MOOD" and it's square. Looking at the adjusted R squared we get values of 0.5611, 0.4675 and 0.6223 indicating that our variables provide new information. Finally, we get that the F-statistical is statistically significantly different from zero at a 1% level of confidence for the S&P 500 and the Nasdaq, and at a 5% level of confidence for the DJIA, indicating that our data follows a quadratic relationship. At first when the "MOOD" increases the closing price will go down, but after a minimum the better the "MOOD" on twitter the higher the closing price of the stock market. With these results we can reject the null hypothesis and conclude that there is correlation between the closing price of each index and the daily "MOOD" of Twitter.

Our results provide a stronger support for the hypothesis tested that those in Nisar and Yeung (2018) who although do not get statistically significant results, they claim to be really close to do so, and they conclude that there is a relationship between these two variables.

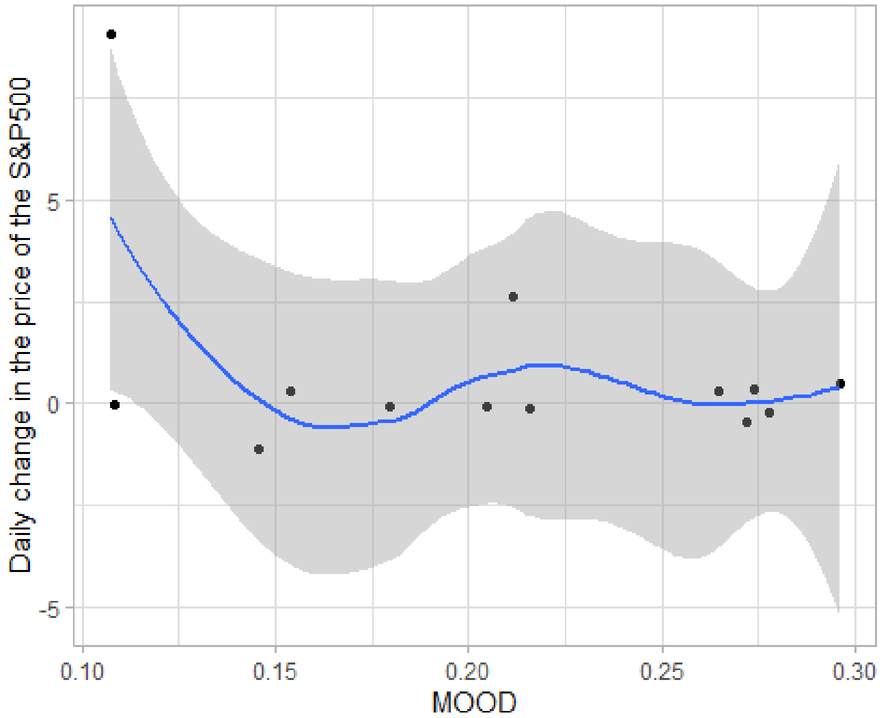
Our last analysis tests the relationship between the daily change in the price of the three indexes and the variable "MOOD". In Table 6 correlations are presented, and interestingly, we see differences between the indexes. The relationship between these two variables for the S&P 500 and the Nasdaq seems to be negative, nevertheless we cannot get clear results as the relationship between our variables is not statistically significantly different from zero at a 10% level of confidence. In contrast, the daily change in price of the DJIA and the general "MOOD" on Twitter are positively correlated and are statistically significantly different from zero at a 5% level of confidence. To see in a more visual way why the results are different for the DJIA index, we are going to look at the scatterplots of the variable "MOOD" and the daily change in price of each index.

Table 6: Correlation matrix between the daily change in the closing price of each index and “MOOD”

	S&P 500 Change	**DJIA** Change	**NASDAQ** Change
Mood	-0.4	0.6**	-0.11
Confidence interval	[-0.78, 0.20]	[0.07, 0.86]	[-0.62, 0.47]

*** p < 0.01; ** p < 0.05; * p < 0.1. Values in square brackets indicate the 95% confidence interval for each correlation.

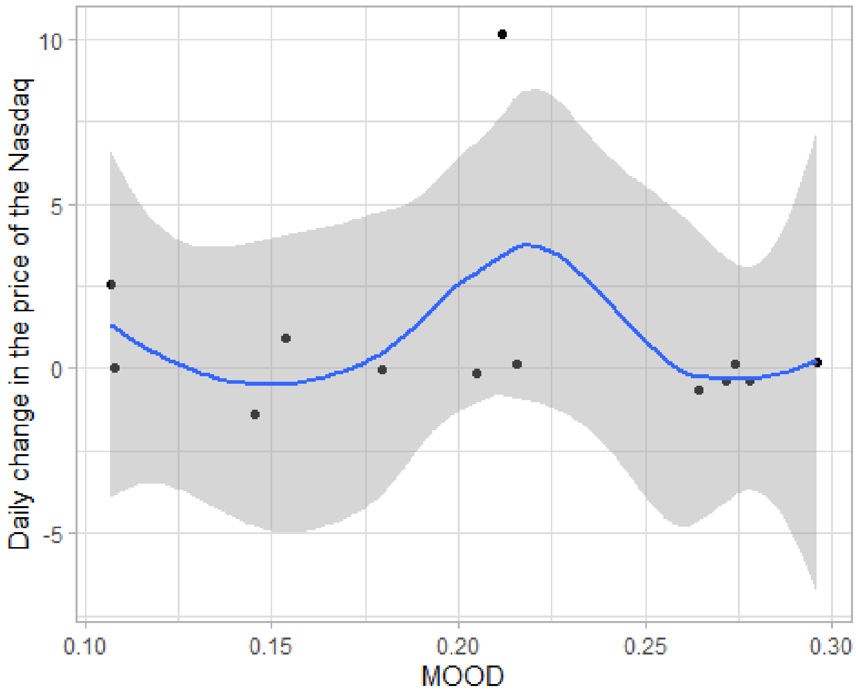
Figure 26: Scatterplot between the daily change in the S&P 500 and “MOOD”



Source: Made with our data on R, “tidytext” package.

Figure 26 shows the relationship between the daily change in the S&P 500 and the general mood on Twitter. For low values of “MOOD” it seems to have a decreasing tendency but it flattens fast, clearly not following a linear relationship. This may be because the relationship in fact is horizontal and the first point disturbs that flat relationship. Nevertheless, we do not have enough data, just 13 observations, to figure out the how relationship looks like.

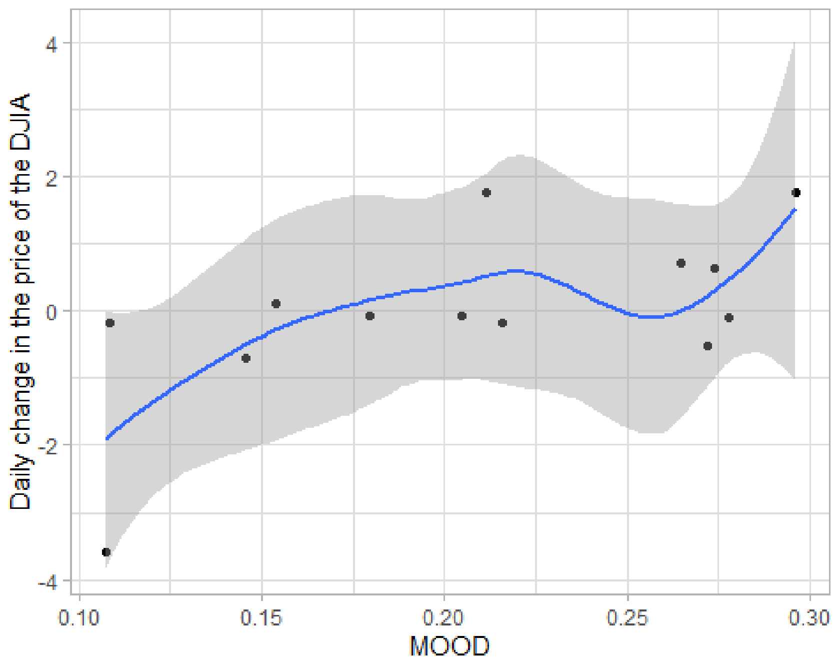
Figure 27: Scatterplot between the daily change in the Nasdaq and "MOOD"



Source: Made with our data on R, "tidytext" package.

Figure 27 shows the relationship between the general mood on Twitter and the daily change on the price of the Nasdaq. From this we cannot say a lot, because like in the previous one we do not have enough data, nevertheless we can see that with the data we have it is not a linear relationship.

Figure 28: Scatterplot between the daily change in the DJIA and "MOOD"



Source: Made with our data on R, "tidytext" package.

Figure 28 shows the relationship between general Twitter "MOOD" and the daily change of the price of the DJIA. In it we can see that the relationship between the daily change in the price of the DJIA and the daily "MOOD" is slightly positive correlated, as a larger value of the "MOOD" means a larger value of the daily change in the price of the DJIA.

Table 7: Regression results for the daily change in price of each index and "MOOD"

	S&P 500	**DJIA**	**NASDAQ**
	Change	Change	Change
(Intercept)	4.123	-2.515**	1.869
std. Error	2.389	1.053	2.916
Mood	-15.703	11.922**	-4.847
std. Error	10.969	4.835	13.389
R Squared	0.157	0.356	0.01177
Adj. R Squared	0.08041	0.2974	-0.07807
F-statistic	2.049	6.08**	0.131
Num. Obs.	13	13	13

*** p < 0,01; ** p < 0,05; * p < 0,1

In Table 7 we present the results for the regression analysis for these variables. The results for the S&P 500 and for the Nasdaq show that our model is not a good fit for the data. But for the DJIA, the values of the coefficients of the intercept and the variable "MOOD" are statistically different from 0 at a 5% level of confidence. The positive coefficient for the "MOOD" variable indicates the normalized daily change on the DJIA will increase 11.922 units, per unit increased in the normalized "MOOD" of Twitter. Finally, looking at the R squared we get that 35.6% of the variance of the daily change of the price of the DJIA are explained by the mood on Twitter. Finally, looking at the F-statistic of 6.08, it is statistically significantly different from zero at a 5% level of confidence, indicating that our model is a better fit for the data than the mean value of the daily change in the price of the three indexes.

(Nisar and Yeung 2018) Claim that a linear model is not a good fit for their data and that they cannot make clear conclusions regarding this model.

7. Conclusions

The aim of this work is to analyze if there is a relationship between general political sentiment extracted from Twitter and stock market movements, so a short window event study of the U.S. 2020 elections was made. To do so, we tested if there is a correlation between twitter sentiment and volume and stock market movements.

With the results of the models that had the volume and the daily closing price of each index as dependent variables, and "Vtweets", "Vpos" and "Vneg" as independent variables, we cannot reject the hull hypothesis that holds that there is no relationship between these variables. Our second analysis, which had the daily closing price of the three indexes as dependent variables and the general mood of Twitter as independent variable, validated that there is a quadratic relationship between these variables. Finally, the model that had as the dependent variable the daily change in the price of the three indexes, and as the independent variable the general mood on Twitter, has shown that is not a good fit for the data of the S&P 500 and the Nasdaq, but that our regression model fits better our data for the daily change in the price of DJIA better than the mean value of the daily change in the price of DJIA.

In general, we cannot reject the null hypothesis of no correlation due to the small amount of data that we were using, although we used around 400,000 tweets, much more than the 60,000 used in (Nisar and Yeung 2018), the small window of time we have used to analyze the political influence makes the sample too small. Another issue happens because of the topic we are analyzing, politics. As we have mentioned earlier, sentiment analysis tools have issues addressing sentiment to political topics, in our sample 46.81% of tweets are recorded as neutral, maybe this percentage would be smaller helping us make a better analysis.

8. References

Alpert, Gabe. 2021. "Top 10 S&P 500 Stocks by Index Weight." *investopedia*. https://www.investopedia.com/top-10-s-and-p-500-stocks-by-index-weight-4843111 (June 13, 2021).

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2(1): 1–8.

Burggraf, Tobias, Ralf Fendel, and Toan Luu Duc Huynh. 2020. "Political News and Stock Prices: Evidence from Trump's Trade War." *Applied Economics Letters* 27(18): 1485–88. https://doi.org/10.1080/13504851.2019.1690626.

Chen, James. 2021. "NASDAQ Composite Index." *Investopedia*. https://www.investopedia.com/terms/n/nasdaqcompositeindex.asp (June 14, 2021).

Clement, J. 2021. "• Global Top Websites by Monthly Visits 2020 | Statista." *2021*. https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/ (June 13, 2021).

Ganti, Akhilesh. 2020. "Dow Jones Industrial Average (DJIA)." *Investopedia*. https://www.investopedia.com/terms/d/djia.asp (June 13, 2021).

Goel, Arpit, and Anshul Mittal. 2012. "Stock Prediction Using Twitter Sentiment Analysis. Standford University, CS229." *Cs229.Stanford.Edu* (December): 1–5. http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf.

Grosz, Barbara J. 1982. "Natural Language Processing." *Artificial Intelligence* 19(2): 131–36.

Hayes, Adam. 2021. "Nasdaq Definition." *Investopedia*. https://www.investopedia.com/terms/n/nasdaq.asp (June 13, 2021).

Hutto, C. J., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* (May): 216–25.

Kenton, Will. 2021. "S&P 500 Index – Standard & Poor's 500 Index." *investopedia*. https://www.investopedia.com/terms/s/sp500.asp#sp-500-vs-djia (June 13, 2021).

Liu, B. (2012). Sentiment Analysis *Sentiment Analysis and Opinion Mining Opinion Mining*. Morgan & Claypool.

Mao, Yuexin, Wei Wei, Bing Wang, and Benyuan Liu. 2012. "Correlating S&P 500 Stocks with Twitter Data." *Proceedings of the 1st ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, HotSocial 2012*: 69–72.

Nisar, Tahir M., and Man Yeung. 2018. "Twitter as a Tool for Forecasting Stock Market Movements: A Short-Window Event Study." *Journal of Finance and Data Science* 4(2): 101–19. https://doi.org/10.1016/j.jfds.2017.11.002.

Norris, Emily. 2021. "QQQ Stock Trading Risks and Rewards." *investopedia*. https://www.investopedia.com/ask/answers/061715/what-qqq-etf.asp#citation-5 (June 13, 2021).

Rao, Tushar, and Saket Srivastava. 2012. "Analyzing Stock Market Movements Using Twitter Sentiment Analysis." *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*: 119–23. http://dl.acm.org/citation.cfm?id=2456719.2456923.

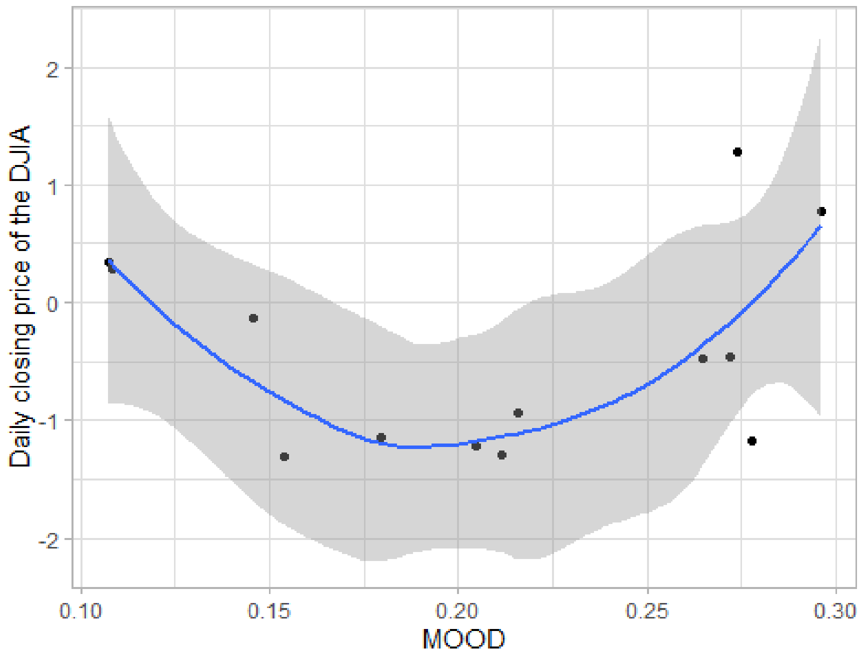
Silgo, J. & Robinson, D. (2017). Text mining with R: a tidy approach. O`Reilly Media.

Souza, Thársis Tuani Pinto, Olga Kolchyna, Philip C. Treleaven, and Tomaso Aste. 2015. "Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry." http://arxiv.org/abs/1507.00784.

Tankovska, H. 2021. "Twitter: Monthly Active Users Worldwide | Statista." *Statista*. https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ (June 13, 2021).

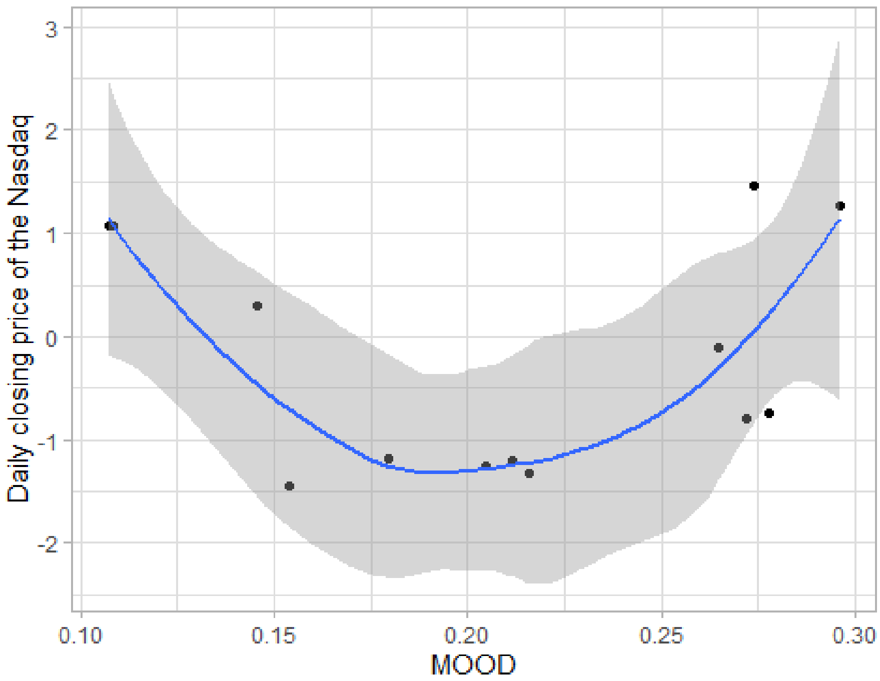
9. Appendix

Figure 29: Scatterplot between the daily closing price of the DJIA and "mood"



Source: Made with our data on R, "tidytext" package.

Figure 30: Scatterplot between the daily closing price of the Nasdaq and "mood"



Source: Made with our data on R, "tidytext" package.