universität
wien

MAGISTERARBEIT / MASTER’S THESIS

Titel der Magisterarbeit / Title of the Master‘s Thesis

„Media and Bitcoin Bubble

On the role of media and its reporting throughout the 2017 Bitcoin bubble“

verfasst von / submitted by

Valentin Valentinov Penev, Bakk.phil.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Philosophie (Mag. phil.)

Wien, 2019/ Vienna, 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066841
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Magisterstudium Publizistik- und Kommunikationswissenschaft
Betreut von / Supervisor:	Univ.-Prof. Hajo Boomgaarden, PhD

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die Masterarbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Valentin Penev Wien, August 2019

Danksagung

Es gibt viele Menschen, die auf die eine oder andere Weise dazu beigetragen haben, dass diese Masterarbeit zu einem Ende kommt oder generell zu meinem Studium.

Mein Dank gilt Prof. Boomgaarden, der sich bereit erklärte mein Betreuer zu sein, und der auch das Thema meiner Diplomarbeit für sehr interessant fand.

Zweifellos geht das größte Verdienst an meine Familie, die mich auf dieser Reise unterstützt hat. Sie dienen als gutes Beispiel, das übernommen und weiterverfolgt werden sollte.

Valentin, Polly, Ivo

TABLE OF CONTENT

FIGURES

TABLES

INTRODUCTION

"Everything you don't understand about money, combined with everything you don't understand about computers" - that is how the British comedian, John Oliver, summed up Bitcoin (henceforth called BTC) (Oliver, 2018; Sommerlad, 2018). The cryptocurrency, whose creator will likely remain forever unknown (Bernard, 2018), was initially interesting only for a relatively small group of cypherpunks (Shin, 2017). While there is a steady growth of daily transactions ever since its inception in 2009 (Olszewicz, 2019), suggesting the number of adopters to have increased, it is more likely the wider audience to be aware of its [of Bitcoin] existence, for the fact that at the end of 2017 the cryptocurrency hit just shy of 20, 000 USD - event widely covered by [mainstream] media (Mack, 2017; Morris 2017; Oliphant, 2017; Suberg, 2017; Wearden 2017).

This sudden appreciation of BTC turned it into the perfect speculative bubble, which Nobel-Prize winner Rober Shiller (2000, p.5) briefly defines as *"[…] unsustainable increase in prices brought on by investors' buying behavior rather than by genuine, fundamental information about value".* A computer bot, likely ran by and on then-very-popular-exchange, was to be blamed for a previous, similar explosive increase in valuation in year 2013 (Madore, 2018), and it was not long before doubts have arisen for another exchange to be in a foul play this time too (Kharpal, 2018). These suspicions were confirmed by researchers from Texas University, who claim that the exchange in question does indeed manipulate the price (Griffin & Shams, 2018). Others considered The People's Bank of China and the country's officials to be in command of the market (Durden, 2017).

Bitcoin is a rather new phenomenon, but there has been a growth in studies towards it (Estrada, 2017), with its price formation and dynamics also being closely looked at. For example, Buchholz, Delaney and Warren (2012) found out that volatility (price changes within certain interval, for example 24 hours) plays statistically significant positive role on price before but loses importance once the bubble has imploded. The CBOE Volatility Index turned out to be variable of importance in a study by MacDonell (2014), and van der Burgt (2018) explained the recent 20,000 USD price tag of Bitcoin with the help of the financial instability hypothesis developed by Minsky.

While this thesis does not deny or disregard those findings, *it simply distinguishes from them.* The reason is that bubbles "[...] *lie at the intersection between finance, economics, and psychology"* (Garber, 2000, p.ix), with the latter one, especially in the case of Bitcoin, being overlooked.

The aim of this master thesis is to explore what role media and its reporting might have played during the 2017 Bitcoin bubble. It presents concepts of media effects on the public and their relevancy within the financial world. On a later stage, researches looking at mass media's part in the formation of previous speculative bubbles among various commodities and assets have been reviewed. Both the theoretical framework and the previous empirical work in the field, serve as foundation in building a case to study the media aspects of the growing popular cryptocurrency's boom. What is more, this thesis is written with **ELI5** in mind. **"Explain Like I'm Five" (ELI5)** is a subreddit on the popular message board site Reddit, where visitors ask complex questions in any field (including economics, science or history), which are then answered by other users in layman language (Pflugfelder, 2017, p. 25 - 26). With over 16 million subscribers, ELI5 is a popular form to gain knowledge. Just like the discussion platform, this current paper is not really aiming for the youngest ones, however it strives to achieve full comprehension in adults, without explicit previous knowledge in computers, finance, media or cryptocurrencies. This is accomplished, without having impact on its academic eligibility. In order to further ease the process of reading, this paper does not just impart information, but is written in narrative style instead.

The first chapter is called and goes through ***"The Evolution of Money"***. It starts from the early ages of bartering, takes a look at the money as used by the majority of people nowadays and ends up with Bitcoin. Only by being aware of the basic history and idea behind money, one can grasp the motivation behind the creation of BTC, its advantages and flaws.

The following chapter established the meaning behind "***Financial Bubbles"*** and introduces three historic speculative mania examples across different assets and commodities. Subsequently the last three bubble formations of Bitcoin are outlined, depicted through charts, with special attention paid to the last one from 2017, which this thesis strives to explore.

Chapter number three introduces the effects that media has on audiences - 1st and 2nd level agenda setting, as well as some of their extensions. Only by having understood these, the reader can connect the dots of how media and reporting can have impact within the financial world. Subsequently, studies of the relationship between reporting and bubbles are discussed.

The 4th chapter outlines the research questions. Motivation behind each of them is defended briefly, based on theories introduced earlier. Following that, there is a section devoted to the argumentation behind the empirical method chosen – manual sentiment analysis on document level, as well as the publications' selection process.

"Statistical analysis and first results" is the title of the sixth chapter, which goes through the methods used for exploring the various relationships between the data variables. Descriptive statistics summarize the data set collected, followed by outputs of the (un)successful trials with correlation, regression and random forest among others.

The second to last section applies the statistical findings into answering the research questions. The limitations of the study design, which left some of the RQs not being explored at all, are closely looked at, providing guidelines for future researches on the topic. Last but not least, ***"Conclusion"*** summarizes the works' fundamentals, motivations and its outtakes.

LITERATURE REVIEW

EVOLUTION OF MONEY

Monies are not new to humankind. While in the beginning bartering of goods has been common, during the Hellenistic period precious metals were adopted as standard mean of exchange. In order to avoid weighting and to ease the transactions, standardized coins with exact drawing and weight were introduced (Albuquerque & Callado, 2015). Later on, despite few failures in history, paper money was put into circulation – it was easy to manufacture and hide, and lightweight enough to move around (Horesh, 2012). Since initially coins were minted from precious metals, their value was tied to that of the metal. Paper money on the other side, were backed by gold, held safe by the authorities (Bariviera, Basgall, Hasperué & Naiouf, 2017). At the same time however, as of Albuquerque and Callado (2015), the introduction of banknotes brought levels of inflation never seen before - essentially, the phenomenon of money losing their value. Arguably, the biggest factor to prevent it, is to control the money supply, thus its issuance (Labonte, 2011).

Humanity has made yet another step in its evolution and now is in the state of converting to *"cashless society"* - that is, cash losing its importance and moving to checks, credit / debit cards and electronic transfers as means of payment (Garcia Swartz, Hahn, Layne-Farrar, 2004). One can say, that we are shifting towards *digital money*. *Digital money is different from digital currency* (to be discussed on a later stage in this thesis), and represents the possibility, with the help of technology, for real money (no matter of the currency - EUR, RUB, USD) to be a *digital object*. If person A wires money to person B via his bank account, the latter is going to receive *digital money*. Only if person B withdraws it, on ATM for example, he will have *real, physical money*. The same way, payment for goods with credit card is done with *digital money* (Albuquerque & Callado, 2015).

No matter their state of evolution throughout the centuries, the country or the currency, money has always been *highly centralized* - only the respective authorities are allowed to create more of the it, granting decision makers with the full (monetary) power. On top of that, the majority of both developed and underdeveloped countries are freely printing more, without any kind of financial backing (Albuquerque & Callado, 2015). As De Heij (2012, p.10) notes in his book, cash is solely *based on trust* - the system functions on the faith between state and its community to accept and use the money. This became even more valid after 1971, when United States President, Richard Nixon, closed *the gold window*, which permitted holders of United States Dollars to exchange them for the gold they have been theoretically backed up with (Irwin, 2012).

BIRTH OF BITCOIN

In 1998, Wei Dai published his idea about "*a scheme for a group of untraceable digital pseudonyms to pay each other with money and to enforce contracts amongst themselves without outside help*", called "*b-money*" (Wei Dai, 1998; Okhuese, 2017). Some 10 years later, Satoshi Nakamoto (2008), inspired by Wei Dai's publication, released the whitepaper of *Bitcoin (BTC) - a form of digital cash*, which would allow one party to send payments to another without a third, trusted one in between. Nowadays, a payment, would it be a wire transfer, credit card one or direct online purchase requires a mediator, most often a financial institution. Transactions are rarely nonreversible and the level of privacy is questionable. The solution of Nakamoto (2008), excludes this third party, relies on cryptographic proof instead and makes the transactions impossible to revert.

The present *trust-based model* is a problematic one. As from the previous paragraphs, there is *trust* that the monetary policies, shaped by the authorities will be with the community interest in hindsight. However, with increased money issuance and actions similar to the ones of Nixon, the system depends on the social convention between state and society.

Bitcoin on the other side, being *digital currency*, is free from these drawbacks - there is no central authority, dependency on banks or various institutions. Main issuer does not exist - it is its core protocol, or predefined rules, that functions as *regulator* (Segendorf, 2014). The inflation is predictable too - the total supply of Bitcoin is set within its code (21 Million) and the last one is expected to be minted around year 2140 (Albuquerque & Callado, 2015).

Trusted mediators are excluded from the process. Transactions are sent to all participants in the network - computers, also called *nodes*. These nodes, function as bookkeepers and record all valid transactions in an accounting book called *blockchain.* A copy of this book is held by each node (Lansky, 2017). Currently there are almost 10,000 active full nodes (Bitnodes, 2018). Full nodes keep the whole history of the blockchain - from day 1. However, this database is rather heavy (40GB+), and not everyone has the spare resources to support it. *Light nodes*, which can run on mobile devices too, hold only the latest information of the blockchain - or, as of the accounting book example, they only keep information about the last 100 valid transactions (Parker, 2015).

While running a node does not require huge financial weight (Parker, 2015), *miners* which *broadcast transactions and mine new coins (albeit digital)* need to be incentivized in a way, for their considerable investment to support the network (Nakamoto, 2008). In the beginning this could have been done with consumer-grade computers, but these became uncompetitive to the pricey hardware, developed exclusively for mining purposes (Lewis, 2015). This comes on top of the process requiring so much electricity (*additional costs for the miners*), that some analysts are afraid a possible exponential growth of Bitcoin might suck in the entire world's electricity in 2020. The motivation for those who support the network lies within the voluntary fees paid for every transaction and newly mined coins. The code of Bitcoin is set so, that every 10 minutes there is new block added to the blockchain, or as of the layman example from earlier - new entry in the accounting books of the nodes. Transactions are sent to their recipients and new coins are being minted. Currently 12.5 Bitcoins are put in circulation for every block (ten minutes), but the core protocol is designed to reduce this reward by 50% every 4 years. It will be 6.25 BTC per block in mid 2020, 3.125 BTC in 2024, with the drop of 50% every 48 months to continue until all coins are released (Ankalkoti & Santhosh, 2017; Gauer, 2017).

It is impossible to know the exact number of miners, but some researchers suggest they might be as high as 100 000 (Parker, 2015). With the evergrowing number of nodes spread across the globe in mind, *Bitcoin seems to be fully functioning decentralized payment system with no-single point of failure* (Nakamoto, 2009). Among its various advantages over the traditional payment system are the swift, low fee and anonym transactions (Segendorf, 2014).

FAST, CHEAP AND ANONYMOUS TRANSACTIONS?

As from a paragraph earlier, transactions are validated, i.e. recorded in the accounting book / on the blockchain, every ten minutes (each reward-block). At the time of writing, the fee for the fastest transaction possible would be 0.11 USD (11 cents). If a payment is not that urgent and can wait for up to 60 minutes, the fee would drop over 50% to just 0,05 USD (5 cents). During busy times for the Bitcoin network, i.e. when there are a lot of transactions pending, fees are higher. One such period was late December 2017, when the average fee for the fastest transaction was just under 40 USD (BitcoinFees, 2018). Those are rather extreme occasions and most of the time, a fast payment can be executed for less than 30 cents. In 2018, 194 Million USD worth of Bitcoin were moved for just 10 cents. With TransferWise, service known to offer considerably cheaper wire transfers than banks, this very same transaction would have netted fees of 7,500 USD, and required few days for clearing (Young, 2018). In 2013, sender took advantage of the non-congested network and transferred close to 195 000 BTC, worth around 145 Million USD at the time, for free – since, as mentioned earlier, *fees are voluntary* (Southurst, 2013).

Another convenience of using Bitcoin is its *anonymity*. As of the critics however, this is also its biggest drawback. Economist Joseph Stiglitz, Nobel-Prize winner, suggests that since the digital currency opens a blank hole for criminals to transact with each other, sooner or later, Bitcoin will be brought down by the authorities (Montag, 2018). Australian study reviewed all valid transactions on the network between 2009 and 2017, and using novel technique, concluded that 25% of Bitcoin adopters are using it for illegal activities, with nearly half of the transactions being crime-related (Foley, Karlsen & Putnins, 2018). Another report claims, some $2.5 billion have been laundered through Bitcoin (Canellis, 2018).

This anonymity however is not aimed at but is rather a consequence of Bitcoin's design. Since it is *decentralized*, there is no authority which assigns wallets to individuals, similar to bank institutions for example, but addresses (also called *public keys*) are generated in cryptographic manner instead. They consist of numbers, upper- and lowercase letters and are 34 characters in length – the first address to ever receive a transaction, for example, was *1A1zP1eP5QGefi2DMPTfTL5SLmv7DivfNa. Along this public key, which equals an IBAN*, and is used to receive funds, there is *private key* generated in the same way, but even more complicated - with 64 characters. It has the functions of a password - whoever has the private key, can control the address associated with it, even without exclusively knowing it. There are about 10 ^ 48 possible bitcoin addresses (that is a number consisting of 49 numbers!), and despite an individual being able to create multiple ones, chances of the same [address] being generated twice are slim to none (Guegan 2018, Hileman & Rauchs, 2017).

Another feature of Bitcoin is its *auditability* – addresses and transactions are anonymous, but the blockchain, i.e. the accounting book, can easily be reviewed by everyone. On *BlockChain.info* one can enter a public key and do a follow up of all outgoing and incoming transactions, including information about the transaction date, amount of BTC moved, their approximate fiat value at the time and the fee paid. If a particular transaction is of high interest, it can be reviewed in details too (BlockChain.com, 2019; Er-Rajy, El Kiram, El Ghazouani & Achbarou 2017). This unrestricted access to the blockchain however, *has raised concerns of the actual anonymity Bitcoin can offer*. Juhasz, Steger, Kondor & Vattay (2016) suggest complicated, but cheap and easy to implement model, that would allow to identify the exact IP or approximate location of some users. However, there are far easier ways too – something as simple as third-party web cookies can de-anonymize the user (Goldfeder, Kalodner, Reisman & Narayanan, 2017), and popular websites and wallets store detailed users' information (Herrera-Joancomart, 2014). This comes on top of the fact, that nowadays all major and legitimate exchanges require for its users to go through KYC / AML procedure (Underwood, 2018). For the really hard cases to crack, as revealed by Snowden, NSA (National Security Agency) – the same US agency, that tapped german chancellery for decades, even Angela Merkel – has been working on, supposedly, successful solution to put names to Bitcoin transactions (Biddle, 2018; Carrel, 2015).

IS BITCOIN MONEY AT ALL?

Bitcoin remains experimental technology (Kraslawski, 2017). There are discussions, if it can be adopted as money on the first place. For this it has *to be medium of exchange, unit of account and store of value. Its effectiveness as medium of exchange (1) is limited* - for this, a good number of common, widely-used goods should be purchasable for BTCs, which as of today, is not the case. (Wolla, 2018) BitPay (www.bitpay.com - https://bitpay.com/docs/) has a solution, which would allow for Bitcoin to be easily accepted by merchants, however, as discussed earlier - when the network is busy, fees are high, and transactions usually need some minutes to confirm (BitPay, 2019). Another industry leader, Coinbase, offer a similar solution (Bulat, 2018), but the company aims at increasing the adoption among users too, by rolling out debit cards which are to be topped up with cryptocurrencies (Feroz, 2019).

It theoretically could be unit of account (2), but most sellers still prefer the traditional, fiat currency. If we consider that Bitcoin is priced at 5000 USD, vehicle's $40 000 price tag can be displayed as 8 BTC, but a coffee worth $2.9 would be the confusing 0.000597 units of a Bitcoin, called Satoshis. Another issue is *its volatility, which also questions its ability to be a store of value (3).* There is a popular joke where a kid asks his Bitcoin-invested dad for 0.1 of the digital currency and the father answers *"147 bucks?! What do you need 155 dollars for?"* This is why, even if a merchant would accept BTC, it is often converted straightaway to traditional currency in order to avoid potential losses caused by price fluctuations. T*o be considered as a store of value*, BTC would need to hold its purchasing power within some reasonable limits (Lo & Wang, 2014; Wolla, 2018). And while Bitcoin went from under $2 (Arthur, 2011) to average price of roughly $3700 for the month of December 2018 (Statista, 2019), one should also remember that this is more than 80% down from its $20 000 price in late 2017 (Suberg, 2017).

Another potential difficulty is the storing of Bitcoin. While many would choose the ease to keep their coins on an exchange, this is far from being optional solution. In 2018 alone, exchanges were hacked of close to $1BN worth of cryptocurrencies (Khatri, 2018). Mt.Gox remains the most recognizable one, when in 2014 close to 850 000 BTCs got stolen (Norry, 2018). The issue here is that by storing BTC, or any other cryptocurrency on custodial exchanges, it is them [the exchanges] knowing and controlling the private keys, i.e. the passwords that give control over the digital assets. As Andreas Antonopoulos, famous Bitcoin supporter says - *"Not your [private] keys, not your Bitcoins"* (Dickinson 2018, Ogundeji 2016). While those attacks are rather sophisticated, having in mind the security teams engaged in such trading platforms, the sage malicious actors started targeting individuals with *phishing attacks* - impersonating emails, exchanges, online wallets and projects in order for neglectful users to enter their private information or login data, enabling the hackers to gain access to their profiles and funds (Drozhzhin, 2018).

Hardware wallets, such as the Ledger Nano S, are the most viable alternative so far. It is protected via a secure chip, actual physical buttons need to be pressed to confirm a transaction, and private keys never leave the USB-like device. However, it is fairly uncomfortable for daily use and requires above-average computer knowledge to work with. Also, one needs to safely keep the randomly generated 24 words *(seed)* in order to regain access to the funds, in case the device gets lost or damaged. It has to be kept offline, f.e. on a piece of paper, and if it [the piece of paper] gets stolen or lost, the funds can never be recovered (Agrawal, 2018; Ledger Wallet, 2019). The problem is far from marginal, with the estimates of Chainanalysis (2018) showing the access to between 2.3 and 3.7 million Bitcoins to be permanently lost.

Last but not least, the situation of Bitcoin is not as dull as many might find it. At the end of the day, this is experimental payment system, created by an individual whose identity remains unknown (Hodge, 2018), with its first transaction being merely 10 years ago (Campbell, 2019). In 2010, a programmer could barely find volunteers to exchange 10 000 BTC of his (worth close to $200 Million at Bitcoin’s 2017 peak) for two large pizzas (Wong, 2018), however nowadays it is possible to make real-estate purchases using Bitcoin (O’Brian, 2018) and even pay your taxes with it (Vigna, 2018). And while Thailand and China have enforced hefty restrictions against cryptocurrencies as a whole (Chong, 2018; Helms, 2018), venezuelans are readily buying Bitcoin, escaping from the rapid devaluation of their local fiat currency (Comben, 2018). In his speech, Valdis Dombrovskis, vice president of the European Commission admitted that *cryptocurrencies are here to stay,* and it is time for their legal status to be defined (European Commission, 2018). The US Stocks and Exchange Commission is also looking closer at them (Clayton, 2017), and giants like NASDAQ and the owner of NYSE (New York Stock Exchange) are expected to launch crypto-trading rather sooner than later (Huillet, 2018; Tully, 2018).

Bitcoin is the first, and so far, most popular cryptocurrency (DeVries, 2016). Brave New Coin (2019) estimates that the during the first week of 2019 close to $3 Billion have been moved on the blockchain via more than 1 862 000 transactions. It [Bitcoin] paved the way, and as of today there are more than 2000 cryptocurrencies in existence (CoinmarketCap, 2019). Some, like Monero have been created with the idea to offer complete anonymity (Alonso, 2018), others, like Ethereum allow for everyone to launch applications on its decentralized network (Vujičić, Jagodić & Randić, 2018). And while every cryptocurrency supposedly functions on a blockchain, *only a* *small percent of the projects are in fact truly decentralized,* in the sense transmitted earlier (Mizrahi, 2018).

FINANCIAL BUBBLE

WHAT IS A FINANCIAL BUBBLE?

Widely used in media and common explanation of *an economic bubble is the situation, where an asset experiences a dramatic price rise in such a short period of time, that is it very likely for the price to go through equally sudden downfall.* While understandable, that explanation faces unclarities. For example, it does not define what a short period of time is, and how much is too big of a price rise. What is more, vivid price fluctuations may happen due to simple supply and demand correlation - when a new fashion accessory is introduced, strongly desired and limited, it will enjoy a high(er) price tag. However, once mass production hits the market or another fashion line is presented, its [the current fashion accessory one] price will fall (Barlevy 2007).

That is why, a more suitable description of *a bubble, would be "the situation where an asset price moves significantly away from its fundamental-based value."* (Kubicova & Komarek, 2011, p.34) Such classification, with a notable gap between the asset's intrinsic value and its price, is widely used in the modern economic theory (Barlevy, 2007). This definition of a bubble has also been adopted by recognizable institutions such as Bank of Canada (Giusti, Jiang, & Xu, 2014), The Federal Reserve Bank of Chicago (Evanoff, Kaufman & Malliaris, 2012) and associates of the International Monetary Fund (Jones, 2014).

Theoretically, the *fair* fundamental-based valuation is a calculation, that considers all future possible returns and risks (Girdzijauskas et al., 2009). Asset is a claim of future payments, and as such, its value is, or should be, correlated to these dividends. Having this in mind, a sharp increase of its price, without any news of payout changes for example, would signify that this particular asset is likely overvalued, i.e. in a bubble (Barlevy, 2007). As the author [Barlevy] points out, such calculation is straightforward for assets, that yield known stream of dividends, and gets rather complicated when periodical proceeds are uncertain. How to figure out the intrinsic value varies on the asset - with real estate for example, one will likely need to use complicated mathematical models coupled with deep market knowledge. (Hott & Monnin, 2006; Smith & Smith, 2006) Gold, commodity with the ultimate *store-of-value* characteristics, which has been traditionally seen as *safe haven* for banks, governments and smaller-scale investors (Baur, 2013), has no agreed determinant of its fundamental value. By lacking this important metric, some question if gold can even experience being a bubble (Lucey & O'Connor, 2013).

The literature on economy is practically immense. If one dives deep enough into the topic, eventually Townsend's models of money will turn up, where money is described as *"intrinsically useless"* (Townsend, 1980, p.265). Barlevy (2007, p.46 & 49) goes even further, and using the projections of Townsend states *"[...] the fundamental value of money should be zero, and the fact that buyers and sellers are willing to trade valuable goods for money implies its price exceeds its zero fundamental value"*, calling it a bubble. This serves to illustrate, that in case all interrelated phenomenon are reviewed, there is possibility for the paper to shift into economic one, losing its primary focus. This is why, on the following pages explanatory additions for the needed terminology will be included, only to achieve a more complete understanding.

This thesis adopts the more sophisticated definition of a bubble, presented on the previous page - *a notable disparity of asset's price and its intrinsic value,* with the small remark that a bubble could also be a *negative one - essentially the price of an asset staying below the one implied by its fundamentals* (Jones, 2014). It is also worth noting that a bubble may occur in any traded commodity or instrument (Conerly, 2013). To gain a better understanding, three, likely the most explanatory cases of speculative bubbles are introduced. All of them occurred in various [economic] epochs, featuring different durations and across contrasting type of assets.

TULIP MANIA (1634 - 1637)

The Dutch tulip mania in the 1630s is likely the first known financial bubble in our history. This particular event enjoys such popularity, that is nowadays a synonym in our jargon for economic bubble (van der Veen, 2012). Peter Graber, who is considered to be the tulip mania expert (French, 2006), unironically put two mosaic-colored tulips on the cover of his book, *Famous First Bubbles* (Garber, 2000).

The first tulip in Europe blossomed in the Netherlands, in 1594, after its bulbs were gifted by ottoman officials to Carolus Clusius, a famous botanist. Flower gardens were not uncommon at the time and having from the scarce tulips was the way to claim exquisiteness (Garbarino, 2011). Soon enough, breeders started working on new varieties with spectacular colors and patterns (Garber, 2000), for which the wealthy ones were paying hefty amounts (van der Veen, 2012).

Normally, the flower has been purchased during summer, after it has blossomed. However, a trend occurred, where tulips were bought in winter, with vague expectations of their future visual characteristics. This allowed for *futures market* to occur in 1636, where investors would buy, or bet, on selected bulbs in the colder months, expecting them to be worth much more once they have bloomed. At the very frenzy of the mania, even common bulbs were sold for twenty times their original value. *Semper Augustus* was supposedly the most sought after tulip, with its flowers being a rare mix of fire red and white. It sold for thirty (30) average yearly salaries, and one buyer even swapped his house for a single bulb of the sort. This hysteria came to an end on February 5th, 1637, when the flower market crashed. The Dutch stock market was not involved in the tulip trade, and the sell-off caused no damage to it or the economy of the country, however, separate individuals and businesses went bankrupt. (Garbarino, 2011).

Gisler (2012) argues, that the tulip mania was a demand and supply occasion, which, as of Berlevy's opinion outlined earlier, would not classify it as a bubble. However, this does not explain why even a common bulb, such as the Witte Croonen rose 26 times in January 1637, only to lose 95% of its value a week later (French, 2006). At the peak of the mania, people were selling their possessions to buy from the *"precious"* bulbs (Garbarino, 2011), however only *losers* (Graber, 2000, p.3) could not see what was coming.

There are many suggestions as of why the tulip mania occurred in the first place. Graber (2000) calls it crowd irrationality, while van der Veen (2012) has other rather valid suggestions - namely, more prominent middle class, general rise of wealth and free capitals, which allowed for a higher risk tolerance. There is one more interesting observation to be made - some 30 years before the tulip mania occurred, The Amsterdam Stock Exchange was established, the first ever stock exchange in the world, and curiously enough a big part of the wealth creation, that made the frenzy possible materialized namely through that exchange. (Goetzmann, 2015).

Shiller (2000, p.245-246) outlines early stories about speculative price movements, suggesting that tulip mania might not be the first bubble in history. Supposedly, both pepper and some grains surged in valuation, with the former one experiencing volatility too. Written correspondence point at the land in ancient Rome rising so much, that it caused discussion among local people. However, the first widely covered bubble remains to be the tulip one, partially because it coincides with the first regularly published newspapers, which may have helped drive investors to a state of irrational exuberance. As Shiller concluded *"The history of speculative bubbles begins roughly with the advent of newspapers."* (Shiller, 2000, p.71)

THE GREAT DEPRESSION (1929 – 1939)

While the Dutch tulip mania affected a rather small number of people, the same cannot be said about the Great Depression, which hit not only the United States and Europe, but Japan and Latin America to a certain degree too (Romer, 2003). In fact, it was not until the end of the 1980s, until its emotional, intellectual, cultural and social consequences were overcome by the humanity (Rothbard, 2010). It is said to be the economic crash of the 20th century, the worst in the history of the US (Wheelock, 2007), and Romer (2003) suggests that this was one of the events which set the stage for Second World War. The author means that only the Civil War resulted in more deaths than the Great Depression.

The Depression started in late 1929, with the US markets crashing severely and lasted for 10 more years. For that period, quantity of production by the US fell by almost 50 percent and real

GDP - 30 percent. A banking panic [that is, regular depositors withdrawing their money from the bank, in fear that it might not be liquid enough] occurred, and by 1933, 20 percent of the banks [7000 banks as of some reports] were out of business. In that same year, then President Franklin Roosevelt, announced *"bank holiday"*, in an attempt to prevent domino effect. Worldwide, output and standards of living dropped significantly, with 25 percent of the people in the industrialized world left jobless for at least a few years to come (Romer, 2003).

As of today, there is still not agreed-upon understanding what exactly might have caused the Great Depression. Many theories have been proposed over the years, however, lately economists have been looking at the stock market crash, which has been largely ignored by now (Cecchetti, 1992; Wheelock 2007).

At the end of 1929, the stock market experienced such a big drop, that it has been labeled as *"The Great Crash of 1929"*. For the 8 years prior that, it [the stock market] experienced a tremendous growth of 400 percent. With the time, investors started feeling insecure and those, who could see the alarming signals, decided to flee the market by liquidating their holdings, on the day known as*"Black Thursday"*, marked with a severity of panic selling unseen by that time (Romer, 2003). Besides lost wealth, the burst of the bubble also made people prudent, question the state of the economy and shrinkage in both consumer and company spending followed (Wheelock, 2007).

Interesting detail is that the 1920s have been generally flourishing years for the US economy, where prices have stayed largely the same with temporary negligible recessions. Only the stock market experienced a dramatic boom. US officials may have realized the exuberance speculation going on and tried various techniques to limit it (more firm monetary policies and restrictions on the credits to brokers), likely helping the bubble to burst. Back then, the countries used to hold on to the gold standard, and central banks had to react and cushion gold outflow from other countries in exchange for the American surplus, which essentially allowed for the Depression to spread even to the far east (Cecchetti, 1992 & Romer, 2003).

DOT.COM BUBBLE (1998 - 2000)

The dot-com bubble [called so because companies with *".com" (dot com) or "e-"* in the name were perceived as valuable by default] is another prominent one, which, similar to The Great Depression was hinged to the US stock market. While it did not result in such a severe impact, neither national nor international, by late 2002 the NASDAQ10 index has fallen by almost 80% and investments were put on hold worldwide. Back in 1995, Netscape held its Initial Public Offering (IPO) and its managers decided to price it at $28 per share, which was more than double the $12 - $14 price range suggested by Morgan Stanley, the investment bank that cooperated in the IPO. The sale was a success, and on its first trading day, Netscape touched $71 per share. Soon enough, a lot of similar offerings occurred, by companies venturing in the new revolutionary technology - the internet (Joosten, 2012).

For the years between 1997 and 2000 the new technology stocks experienced a growth of over 500 percent (Griffin, Harris, Shu & Topaloglu, 2011). Analyst bankers did not question the rapid growth promises (Howcroft, Richardson & Wilson, 2001), and people without experience in trading or investing rushed into the stock market, fearing they might miss the boat of high returns (Joosten, 2012). Absurdity reached levels, with economists speculating humanity is on the verge of a new era, where inflation and recessions could not exist (Medipally, 2018). They however were soon to be proven wrong.

The dot com bubble imploded in March 2000, and by year's end, the gains from the previous years got erased (Ofek & Richardson, 2003). In hindsight, the reverie was evident. Etoys and PlanetRX.com for example, an online pharmacy, were valued at more than $10 billion, while controlling only a few millions of assets and turning no profits. Kozmo, an online delivery company used to spend more than $30 million a month, without generating any profits too. Boo.com, which specialized in selling apparel online, managed to waste $120 million in little over a year, achieving no meaningful results (Friedman & Hirakubo, 2002). In the early months of 2000, a number of companies started to run out of cash, and investors realized the multibillion dollar enterprises did not manage to secure even the minimum for their existence by themselves, but relied heavily on outside financing to stay afloat (Howcroft et.al., 2001). When prices plunged, investors became insecure to buy back. The companies needed to restructure in order to survive. Kozmo for example, laid off 2,200 works and brought down its monthly expenses to $2 million, only to be liquidated few months later (Friedman & Hirakubo, 2002). Legitimate companies were not left unaffected - Amazon lost close to 70% of its share price by the summer of 2000. Lehman Brothers publicly advised interested parties to avoid its stock, and by 2002 Jeff Bezos needed to fire 15 percent of his workforce, close warehouses and customer-service center, to counteract a financial loss of $1.4 billion US dollars (Frey & Cook, 2004).

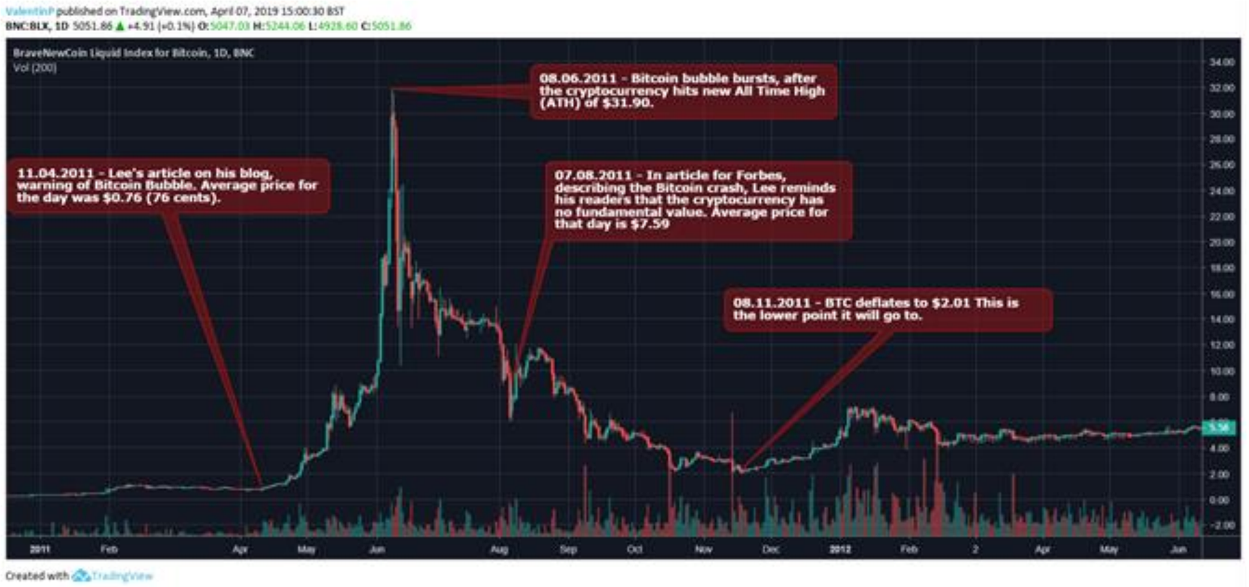
The bubble was predisposed to occur due to low interest loans and the relative ease for one to engage with stocks (Joosten, 2012). Another factor is that, a lot of the new companies to come into existence were competing in the same niches, causing oversaturation. What is more, the newcomers lacked basic business models, were growing too fast and many of them did not offer anything valuable for the customers to start with (Friedman & Harikubo, 2002).

BITCOIN BUBBLE(S) (2009 - 2019)

As mentioned earlier, every traded asset can go through a financial bubble. If we take a closer look at the price history of Bitcoin, there are a few to be identified, despite the fact it has been in existence for little over 10 years.

The first more notable one is from 2011. Back then, the cryptocurrency did not enjoy as much popularity, however there is still some coverage to be found. Timothy Lee (2011) for example, famous tech and blockchain journalists, outlined in his blog the inability of Bitcoin to get any demand outside its tight support group and the speculators. He warned the readers that the virtual currency has no fundamental value, and the recent price surge, albeit small, has been driven by irrational exuberance, which, as in all other cases, will result in the bubble popping. On a side note, the day his blog was published, 11.04.2011, Bitcoin averaged a price of 0.76 USD (76 cents).

The bubble did indeed implode – in June 2011, after hitting a new high price of 31.90 USD per Bitcoin. In an August article for Forbes, Lee (2011) again reminded of the issues that the revolutionary currency is facing. He meant, real estate could be overvalued, but provides a roof to live under. Gold might burst as well, but at the end of the day jewelry still will be made out of it, implying these assets are unlikely to go to zero. Bitcoin on the other hand, 7 USD at the time of publishing, featured no intrinsic value whatsoever, and was doomed to end up worthless. Lee turned out to be correct once again. BTC continued to slide down, however it bottomed, that is, the price not going further down, in November 2011, at 2.01 USD. While that was more than 93% decrease from the top, until the publishing of this thesis (August 2019), Bitcoin is yet to touch the 76 cents price tag from mid-April 2011.



08.06.2011 - Bitcoin bubble bursts, after the cryptocurrency hits new All Time High (ATH) of $31.90.

11.04.2011 - Lee's article on his blog, warning of Bitcoin Bubble. Average price for the day was $0.76 (76 cents).

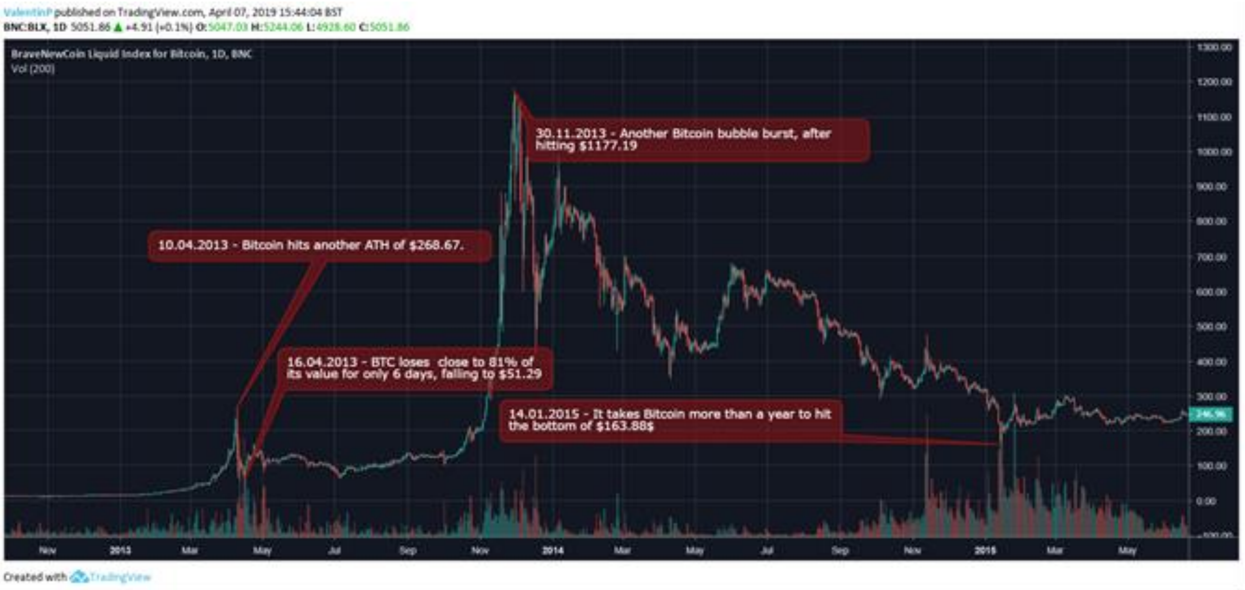
07.08.2011 - In article for Forbes, describing the Bitcoin crash, Lee reminds his readers that the cryptocurrency has no fundamental value. Average price for that day is $7.59

08.11.2011 - BTC deflates to $2.01 This is the lower point it will go to.

Figure 1. Daily chart of BTC bubble in 2011 from < $0.7, up to $31.9 and its bottom in 2011 at $2.01. The explanatory texts pinpoint to the exact dates of the price and Timothy Lee's articles. Source: TradingView.com

The next speculative period took place in 2013. It started forming in late 2012 and peaked at 268 USD on 10th of April 2013, only to hit the bottom at 51 USD six days later! It picked up steam later that year again, and the run up ended at close to 1200 USD at the end of November 2013. This bubble phase needed a little longer to fully deflate, and the depreciation of Bitcoin ended in mid-January 2015 at 163 USD. We can only speculate of the exact reasons why, one of them possibly being that it broke the psychological level of 1000 USD, but at the time, the cryptocurrency started getting attention from both mainstream media and influential people. In interview for Huffingtonpost (Zoldan, 2013), Felix Salmon, renowned financial journalist, expressed his fear of Bitcoin being commodity in a bubble state, with significant acceptance and growth potential limitations, mainly due to it being independent. In an experiment for the VICE magazine, Baumann (2013) covered living in Panama solely relying on Bitcoin, calling it *"the*

Mickey Mouse currency" and *"funny money"*. Kaminska (2013) published a rather negative report for the Financial Times, while the growing popular ZeroHedge had positive short – term expectations, but a rather dark overall outcome (Durden, 2013). Nout Wellink, who was then already retired head of the Dutch central bank, compared Bitcoin to the country's Tulip Mania, hinting that while speculators from the latter one were left at least with flowers in their hands, the cryptocurrency is a pure vaporware (Hern, 2013; Worstall, 2013). In interview for Bloomberg, ex Federal Reserve Chairman Alan Greenspan also called it a bubble, with questionable intrinsic value (Kearns, 2013). The professor of Finance at Boston University Mark Williams (2013), warned of the flawed DNA the alternative currency has, and that it will easily fall to single digits by mid 2014.


ValentinP published on TradingView.com, April 07, 2019 15:44:04 BST

BNC:BLX, 1D 5051.86 ▲ +4.91 (+0.1%) O:5047.03 H:5244.06 L:4928.60 C:5051.86

30.11.2013 - Another Bitcoin bubble burst, after hitting $1177.19

10.04.2013 - Bitcoin hits another ATH of $268.67.

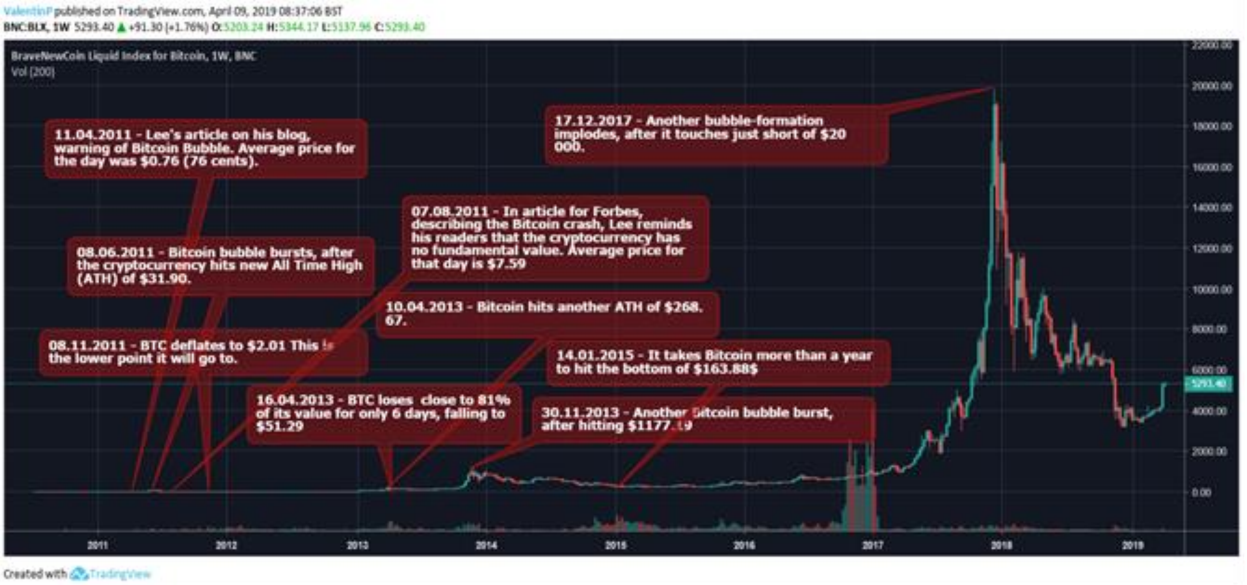
16.04.2013 - BTC loses close to 81% of its value for only 6 days, falling to $51.29

14.01.2015 - It takes Bitcoin more than a year to hit the bottom of 163.88

Created with TradingView

*Figure 2.*Daily chart of BTC's bubbles in 2013. It hit $268 and fell to $51.29 six day later, only to climb to $1177.19 on 30.11.2013. The subsequent deflation of Bitcoin was its longest in duration, bottoming out at $163.88 on 14.01.2015. Source: TradingView.com

Bitcoin managed to prove the sceptics wrong once again. From the 163 USD low in mid-January 2015, it climbed, first slowly, then more rapidly, all the way up to little under 20 000 USD on 17th of December 2017, when the bubble popped. A sell-off of BTC occurred, with the cryptocurrency losing more than 45% of its value within only 5 days. This bubble period was no less interesting. Bitcoin was "banned" by China (Rapoza, 2017), called *"fraud that will eventually* *blow up"* by Jamie Dimon (Imbert, 2017) only for the bank which CEO he is, JPMorgan, to heavily invest into the novel monetary concept (Buck, 2017). While BTC lost 84% from its $19764 peak until mid-December 2018, there is no way to confirm that its depreciation has come to a stop, until the current top is taken away. Nonetheless, this paper focuses on the way up in this 2017 bubble phase. While it might be challenging to come up with even rough calculation of Bitcoin fundamental valuation, as discussed in a previous chapter, every traded instrument could experience a state of bubble. What is more, the sharp price movements of Bitcoin remind of the ones in the examples presented before and match the bubble understanding introduced earlier.



11.04.2011 - Lee's article on his blog, warning of Bitcoin Bubble. Average price for the day was $0.76 (76 cents).

08.06.2011 - Bitcoin bubble bursts, after the cryptocurrency hits new All Time High (ATH) of $31.90.

08.11.2011 - BTC deflates to $2.01 This is the lower point it will go to.

07.08.2011 - In article for Forbes, describing the Bitcoin crash, Lee reminds his readers that the cryptocurrency has no fundamental value. Average price for that day is $7.59

10.04.2013 - Bitcoin hits another ATH of $268. 67.

16.04.2013 - BTC loses close to 81% of its value for only 6 days, falling to $51.29

14.01.2015 - It takes Bitcoin more than a year to hit the bottom of 163.88

30.11.2013 - Another Bitcoin bubble burst, after hitting $1177.19

17.12.2017 - Another bubble-formation implodes, after it touches just short of $20 000.

Figure 3. Weekly chart of the entire lifetime of BTC. It shows the grand scheme of things, and how previous bubble formations look unimpressive to the current ATH and pop at close to $20000. Source: TradingView.com

MEDIA EFFECTS

This chapter focuses on some of the concepts that outline how and what effects media might have on the public. The respective theories, while still in their infancy, were mostly studied in regard to the broad topic of politics. Namely that is the reason why on the following pages there will be a number of cases referring to candidacy and elections. These works did not only tramp the trail for many others to follow but are well-suited enough to function as illustration for when presenting the models. At a later stage, their relevance within the financial world is discussed. The *multi-facette agenda-setting theory* is fundamental and is introduced in details, due to its "[...] *compatibility with a variety of other communication concepts and theories*" (McCombs, Lopez-Escobar & Llamas 2000, p.78), only to be tied to the *two-step flow of communication concept and opinion leadership*, which are briefly explained. On top of that, a mere decade after the inception of agenda-setting, the inborn media effect got the attention of experts to the level, that it became a major topic on every communication science conference (Blood, 1982), with more than 400 associated studies by the 2000s (Strömbäck & Kiousis, 2010). Some rushed to replicate the work of its founders McCombs and Shaw (Iyengar, Peters & Kinder 1982), while others were genuinely motivated to theoretize, hypothesize and experiment with it, resulting in further findings. Having this in mind, the current section's purpose is not to review all of its *extensions*. Some are fully introduced, others briefly mentioned and the rest - deliberately left out, focusing on the ones of importance for the thesis. Among the recommended works [partially used in this paper] for in-depth look of agenda-setting theory, its history and interrelated phenomenon are : *"The Evolution of Agenda-Setting Research: Twenty-Five Years in the Marketplace of Ideas"* by McCombs and Shaw (1993), the book *"Agenda-Setting"* by James Dearing and Everett Rogers (1996), *The Agenda Setting Journal* and its 2018 article by Chris Vargo - "*Fifty years of agenda-setting research: New directions and challenges for the theor*y".

FIRST LEVEL AGENDA-SETTING THEORY

Scholars have been for long debating of the influence mass media might have on the public. The first half of the 20th century was marked by the popular belief that media directly *injects* information and opinion into the helpless recipients - view, known as the *hypodermic needle theory* (Neuman & Guggenheim, 2011). These concerns become even more valid, if we consider the assumptions that media is created and financed by the elites, who may [actively] direct the content that comes out, i.e. the agenda. Some early studies, albeit with focus on radio, did refute these fears, suggesting the supposed effect on listeners is somewhat weak (Rogers & Dearing, 1988).

The second half of the 20th century set the beginning of more serious research into probable media effects. In a very popular review on the topic, the political scientist Bernard Cohen (1963, p.13) wrote - *"The press may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about".* McCombs and Shaw (1972), partially influenced by Cohen's view, performed a well-executed study, where they indeed matched North Carolina voters' beliefs of what they find important, to be the same as these [beliefs] made prominent by media earlier during the campaign itself, demonstrating the ability of mass media to define the agenda. This work gave the reasoning behind the ***[first] level agenda-setting function of media***, with the authors arguably being the most well-known propagators of the concept. Until today, Cohen's famous phrase remains simplistic, but accurate way to portray the meaning behind it. The results of McCombs and Shaw imply for news to not simply function as information stream to society, but to largely influence what topics it will consider important. In another early experiment, Funkhouser (1973, p.74) also concluded that *"[...]the amount of media attention given to an issue strongly influences its visibility to the public"*.

With the time however, it became clear that agenda-setting is not as straightforward as it seems. For example, *low media credibility and personal experiences*, could hinder recognizing the issue as important. The recipient not holding the topic for meaningful at all, might neutralize the setting too (Rogers & Dearing, 1988). In a study within the South Korean newspaper landscape, Lee and Hahn (2014), found out that it is less likely for media to set the agenda for older and more educated people, compared to younger readers, who lack prior knowledge and critical thinking. Information presented with depth and additional research would rather force issue recognition within the audience, unlike short, sensational reports, often seen as untrustworthy. There are critics too. In 1985 Hill performed a study on agenda-setting function and television. He concluded that viewers' *news awareness and education* are major factor for them being able to comprehend and recall the news on the first place. However, his results show the quantity of TV news to be of a marginal effect on salience, which speaks not only against the basic understanding behind the agenda concept, but is also in discrepancy with McCombs and Shaw (1972), who found amount of news along their relative placement to be a major factor in their setting confirmation. Erbring and Goldenberg (1980, p.45) also came to the conclusion that *"it would be unwarranted to assume that any increase or decrease in media coverage invariably produces a corresponding increase or decrease in individual concerns",* calling for more research.

At some point dedicated agenda-researchers emerged, who looked into the agenda interrelationships between media, public and policy actors (Berger, 2001). Some of them decided to focus on *"[...]why information about certain issues, and not other issues, is available to the public[...]"* (Dearing & Rogers, 1996, p.2) in attempt to find out *"who, or what, sets the media's agenda"* (Turk & Franklin, 1987, p. 29). This smaller field of study became known as *agenda-building*, where researchers investigate how might the articles of journalists, and therefore public opinion be influenced by outside forces. Especially in politics, there are many individuals who would and theoretically could meddle into media coverage, along foreign countries and the corporate elite to a certain degree too (Parmelee, 2014). Political public relations for example include speeches, conferences, interviews and news releases and there are convincing evidences for the latter one to successfully steer media reporting towards their initial sender (Kiousis, Park, Kim & Go 2013, p.654). Other studies show that the majority of stories on newspapers and television originate from external sources, rather than being in-house work. The reasoning behind that phenomenon could be economical one - publishers, no matter of their medium, need to assure the availability of enough manpower (for example reporters and photographers) to provide content. In that sense *"[...] news sources who are able to reduce the costs of reporting news will be able to exert greater influence on the news media agenda.*" (Berkowitz, 1992, p.86). Since media could be seen as the connection between the government and the citizens, one can argue that it is also able to interfere in the policymaking process, i.e. *to set the policy agenda*. Media defines, draws and sustains the public attention to an issue. There is enough evident literature that media can focus the public awareness on politically related turmoils and their respective solutions, which then leaves it up to politicians to handle the situation (Soroka, Lawlor, Farnsworth & Young, 2012).

SECOND LEVEL AGENDA-SETTING THEORY

One of the bigger leaps forward in the theory of agenda-setting however, comes from one of its founders - McCombs. He suggested, slightly tweaking Cohen's prominent phrase, that media tells us not only what to think about, *but also what to think about it*, a concept known as *second level of agenda-setting theory*. Every object that is being communicated also features various *attributes* - these are, its unique characteristics and properties. While first level agenda-setting defines what objects (f.e. political party) the public to be aware of, *second level agenda-setting lies within the salience transmission from the media to recipients of the attributes used to present that political formation,* hence it is also sometimes being referred to as *attribute agenda-setting*. Undoubtedly, the choice of both objects and their respective characteristics to be under the spotlight is a powerful tool. While the selection of what to be covered can be explained with news values (Helfer & Aelst, 2016) or the news selection filter, i.e. *"Gatekeeping"* (Barzilai-Nahon, 2011), the assignment of their individual attributes is largely arbitrary. Here, second level agenda-setting theory draws connection to another communication model - *framing* (McCombs, Llamas, Lopez-Escobar & Rey, 1997).

Coming up with exact definition of framing *"[...] has been notoriously slippery"*. (Boydstun, Gross, Resnik & Smith, 2013, p.2). Widely accepted one in the [political] communication field, which also resonates with the current thesis comes from Entman (1993, p.52), as of whom *"Framing essentially involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described"*. It represents a technique, where an issue is explored from a certain perspective, while [purposefully] excluding alternative ones (Boydstun et al., 2013). Coupled with agenda-setting, the potential of framing to call the attention to certain details turns it into discrete instrument for manipulation (Field et al., 2018). Adequate example of framing would be the image of contemporary Russia in American press. In a study for the San Francisco State University, Tsygankov (2017) analyzed editorials for the period between 2008 and 2014 of the leading U.S. newspapers, which have their focus on Russia's internal politics. Alongside the dominant negative image that Russia was awarded with, the author identified key frames used by the press to establish portrayal of modern Russia as neo-Soviet autocracy. To maintain that stance, aspects and facts that did not fit into that narrative were excluded. Semetko and Valkenburg (2000) also found widespread usage of frames in the Dutch national news media, when covering european politics.

It might be surprising, but many studies before 1977 already hint at significant correlation between the salience of attributes in media and the public, without being exclusively meant to investigate such a dependence. However, the work of McCombs and his fellow researchers has been specifically designed to compare the relationship between political candidates' image in mass media with the one perceived among voters. While political parties would select one set of characteristics to establish the *image* of their nominee, media might go for another one in their stories, and since it is highly likely for majority of the public to learn about candidates through the media, their representation might later impact recipients' voting behaviour.

Another feature of the second level agenda-setting is that the attributes of the objects communicated, or the candidates, as in the case of McCombs' study, should be further analyzed in *two dimensions - substantive and affective*. **(1)** *The agenda of substantive attributes* includes the selection of facts about the objects, or as in the elections example - the ideology, issue position, qualification, experience and even personality of candidates. **(2)** On the other side, *the agenda of affective attributes* covers the way these substantive attributes, facts, are presented, the tone used, their valence - positive, neutral or negative one.

The formal study derived some curious conclusions - overall, media reported overwhelmingly positive, with state-controlled media staying largely neutral, while newspapers exerted greater influence than television. The most valuable takeouts for the extended agenda-setting model however were: **(1)** There is sufficient evidence to confirm second level agenda-setting on *both the affective and substantive dimensions* on how voters interpret the candidates and **(2)** the affective dimension features greater effects. Results indicate the voters' affective descriptions to firmly echo the ones of the newspapers articles (McCombs et al., 1977).

To summarize, agenda-setting as presented here postulates two effects that media has on recipients. On its first level, topics salient on the media agenda become salient on the public agenda. On its second level, the characteristics used by media to portray these topics, are later being reflected on the public agenda. This is valid for both the selected substantive attributes, i.e. the supposedly objective facts on the topic, and the neutral, negative or positive valence they have been described with. To give an example, there is a high likelihood that a mayor's vision about the city will be perceived by the citizens the same way it was presented by media, along with the expressed negative, neutral or positive tone of description. There is of course some criticism, yet the first level agenda-setting theory has been backed up by numerous empirical studies, on various political topics, not limited to certain media channel, decade or region (McCombs et.al., 2000), and there is growing interest towards the attribute agenda-setting concept to be observed too (Golan, Kiousis & McDaniel, 2007). It also should be underlined, that neither of the concepts is limited to the field of politics. They could very well examine the perception of foreign nations (Wanta, Golan & Lee, 2004), religion issues (Bowe, Fahmy & Wanta, 2013) or media coverage of natural disasters (Cutter et al., 2008).

MEDIA AND ITS FINANCIAL IMPACT

Media has been said to impact the financial world too. Notable market events occur when there is homogeneous way of thinking among a larger group of people, and media is the one spreading the ideas. Many may be surprised that financial and sports news sometimes comprise 50% of the newspaper articles, however the financial market is a lucrative source of news - it not only has daily movements across various assets, but is also the place where fortunes are being made and lost, granting it a bit of sensationality (Shiller, 2000, p.71 - p.95). One could argue, this may be a reason why there is noticeable growth of business-related news in mainstream media, with the coverage of NYSE (New York Stock Exchange) doubling in a decade (Carroll & McCombs, 2003). British housing market for example, received increased media attention of up to 1200% during its booming years (Walker, 2014). The author also concluded for media to have Granger-caused changes in house valuations, with news tonality likely influencing future returns' expectations. Overly optimistic beliefs about the prospective house prices might have pushed the American housing bubble earlier too (Foote, Gerardi & Willen, 2012). The study of Walker (2014) builds on t*he two-step flow of communication theory*. In yet another politically related work, Lazarsfeld, Berelson & Gaudet (1944), observed two type of people - some admitted their voting decision to have been influenced by mass media, consisting of print and radio at the time, while the rest, especially those who decided for a candidate at the very late stages, were largely affected by people in their close(d) circles. The first group of people, who also are more active in political discussions, could be identified as *opinion leaders.* Information flows towards them and is subsequently resonated to the less active part of the public. As of Walker (2014) the British housing market might have suffered from media influencing the opinions leaders, which then transmitted the agenda to their followers. It is important to note, that Walker does *not* speak of agenda-setting, however media influence and sentiment is a prime concern of his.

Sentiment plays a major role in both of the studies by Kräussl and Mirgorodskaya too. They hypothesize that investors sentiment mirrors media sentiment. Their earlier work concludes that "*[...] news media can have a prolonged effect on market sentiment and on long-term financial performance [...]*", with pessimistic media views causing downward market pressure for up to 24 months (Kräussl & Mirgorodskaya, 2014, p.17). Their later study, based on theories from behavioral finance, provided evidence for media sentiment to impact market performance with a lag of up to 25 months (Kräussl & Mirgorodskaya, 2017). Ruscheinsky, Lang, and Schäfers (2017) report for 3 – 4 months of delay between positive / negative real estate media sentiment change and respectively the upward / downward returns on the market. Tetlock, Saar-Tsechansky and Macskassy (2008) also note a bit of time is needed until the stock market price reflects the negative word content of firm's coverage. To deviate a bit, as from our previous pages, this lag can be due to the physical time needed, even though largely undetermined, for opinion-leaders to influence their *"followers"*. What is more, opinion leaders themselves need to be influenced by the news media on the first place. How long would this take, is also unknown, however as of Carroll and McCombs (2003, p. 37): *"[...]repeated attention to an object day after day is the most powerful message of all about its salience",* suggesting that agenda transmission does not happen overnight. This rather valid assumption is contrary to the widely-criticized *efficient market hypothesis,* that the price fully and nearly instantly incorporates all available information, but is in line with the view of conservatism in human information processing by Edwards (1968), who suggests people's beliefs are gradually adjusted to the new information available.

Tetlock (2007) concludes that high media pessimism predicts strong downward pressure on prices, which then return to their genuine fundamental level. As from the bubble chapter earlier, Tetlock's findings could be formulated in a way, that extremely negative sentiment expressed through media may cause the market to experience a temporary *“negative bubble”*. The author also discovers extremely high or low levels of media pessimism to cause high trading volume, and that high media pessimism is preceded by low market returns. While Tetlock investigated relationships within the US stock market, Yoshinaga and Junior (2012) had their focus on the Brazilian one. Their results, as opposed to the ones of Tetlock, statistically confirm that the levels of returns are higher after a period of negative sentiment, than after one with positive sentiment. Bathia’s and Bredin’s (2013) study also confirmed that high sentiment is followed by low future returns and vice-versa. As of the Chinese stock market, positive media reports tend to lessen the likelihood of it crashing (Zhu, Wu, Zhang & Yu, 2017). Cahill, Wee and Yang (2017) noticed a contrast of how media sentiment affects institutional and retail traders and their reaction to good and bad unexpected earning news. Strauß, Vliegenthart and Verhoeven (2016) on the other side, found newspapers to rather react to movements, with their reports consisting of more negative words after the price has increased. Barber and Odean’s (2008) results attest for individual, layman investors to choose stocks that have simply caught their attention, such covered in the news, experiencing abnormal appreciation or trading volume. A large number of similar investors, making purchases based on awareness, might temporary inflate the stock’s price, causing a bubble. Study on the Indian stock market, suggests that media likely features bigger effect than technical and fundamental analysis, when taking a buy or sell decision (Malhotra & Malhotra, 2012). It was not long before corporations acknowledged, or at least suspected of mass media potential to affect investors and price movements. Assigned Investor Relations (IR) professionals cautiously plan the timing and content of their company’s communications (Nielsen & Bukh, 2014). Properly executed IR activities positively affect share price, promote genuine media attention (Bushee & Miller, 2012), and company’s messages are granted with greater credibility, when shared through media, rather than when communicated using press releases (Carroll & McCombs, 2003).

MEDIA & BUBBLES

There are relatively few works which directly look at the probable relationship between mass media and various speculative bubbles. One of them outlines that bubbles are more common through assets which get public attention based on their price rise. A basic observation is that media pays more attention to the movements of some assets while neglecting volatility in others, essentially causing "hype" for the former ones – a sudden surge of 10% in stocks would draw vast media awareness than similar increase in textiles (Chinco, 2018). For its role, Shiller (2000, p.95) calls media *"[...] fundamental propagators of speculative price movements [...]".* Apart from Walker (2014) and Foote et.al., (2012), Mercille's (2014) focus is also on the property market. His findings point at news organizations largely assisting the Irish bubble existence till its implosion, sending the country in financial crisis. Besides having to meet the needs of its real estate company advertisers, Irish mass media shared views similar to the corporate and government elite. Coverage of possible bubble existence was negligible and mostly negating it. Squires (2012) also blames indifferent media for not giving signals about the inflating bubble in 2007 and 2008. Page (2002, p.50) criticized publishers for perhaps purposefully, overlooking what is newsworthy and by that failing to report on *"[...] the collective insanity of our ruling financial elite".* Stock bubbles are known to be positively correlated with the sentiment expressed by investors, which leads to the build-up and evolution of their price deformations (Yao & Liu, 2018). It is namely mass media however, that dictates investor sentiment, affecting their trading decisions. Another valuable observation that likely affects speculators' judgement is that they tend to ignore negative news in bullish market. (Yang, Lin & Yi, 2017). Media was not a crucial factor in altering the investor sentiment during the Railway Mania in the mid-1840s, but managed to stay responsive, providing factual information (Campbell, Turner & Walker, 2012). As of Bhattacharya, Galpin, Ray and Yu (2009), media is not to be blamed for the Internet bubble either, however Internet-related stocks coverage was noticeably more positive during the rise and more negative in their deflation. Their results indicate for untypically high returns to cause more positive media coverage on the subsequent day.

RESEARCH INTERESTS

Bitcoin might have started as a monetary experiment, but it did expand beyond its initial small group of supporters. There are clear signals that its core, the decentralized ledger, is growing with healthy temps, and increased attention from regulators, funds, institutional and individual investors as well as users is to be observed. The aspects of such fintech advancement, that could theoretically become part of everyone's life within minutes, should rightfully so be explored in detail.

One of BTC's distinguishing characteristics is that it is tradeable and highly volatile asset. In its latest bull run, which ended with bubble implosion, Bitcoin went from just under 165$ in January 2015 to close to 20,000$ in December 2017. There are a number of theories within the fields of finance and economics, that are tailored towards explaining such phenomena. This thesis however, attempts to investigate the role of media in Bitcoin's bubble (formation). Media is a powerful actor, and studies show that it can dictate both what its audience to think about (1st level agenda – setting) as well as how to think about it (2nd level agenda – setting). Both of these theories and other types of media influence have been studied in various fields including the world of finance.

The research interests of this paper are constructed gradually, in order to assure a build-up. The first three research questions are:

RQ1: What are the general frequency characteristics of Bitcoin – related publication during the period of study?

RQ2: What are the general sentiment characteristics of Bitcoin – related publications during the period of study?

RQ3: What are the general bubble – implication characteristics of Bitcoin – related publications during the period of study?

These will be answered with the help of descriptive statistics. The aim here is to roughly explore the distribution of these implied variables, such as the months with the most news or the most positive news article genres.

Considering the studies and conclusions introduced earlier, in a similar fashion to Kräussl and Mirgorodskaya, this thesis assumes the overall validity of 1st and 2nd level of agenda setting, and that sentiment conveyed in media dictates the one of investors and traders. As already discussed, sentiment has been found to affect valuation in various ways. For example, excessively optimistic reporting supposedly caused the American and British housing markets frenzies (Foote, Gerardi & Willen, 2012; Walker, 2014) and Tetlock (2007) concluded for extremely negative polarity to lead to negative bubbles. Bathia and Bredin (2013) on the other side found unfavorable reporting to result in higher future returns and also vice – versa (positive attitude to cause lower returns), view, confirmed by the results of Yoshinaga and Junior (2012) too. What is more, there is some lag to be expected, since agenda and attribute transmission does not happen instantly, as per Saar-Tsechansky and Macskassy (2008), Ruscheinsky, Lang, and Schäfers (2017) and the studies of Kräussl and Mirgorodskaya. Having in mind all what has been covered so far, the fourth research question says:

RQ4: To what extend can negative / positive sentiment loaded Bitcoin – related articles predict Bitcoin's price action?

Exploring this one in details, requires the right type and amount of data. Having in mind that lag is to be accounted for, and that the models that looks into such type of interdependencies have forecasting nature, the time intervals should be regular, with equal gaps inbetween. Stocks are tradable Monday to Friday, with occasional holidays, which limits as to what the gaps are. The trading of Bitcoin on the other side never stops. Data being present for every single day, allows for more flexibility when it comes to the duration of periods – 8 days of news, followed by 5 days of no news, or 4 days of news, followed by 4 days without news. Unlike most of the other studies within that field, the majority of this one is to be completed by a single researcher with a set deadline, which would allow for one online medium to be properly reviewed. This issue is discussed further on the coming pages.

Since there is no guarantee that the data to be collected would be sufficient for the fourth question to be researched, a one more is included as precaution:

RQ5: What importance does sentiment of Bitcoin – related news have for Bitcoin – price action?

In order to extract feature importance for this fifth question, supervised, machine – learning algorithm is trained and applied, which functions in a different fashion and is not to be affected by irregularity.

Any other valuable results are also to be accounted for.

METHOD

Sentiment analysis stems from the studies of public opinion at the beginning of the 20th century, however the general curiosity towards opinions expressed by others is likely to date back as far as the emergence of verbal communication (Graziotin, Kuutila & Mäntylä, 2018). It precedes even content analysis, acknowledged as one of the most important research techniques in social sciences, which notion is suspected to have begun with the first conscious use of symbols and writing (Krippendorff, 2004, p. xvii).

Sentiment analysis as known nowadays, is "[...] t*he field of study 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, 2010, p.1) It is a method for extraction of subjective information such as polarity of opinions and evaluations, the not sot manifest perspective of the author towards the object of discussion in various communication sources (Hoffmann, Wiebe & Wilson, 2009; Kumar & Sebastian, 2012). Often referred to as *opinion mining* too, the technique is preferred by those, concerned with the way sentiments are expressed, whether the content conveys positive, negative or neutral stance (Nasukawa & Yi, 2003). During most of the 20th century, the method enjoyed only minor popularity and when so, primarily with political use case (Graziotin, Kuutila & Mäntylä, 2018). That however changed in the early 2000s, when a whole new wave of interest studying the [dis]advantages of sentiment analysis occurred. The improvement of machine learning methods within NLP (natural language processing), growth of the World Wide Web, opinion-sharing websites and the potential benefits of their real – time analysis were all factors for the sudden rise of academic interest towards opinion mining (Lee & Pong, 2018). Graziotin, Kuutila and Mäntylä (2018) also note that while exploring opinions started to become a topic in 1940, it picked up traction in the 90s, with majority of the papers being published only after year 2004. Despite the fact the technique has been properly honored for less than 2 decades, it went through a major shift from analyzing simple product reviews, to the successful application of sentiment analysis into predicting stock market movements and studying depiction of political actors.

Likely reason the rise of awareness for sentiment analysis to coincide with the advent of personal computers and the internet, is that collecting and analyzing something as banal as consumer opinions using the conventional survey method is expensive and time-consuming (Younis, 2015). While in the late 90s and at the beginning of the 21st century computers were mainly used to ease the hurdle of manual coding entries, later on, attempts began to fully *automate* the process of analysis - from gathering the vast amount of available opinions, to the categorization of their polarity (Teraiya & Vohra, 2013). Few other driving forces are to be considered too – for example, to identify the degree of viewpoint’s favorability is a task, that requires a certain level of intelligence, common sense, topic and linguistic knowledge, and it is not rare for people to struggle when interpreting judgements. That latter detail turns into significant issue when there is an entire group of documents to be categorized, since consensus for the exact polarity of each text is hardly achievable, and this even within a small group of evaluators. These are all reasons, which lead to believe that eliminating the human factor is the way to achieve better overall results and higher objectivity (Nasukawa & Yi, 2003). In fact, nowadays, the majority of authors in the field emphasize on the *computer – assisted* characteristics of the method. Torabian (2016, p. IV) for example defines sentiment analysis as “[...] *an application of natural language processing and computational linguistic [...] to capture the evaluative factors such as positive, negative, or neutral, with or without their strength, from plain texts.*” Similar understanding, but with varying wording and style has been expressed by Han, Yang, Zhang, Zhang and Zou (2018), Ebrahimi, Shteth and Yazdavar (2017), Gupta and Kaur (2013) as well as Gama and Rambocas (2013).

Generally speaking, throughout the years two broad computer – assisted methods for sentiment analysis have been developed – (1) *lexicon – based approach* and a (2) *machine learning* one. The *lexicon – based method*, in simplest terms, functions as a dictionary. Each word [phrase] in the dictionary / the lexicon, is associated with certain sentiment orientation – negative, neutral or positive. This technique assumes, that the polarity of a text, would it be a tweet or a whole document is the aggregate sum of the polarities of each word [phrase]. Nowadays there are a number of such lexicons – LIWC, GI (General Inquirer) and SentiStrength to name a few – build to evaluate general language. (Araujo, Benevenuto, Cha & Goncalves, 2013; Han et al., 2018).

The machine learning methodology on the other side, functions in a bit more complicated fashion. It requires data, for example movie reviews, which have been manually annotated. One of the main challenges associated with this adaptable technique however is, the requirement for algorithm, sophisticated enough so it can go through the data, the reviews provided, and analyze them. Subsequently, when it [the algorithm, the automated system] is presented with new, unique ratings, it shall be able to pinpoint them as either negative, neutral or positive, and this with impeccable accuracy (Araujo et al., 2013; Biswas & Bordoloi, 2018; Llombart, 2017).

Some however, consider foolish the suggestion the type of analysis people do to be achievable on a much larger scale with the help of automated technology (Duarte, Llanso & Loup, 2018). While there are studies sharing rather positive outlook towards computational sentiment recognition (Kundi, Khan, Ahmad & Asghar, 2014; Filho, Almeida & Pappa, 2015), some valid counter arguments exist. First and foremost, the complete automation is not fully present [yet]. For both of the methods exists the difficulty to find matching and extensive amount of human – labeled corpus (Tai & Kao, 2013). Roth, Ruppenhofer, Schulder and Wiegand (2017) for example, manually annotated 2000 English verbs. Alkorta, Gojenola, Iruskieta and Taboada (2017) created a dictionary, which got complied by hand too. Five people were needed by Schmidt and Burghardt (2018) to classify 200 speeches, which were later used to train their tool, while Guerini and Staiano (2014) decided to crowd-source the categorization of their 37 thousand emotion terms. Systems emerged that allow for the sentiment of words to be defined in an unsupervised manner, however their accuracy fluctuates greatly and depends on a number of variables (Hatzivassiloglou &

McKeown, 1997; Turney & Littman, 2013). Highly problematic here would be the polarity strength detection for each word – for example, *"beautiful"* and *"good looking"* convey the same meaning, but *beautiful* features more pronounced emotion (Kaushik & Mishra, 2014). Even if we assume the existence or future compilation of extensive dictionary, one should consider that words may vary in meaning depending on context (Duarte, Llanso & Loup, 2018). The conundrum on how to handle negation within the lexicon – based approach remains (Edison & Aloysius, 2017), and general-purpose dictionaries, which happen to be the most widely used, are known to be greatly biased, skewing final results (Goncalves et al., 2014).

These drawbacks are not exclusive, with most of them valid for the machine – learning method too. While the approach is more adaptable in its nature, it requires one training set for the algorithm to study the various sentence / document features, and a test set to verify its [the algorithm's] proper functionality (Thelwall & Prabowo, 2008). One of its main setbacks is the great cost of coming up with sufficient, both in quantity and quality, human – labeled data, but also the fact that even if the automatic classifier ends up being precisely refined, it can rarely be applied to new, fully unique set of data (Araujo et al., 2013). Furthermore, the degree of real – world reference strictly depends on the thoroughness of the corpora applied to tune the decision – making system (Ahmed et al., 2018). Valid example here would be the language use, which varies as of age, education and social status, with the theoretical underrepresentation of minorities in the training data to result in bias (Duarte et al., 2018). Some researchers, recognizing that neither of the methods is polished enough, decided to experiment using extensions from both approaches. Hasan et al., (2018) are supposedly the first ones to validate three popular sentiment analysis lexicons with two adaptable algorithms. Filho and Pardo (2013, p.571) combined three different approaches (rule – based, lexicon – based and machine – learning one), and their modest but encouraging results lead them to the belief that *"[...] hybrid techniques might outperform the current state-of-art in sentiment analysis"*. With another experimental system and model, Fouad, Gharib and Mashat (2018) managed to achieve accuracy of over 90% to tweets manually annotated as negative / positive from the Sanders Twitter Corpus.

Put in a nutshell, sentiment analysis, the way it is perceived and used by most of the researchers nowadays, is the computationally – assisted gathering and analysis of opinions, appraisals, emotions or as in the case of this master thesis, the sometimes disguised negative, neutral and positive valence towards the object of discussion – would it be a product, individual or certain topic. It is seen as the only viable solution for the timely interpretation of the vast amount of opinions made available today, however it is far from being the optimal one. As discussed, it is possible to improve the level of accuracy by combining various methods, however, this is a task that requires knowledge in mathematics and coding skills well above the average. One of the main issues is, that once developed, an algorithm is almost never reliable combined with new data or varying domains (Ahmed et al., 2018). Last but not least, tools specifically built to fit certain content are costly endeavour, and the available lexicons complied with general language use in mind, which are likely to [financially] appear to smaller enterprises or individuals, are liable to deliver biased results (Duarte et al., 2018). Conducting sentiment analysis in the old – fashioned way, primarily by hand, from the collection of materials to their categorization, is labour – intensive and equally expensive. This would have been a fair trade – off if validity was guaranteed, however, there might be disagreement among the annotators regarding category and weight of polarity, which comes on top of the sample-size limitations (Nasukawa & Yi, 2003). Yet, by sacrificing the speed and controversial adaptability of the automated methods, researchers do not have to worry about the problematic detection of fake postings (Siddharth, Darsini & Sujithra, 2018), or their likely failure to properly handle satire, compassion and emoticons among others (Desai & Mehta, 2016). Another advantage for the manual annotation comes from the regular inclusions of new informal and slang terms in our language, one of the many sticks in the wheel of precise automated classifiers (Joshi, Prajapati, Shaikh & Vala, 2016). The greatest acknowledgement for the manual sentiment analysis however, is that the reliability of (semi) – automated systems for polarity detection is often judged based on comparison to human – annotated databases (Fouad et al., 2018; Ghiassi, Skinner & Zimbra, 2013).

Sentiment analysis is widely experimented with among those, who strive to find out meaningful correlations within the financial market. Lima et al. (2016) for example, aimed to understand the shared mood towards an asset, and algorithmically predict investors behavior. The team behind the study faced difficulties with tweets rich in irony and neologisms, concluding that despite the continuous research in the field, automated classifiers should only be considered as complementary in the decision – making process. In a separate study, Vijay, Singh and Malhotra (2018) built a system for the automatic extraction and categorization of news into 3 predefined polarity dimensions. Their research demonstrated stock market returns to be greatly affected by online media, with shocking news causing volatility, even when the information does not concern the company's fundamentals.

It is reasonable to take a look at the sentiment analysis methods, employed by the studies cited in the chapters *"Media and its financial impact"* and *"Media & Bubbles"*. The work of Bhattacharya et.al., (2009), that closely explores the role media had in the Internet IPO (Initial Public Offering) bubble, stands out due to the amount of man – hours required for its completion. Following their defined research criteria, the authors *hand – collected* over 171,000 relevant news pieces for the period from 1996 to year 2000. *Afterwards, they manually annotated the sentiment (bad, neutral, good) of every article, a process that took them a little over 2 years – from the fall of 2002 till late 2004.* As outlined earlier, these are the years when exploration of partially unassisted methods began. Campbell et.al., (2012), also decided to hand – pick the articles that relate to the British Railway Mania in the 19th century. For sentiment analysis they used one of the general lexicons, LIWC (Linguistic Inquiry and Word Count), that calculates positivity score based on the percentage of positive words included in the article, also annotated as favorable in the inbuild 4500 – word corpus of the tool. Additionally, the authors classified a good number of articles by themselves, firstly because language nowadays likely differs from the one used back in the 1840s, and secondly of the software limitations to recognize sentence structure, to distinguish reports of past events or future expectations. On the other side, Walker's (2014) study into the correlation between media coverage and the British housing boom relies solely on a dictionary tool – Diction 5.0, which functions in the same way as LIWC, but features extended corpora of 10 000 words. The author means its utilization guarantees correct assessment and lack of bias, since an evaluator's understanding of negative / positive changes with each article read. Walker (2014, p.3957) justifies the selection of Diction 5.0, by emphasizing that it *"[...] has recently been used extensively in accounting, business and finance literature"*. Strauss et.al., (2014) applied the default dutch version of LIWC to measure *emotions* in news, while Kräussl and Mirgorodskaya (2017) resorted to General Inquirer (GI), a Harvard – developed quantitative content analysis program similar to Diction and LIWC. They however, replaced the default built – in corpora with context adjusted one by Loughran and McDonald (2011), who found out that close to 75% of the terms included in the standard negative word list are not necessarily negative when applied to financial content. The same corpus, but adjusted for real estate, coupled again with GI has been used by Ruscheinsky et.al., (2018). The fame of General Inquirer rose with Tetlock (2007), who concluded a significant relationship between media pessimism and the price change of Dow Jones Industrial Average Index. He did not rely on the default settings of the lexicon, but fine – tuned them instead. Zhu et.al., (2017) studied how media reports in China might impact the expectations of local stock market crash. Since there was a lack of a fitting word – compilation in Chinese, the authors had no choice but to manually create one with the words from 2000 domain – related articles. Following that, they annotated the headlines of equal numbers of negative / positive statements, which were then used to train their machine – learning tool. The work of Cahill et.al., (2017) attracts attention due to the contrasting approach taken. Their data comes from Thomson Reuters News Analytics (TRNA) database, which provides comprehensive analysis on news, including sentiment classification in the categories –negative, neutral and positive.

Experiment was conducted with one of the automatic sentiment analysis tools, LIWC – a dictionary – based solution. As discussed, it has been successfully used in financial research, but also in dream analysis (Bulkeley & Graves, 2018; Nadeau, Sabourin, De Koninck, Matwin & Turney, 2006), changes in college textbooks over time (Sell & Farreras, 2017), as well as suicide notes (Garcia – Caballero, Jiménez, Fernández – Cabana & García-Lado, 2012). For this purpose, a license of the latest edition of LIWC was purchased. It consists of close to 6400 words and phrases, divided in multiple improved categories, and supposedly allows analysis of the so called internet language. Few randomly selected but relevant articles from the Financial Times website (the choice of FT.com is briefly discussed on the following pages) were imported into the software. *"There are many reasons to be cautious about Bitcoin"* is the headline of a rather critical analysis, which particularly stands, since it spreads misconceptions, makes parallel between BTC and the Dutch tulip mania and presents the cryptocurrency as an asset, which valuation is based on no fundamentals but wobbly trust. Strangely enough, the article in question was awarded by LIWC with a *tonality* of 84.02 points, where the closer the number to 100 is, the more positive, favorable the body of text is supposed to be. A score of ~50 would represent uncertainty, neutrality. The current word corpora of the tool had been complied back in 2015, way before the latest peak of Bitcoin or interest towards the field. This could be the reason why terms, present in the articles analyzed and often used in the field as well, such as "*bubble", "volatility", "fraud", "blockchain"* or *"database"* are not part of the lexicon's corpora. The phrase *"mining pools"* has been split in two separate words, where *"mining"* is sorted under no category, and *"pools"* is set under the *leisure* one. *"Bitcoin"* falls under the category of *money*, however its plural form lacks as an entry.

Bizarrely, *"crypto"*, *"cryptocurrency", "cryptocurrencies", "cryptographic"* and *"cryptic"* have been sorted in the list of *death - related words,* while *"geek"* is considered a swear. It is unclear, at least relative to the topic of this thesis, why the publishers of the software have adopted such an approach towards these words, however, one should not forget that LIWC, like other similar solutions, is simply a probabilistic text analysis tool, that cannot differentiate irony, sarcasm, idioms or understand context (Tausczik & Pennebaker, 2010). That is problematic, considering there are various discrete ways to express sentiment. For example, the phrase *"Who would buy Bitcoin?"* consists of no obviously negative words, yet the opinion conveyed is highly unfavorable (Khoo, Nourbakhsh & Na, 2012).

It seems no universal, ideal methodology for sentiment analysis is at our disposal just yet. By going for manual annotation, one would sacrifice the speed and scope of (semi) - automated methods, which however cannot read between the lines like a person could. In any case there will be some kind of compromise. This thesis will adopt fully human – coded categorization. While it is significantly more time consuming and labour – intensive, it is the only way to pursue quality of polarity decoding, rather than quantity of news. Zhu et.al., for example note that *"[...] the most credible classifications could be to assign personnel to thoroughly read and analyse every news report [...]",* however, this would have been hardly doable in their case, with a collection of more than 4.5 million articles. Goidel and Langley (1995), who wanted to explore possible indirect media effects on economy, manually classified in negative, neutral and positive the front – page articles of The New York Times for the years between 1981 and 1992. It is a valid argument, that back in the 90s, automated sentiment analysis was not developed or easily accessible as it is today, however, in 2008, Gong and Gul evaluated the quantity and quality of Chinese media coverage and the impact it had on stock price, by manually coding the articles. For their study, Kalogeropoulos, Svensson, Van Dalen, de Vreese and Albaek (2014) employed three annotators, who had to follow extensive codebook, while De Bruycker and Walgrave (2014) evaluated by hand 24,500 articles covering the financial crisis in Belgium. *In that sense, the current master thesis will not be an exception, but rather part of a smaller trend in content analysis and the study of opinion.*

PERIOD OF STUDY, MEDIUM OF INTEREST AND NEWS SELECTION PROCESS

The period of interest for our study is between 14th of January 2015 (14.01.2015) and 17th of December 2017 (17.12.2017). As illustrated and discussed in the *Bitcoin Bubble(s)* chapter, that is the timeframe that extends starting at the bottom from previous top, until the bubble peak of importance for this research. The medium of interest, the one to collect the articles from is the website of the popular daily financial outlet, The Financial Times – www.ft.com The publishing group describes itself as *"[...] one of the world's leading news organisations, recognised internationally for its authority, integrity and accuracy."* Dyck, Volchkova and Zingales (2008) concluded the credibility and influence of The Financial Times is to be unmatched by the alternative news sources. The web portal of the publishing house shares in a similar fashion the notion of objective reporting and the delivery of high – quality news, market analysis and commentaries across various industries (Reeb, 2010). The decision to select the web as primary source of articles, is that newspapers are notoriously known to provide outdated information when it comes to finances (Davis, 2005), with printed outlets reflecting rather than shaping stock market developments (Scheufele, Haas & Brosius, 2011).

The news selection process began by using the inbuild search engine in FT.com and looking for relevant articles using the keyword *"bitcoin"*. Overall, there are more than 2500 results, however this number drops to 1056 once the timeframe of interest mentioned above is applied as a filter. Worthy remark here is that these are all findings, *where the keyword is simply present* – it could be a link, sentence or something as small as a caption. Financial Times allows for further filtration of the results with predefined categories such as *"Markets", "Bitcoin", "Currencies", "Cryptocurrencies", "Technology sector", "Blockchain", "Fintech" and others.* While in one way or another all these deem to be relevant in our case, no further information is available about their categorization reasoning. What is more, a deeper look shows that one and the same article may fall under multiple of these labels. This is the main reason behind the decision for manual filtration of *all* 1056 results, despite the lengthy manual labour associated with it. Another person, with profound interest and knowledge in the field of blockchain and cryptocurrencies, volunteered to assists in the empirical research of this thesis. We separately examined the 1056 results one by one. By default, FT.com sorts the results with the latest one on top, however we deliberately started with the oldest one first, i.e. from January 2015. The rationale for that was purely suggestive – there could be developing stories during the timeframe of analysis, and better judgements could be made when following the way they unfold. In retrospect, this was the right decision to take, since multiple such cases were encountered – the ruling against Silk Road's mastermind, the MtGox case, the Bitfinex hack, regulations, rumours and launch of futures and others.

The aim of the manual filtration process was to select only these publications, which happen to be relevant to Bitcoin, to have the cryptocurrency or a meaningful related matter as its main topic. One could argue, that purely mentioning or referring to *"Bitcoin"* a certain number of times, for example ≥ 5, would signal its relevance to the topic, however that is not always the case. The stories *"Silk Road trial sheds light on dark web"* and *"Ulbricht found guilty over Silk Road drugs site"* have each 7 mentions of the keyword in their body of text, however, Bitcoin is just a side matter to a different main story – that Silk Road, the anonymous marketplace on the dark web, had the cryptocurrency as its main mean of payments. These and similar articles are not substantial for this research, since they *do not convey any sentiment or meaningful information towards our object of interest. At the end of the selection process, a total 244 (23.11%) publications were found to be fitting.* 76.89% (812) of the initial 1056 results within the specified timeframe, were publications of no importance including few podcasts, which were left out too, since audio stretches out beyond the characteristics of written text.

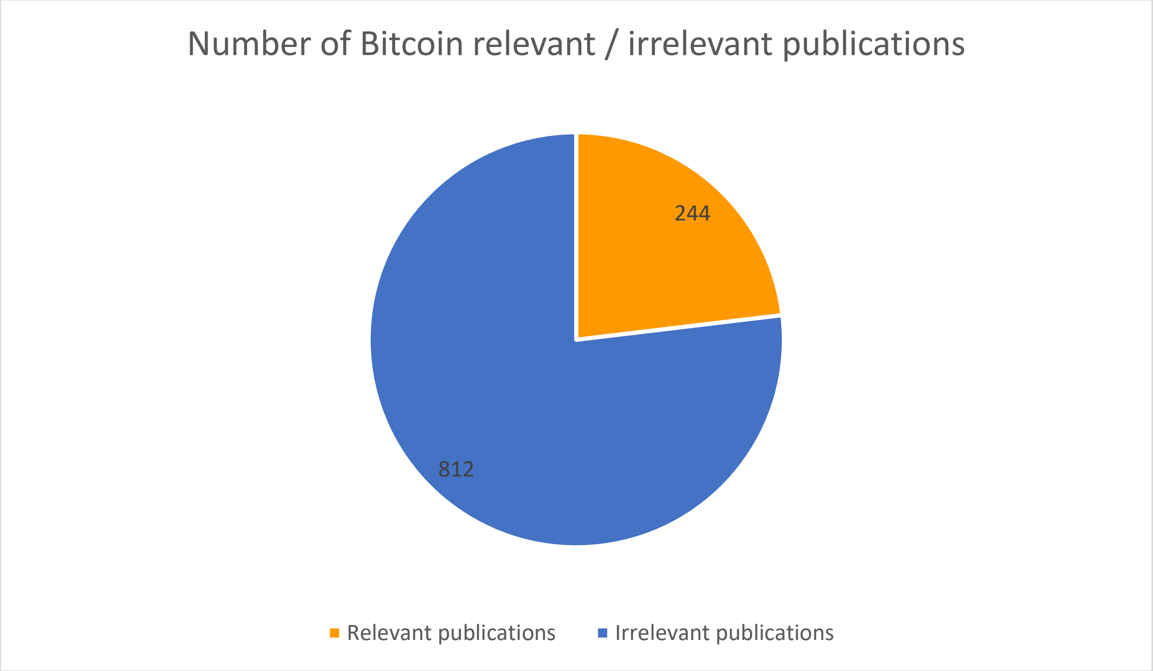


Figure 4. Number of relevant / irrelevant BTC - publications for the period of study, on FT.com, that meet the search criteria for the keyword "bitcoin". Total number of found articles : 1056

Following step was the manual annotation of the selected articles according to predefined variables. Full list, explanation and reasoning behind the variables is to be found in the appendix of this thesis, in the codebook – essential document meant to guide and make sense of qualitative data. A detailed codebook with clear definitions would minimize discrepancies among coders. It is hard to establish one correctly from the get – go, and it is not an exception for changes within codes or their respective definitions to be made once researchers start going through the data (DeCuir-Gunby, Marshall & McCulloch, 2011). It was partially the case in this study as well. The second coder went through part of the articles, to ensure consensus within sentiment categorization and clear out ambiguities. For the reader to be able to make sense of the pages to come, without the need to study the codebook in details just yet, some of the *more important variables* include:

1. **Date** – The date when the article was published
2. **Title sentiment** – Sentiment of the article's title – *Negative, Neutral or Positive*
3. **Article sentiment** – Sentiment of the body of text, of the publication itself – *Negative, Neutral or Positive*

4. **Bubble Implication** – Whether or not the article suggests, that Bitcoin might be in a bubble state – No or Yes (Article being labeled as “No”, does not mean it explicitly mention that Bitcoin is not in a bubble!)
5. **Length of the article in words**
6. **News Article Genre** – Defines, what type the publication is – *News, Article, Analysis, Interview or Other*
7. **Alphaville** – Whether or not the article is part of Alphaville – the blog of FT.com’s financial team – *No or Yes*
8. **Bitcoin Price information for every day where at least one article was published** – (1) Daily Open, (2) Daily Close, (3) Daily Low, (4) Daily High, (5) Daily Average, (6) Daily Change – difference between Daily Open and Daily Close, (7) Percentage Change – Daily Change converted in percentage (%) and divided by 100, and last but not least (8) Change Tag – categorization of Percentage change in 4 levels, ≤ -5%, ≤ 0%, ≤ 5%, > 5%. For fuller explanation and the logic behind each of them, please refer to the codebook.

This study relies on the *Bitcoin Liquid Index (BLX)* for Bitcoin price information. Since Bitcoin is a highly volatile market with debatable liquidity, the index was developed with aim to be the reliable source of Bitcoin’s fair USD value. It sources real – live data from the worlds most trusted and liquid exchanges, that attract high volume. BLX operates under tight requirements, and is found to be the industry standard for Bitcoin’s current valuation (BraveNewCoin, n.d.)

STATISTICAL CALCULATIONS AND FIRST RESULTS

The first easily observable phenomenon the occasional character of articles for the period of study. The timeframe of interest (14.01.2015 until 17.12.2017) equals to 1068 days or 152 weeks and 4 days. The 244 news stories, which fall within the bubble timeline, have been published in 155 days, which means that **mere 14.51% of the days received some attention – worth publications. 63 weeks, which represent 41.18% of the total week count have no relevant publications.**

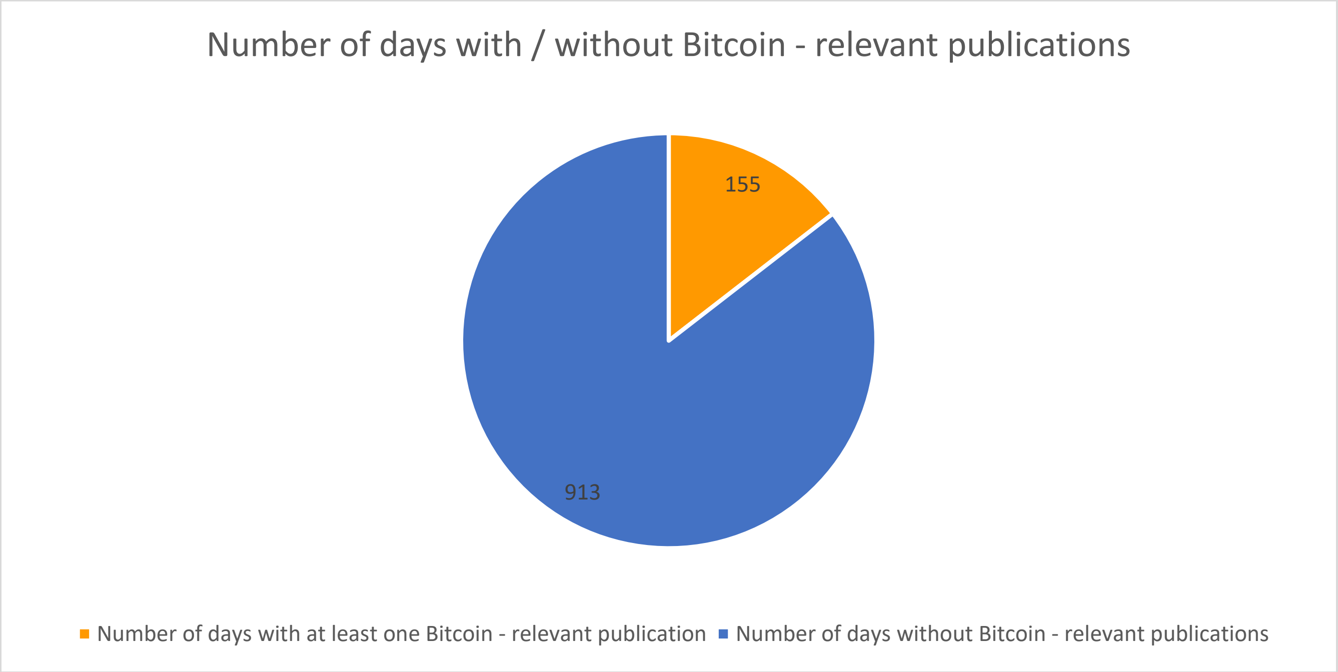


Figure 5. Number of days with / without BTC - relevant publications during the period of study, which equals to a total of 1068 days

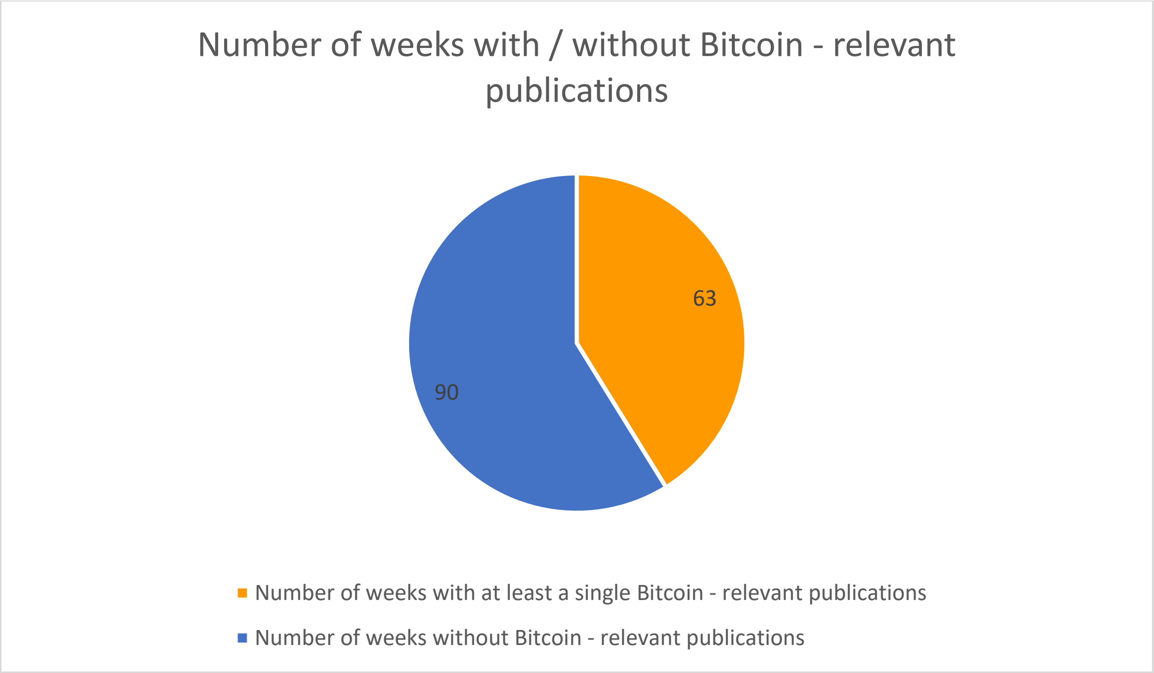


Figure 6. Number of weeks with / without BTC - relevant publications during the period of study, which equals to a total of 153 weeks

However, *no calendar(!) month lacks publication*, with multiple ones having a single relevant article. Curious observation here is the *"activity"* within the month of December 2017, and more specifically its first 17 days, since the selection period ended with the bubble implosion on 17.12.2017, which [December 2017] featured 49 relevant news stories, same amount as for the whole year of 2015.

Table 1. Number of articles for each month through the period of study

I	II	III	I	II	III	I	II	III
Jan 2015	3	1.23%	**Jan 2016**	**1**	**0.41%**	Jan 2017	6	2.46%
Feb 2015	4	1.64%	Feb 2016	2	0.82%	Feb 2017	2	0.82%
March 2015	5	2.05%	March 2016	2	0.82%	March 2017	7	2.87%
April 2015	4	1.64%	April 2016	3	1.23%	**April 2017**	**1**	**0.41%**
May 2015	4	1.64%	May 2016	6	2.46%	May 2017	11	4.51%
June 2015	4	1.64%	June 2016	3	1.23%	**June 2017**	**1**	**0.41%**
July 2015	9	3.69%	**July 2016**	**1**	**0.41%**	July 2017	8	3.28%
Aug 2015	**1**	**0.41%**	Aug 2016	12	4.92%	Aug 2017	11	4.51%
Sep 2015	6	2.46%	Sep 2016	2	0.82%	Sep 2017	24	9.84%
Oct 2015	3	1.23%	**Oct 2016**	**1**	**0.41%**	Oct 2017	11	4.51%
Nov 2015	3	1.23%	Nov 2016	3	1.23%	Nov 2017	24	9.84%
Dec 2015	3	1.23%	Dec 2016	4	1.64%	**Dec 2017**	**49**	**20.08%**
Total 2015	**49**	**20.08%**	Total 2016	40	16.39%	**Total 2017**	**155**	**63.52%**

Note: I = Month & Year; II = Total number of relevant articles for that month / year; III= % of the articles within that month / year from all relevant publications during the period of study

The overall lack of density of the publications and the monthly data presented above hint that there might be difficulties exploring the fourth research question. While it is true that such analyses is to be designed even with infrequent gaps and no clear seasonality, the methods are often an object of experiments and studies themselves, rather than a straightforward solution to use (Eckner, 2014; Gamberini, Lolli, Rimini & Sgarbossa, 2010; Hanzak, 2014). This issue is further analyzed in the chapter ”Discussion, Limitations & Further Research”.

Similar to the overall frequency of publishing, the distribution of news articles genres is also far from leveled out, with the two largest categories, *“story”* and *“analysis”* accounting for close to 82% from all relevant publications. *Combined, all genres feature mostly neutral titles and bodies of text,* with positive publications, less than 19%, being the least prominent. Further noteworthy fact is that *a whole 1/4th from all relevant articles are part of Alphaville* – the financial blog of The Financial Times, specifically aimed at finance professionals. 49 (or some 20%) out of the 244 selected publications, warn of Bitcoin being a bubble.

Table 2. Distribution of article genres as a number and percentage from all relevant publications

News Article Genre (NAG)	As a number from all relevant publications	As percentage (%) from all relevant publications
News	34	13.93%
Story	103	42.21%
Analysis	96	39.34
Interview	3	1.23%
Other	8	3.28%
Total	**244**	**99.99%**

Table 3. Distribution of articles in the categories Alphaville, Bubble implication and Premium as a total number and % from all relevant articles

	Articles, part of FT Alphaville	Articles, suggesting that Bitcoin might be in a bubble	Articles, part of FT Premium subscription plan
Total number	61	49	11
Percentage	25%	20.08%	4.51%

Table 4. Distribution of Negative / Neutral / Positive articles as a total number / % from all relevant publications

	Negative	Neutral	Positive
Title, Total number	66	117	61
Title, Percentage	27.05%	47.95%	25%
Body, Total number	86	112	46
Body, Percentage	35.25%	45.9%	18.85%

Table 5,6 and 7 show the characteristics of the various genres. For example, 56% of the news titles are positive, while close to 62% of analyses feature neutral headlines. Interestingly enough, there is a shift with over 60% of the articles of the latter type conveying largely negative sentiment towards Bitcoin, while close to 65% of the news stay neutral. Least discrepancies between title and text sentiment are observed among the stories. As discussed in the previous paragraph, 20% from all filtered publications imply, that the most prominent cryptocurrency might be in a bubble, with most of them (27 out of 49) being analyses.

Table 5. Article genres and their title sentiment distribution in a total number of articles / as % from all publications within that genre

Title Sentiment	News	Story	Analysis	Interview	Other
Negative	11 / 45.9%	23 / 22.33%	32 / 33.33%	1 / 33.33%	0 / 0%
Neutral	4 / 11.76%	44 / 42.72%	59 / 61.46%	1 / 33.33%	7 / 87.5%
Positive	19 / 55.88%	36 / 34.95%	5 / 5.21%	1 / 33.33%	1 / 12.5%

Table 6. Article genres and their body sentiment distribution in a total number of articles / as % from all publications within that genre

Body Sentiment	News	Story	Analysis	Interview	Other
Negative	3 / 8.82%	21 / 20.39%	58 / 60.42%	1 / 33.33%	3 / 37.5%
Neutral	22 / 64.71%	47 / 45.63%	36 / 37.5%	2 / 66.66%	5 / 62.5%
Positive	9 / 26.47%	35 / 33.98%	2 / 2.08%	0 / 0%	0 / 0%

Table 7. Bubble implication articles among the various genres, as a number / % from all publications within that genre

Bubble Implication	News	Story	Analysis	Interview	Other
No	30 / 88.24%	87 / 84.47%	69 / 71.88%	3 / 100%	6 / 75%
Yes	4 / 11.76%	16 / 15.53%	27 / 28.13%	0 / 0%	2 / 25%

The shortest and the longest publication share a lot in common. Both are part of the Alphaville blog and could not be sorted in any of the first four genres. While the longest article is almost 5000 words in length, conveys neutral attitude and is a strange form of reader – led discussion, the shortest one is mere 57 words long, but loaded with mockery and negativity. On average, an article has the length of 687 words, with the standard deviation being 539 – which means that most publications are between 148 and 1226 words.

First step towards a more concise exploration of the association between these various parameters, is to be accomplished with the help of *correlation matrix*. Correlation represents the relationship between two variables in the form of *correlation coefficient (r),* a number between – 1 and +1. Being in the negative, it would signify negative relationship, where one of the variables decreases with the other one increasing. A positive coefficient would mean that both of the variables increase simultaneously, while a perfect null (0) hints at no interdependencies (Gingrich, 2004; Pham-Gia and Choulakian, 2014).

For better structure of the chapter and to not hinder the ease of readability, the correlation matrix has been placed in the appendix, on page number 8. Gingrich (2004, p.800) himself notes that the size of the correlation coefficient (r) *"[...] can differ rather considerably depending on what type of data is being examined".* In order to avoid confusion, this research adopts the following scale:

- 0 to 0.19 – no correlation
- 0.2 to 0.39 – weak correlation
- 0.4 to 0.59 – moderate correlation
- 0.6 to 0.79 – strong correlation
- 0.8 to 1 – very strong correlation

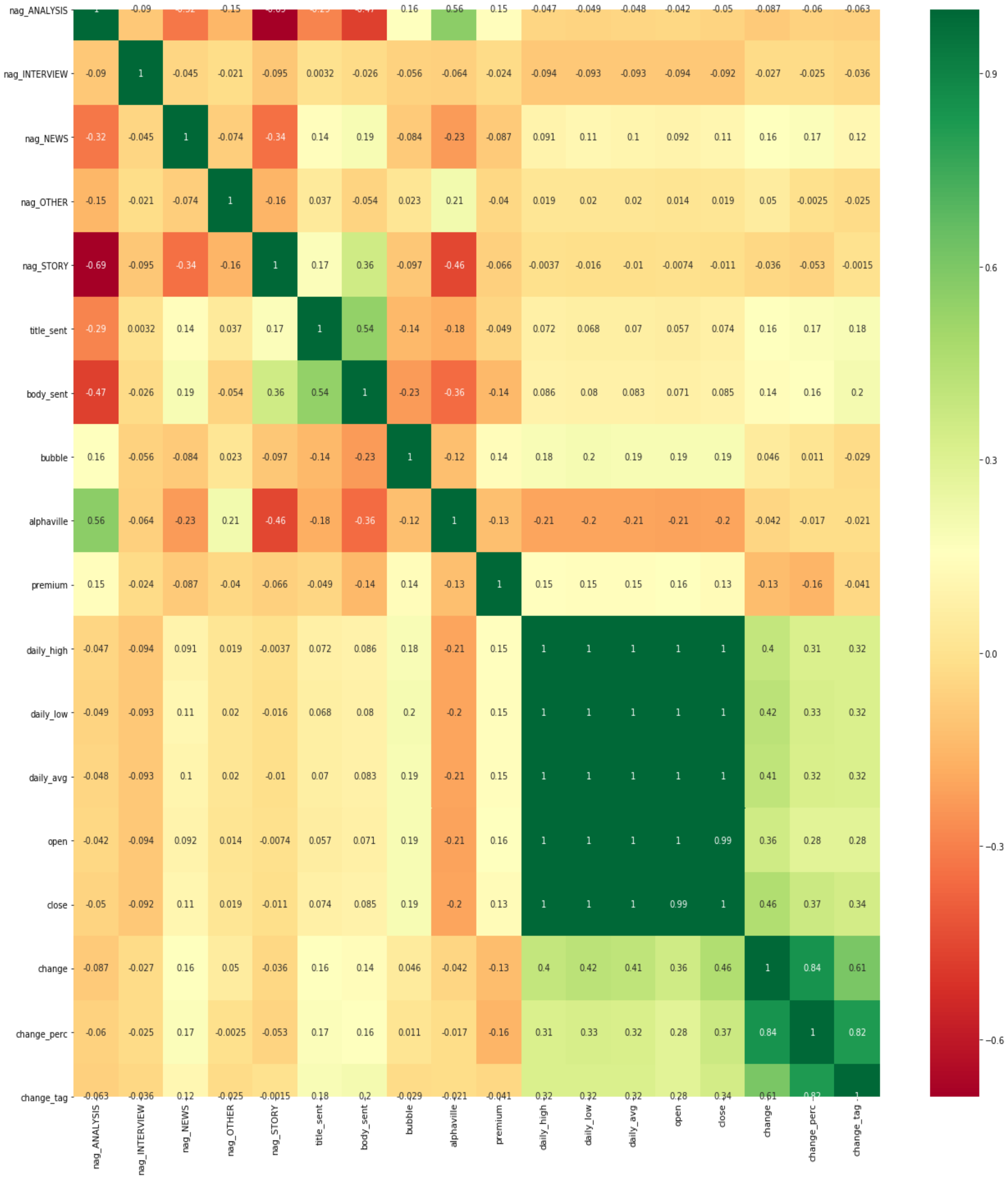


Figure 7. Heatmap with correlation - matrix between all variables

A glance at the matrix shows that there are only a few meaningful moderate correlations. For example, there is positive correlation of *r = +0.56* between a publication being analysis and part of FT's financial blog Alphaville. This cannot be said about the story article genre and Alphaville, where a negative correlation is to be observed, with a coefficient of *r = –0.46*. Both body sentiment and title sentiment are negatively correlated to the analyses article type, with the coefficient being respectively –0.47 and –0.29. Story on the other side has a weak (+0.36), but positive correlation with the body sentiment variable.

There are few interesting observations to be made when looking at price – related variables, however the featured correlation is weak at best. For example, the BLX index parameters – *(1) daily open, (2) daily close, (3) daily high, (4) daily low and (5)* daily average are *positively correlated to the bubble implication variable* with coefficients between +0.18 and +0.20. At the same time, all of these price dimensions are equally *negatively correlated to the Alphaville category*, with weak dependency of r being between –0.2 and –0.21, while *body sentiment and title sentiment show to be not correlated to these 5 price indicators*, with r being in the 0.00xx range. The two sentiment variables however, show higher, yet not even weak correlation to the three price – fluctuation parameters - *(6) absolute change, (7) change percent and (8) change tag*.

Some of the results so far hint at curious interdependencies, yet the issue with general correlations as the ones described above (issue for this thesis, not the method overall), is that *correlation does not imply causation*. While it is true that the significance, measured in *p-value*, of the meaningful relationships could have been tested, *the leading motivation behind this thesis is the extent to which the characteristics of the selected articles affect Bitcoin's price formation.* One way to observe such a relationship would be by relying on **regression** – widely used statistical method, that allows for interdependencies between at least two variables to be observed. In its simplest form, [linear] regression allows the researcher to set a **dependent variable (Y on the axis)** and **independent one (X on the axis)** and look how they both relate on the graph. For this thesis, the manually annotate articles' characteristics are the predictor variables, supposedly affecting the recorded price actions, the response variables. Once ran, a regressive model provides various outputs, such as the overall method's significance (*Significance F*), the already introduced p-value, which marks significance of the independent variable(s), as well as *standard error*, indicator similar to a *standard deviation.* Initially, most indicative would be the *goodness-of fit* *test, and its R2 (r-squared) coefficient.* It lays always within the range between 0 and 1, which is simplified percentage (%) representation of the cases where the dependent price formation could be explained with the help of the articles' attributes. R2 value of 0.525 would mean that the independent variable(s) are the likely cause behind 52.5% of the occurrences in the dependent variable. The higher the coefficient, the more meaningful the predictor's effect on the dependent variable is.

Numerous linear (with one independent variable) and multiple regression (with at least two independent variables) analyses were run. Five price codes served as response variables - [1] daily high, [2] daily average, [3] absolute change, [4] percentage change and the change tag one [5]. In separate linear regression models, the response of each and every one of these was measured against each of the following independent variables [a] title sentiment, [b] body sentiment, [c] bubble implication, [d] length in words and [e] news article genre. Multiple regression plots were run with [I] all independent variables combined, from [a] to [e], and [II] the purely sentiment – loaded predictors, [a] title sentiment and [b] body sentiment.

A total of 35 scenarios were evaluated, however none of them hinted at major predictors for any of the dependent variables. Inclusion of the plots and their parameters, for example in the form of tables, will result in unnecessary bulk for the thesis. The outcomes are available for review in the appendix with highlighted results for better visibility. The R squared values are far from satisfactory, with the greatest one being **0.078705**, which reads that **mere 7.87%** of the changes within the *[2] average daily price formations* are caused by *[I] all independent variables*. This model seems to be overall valid with *significance F* of 0.0014, however out of all five predictors included in the schema *[I]*, only the variables *[c] bubble implication* and *[d] length in words* feature worthy p – values of *[c] 0.000925* and *[d] 0.0212*, respectively.

Examined in linear regressions, *[1] title sentiment and [2] body sentiment* result to be significant predictors for every fluctuation – related price action: *[3] absolute change, [4] percentage change and [5] change tag.* Combined [II], the two are valid model only for the last response variable [5], however, in every of these scenarios, they explain less than 5% of the variations. One could easily ignore results like these, or more likely, conclude, within the current research, sentiment is not a meaningful predictor for considerable percentage for any of the price

– related matters. While such interpretation would not be wrong, it will be valid with remark for the statistical method used.

Both linear and multiple regression are established and widely prominent statistical models. The point is not to neglect or disregard them, but compare, experiment with the assessments of another predictor method – the *Random Forest Tree algorithm*. Random Forest (henceforth called RF), being introduced back in 2001, is relatively recent, but efficient and attracting vast attention statistical method for resolving classification and regression issues (Breiman, 2001; Genuer, Poggi & Tuleau-Malo, 2010). RF stems from and is built with supervised machine learning technology in mind. To grasp the essence behind random forest, or any similar method in that regard, requires not only deep knowledge in mathematics and statistics, but also expertise in the binary computer logic behind the user – friendly facade.

In order to illustrate the principles behind random forest, this thesis adopts simplified interpretation, based on common knowledge, as introduced by Will Koehrsen, young data science engineer, with excellent education, experience and profound interest in machine and deep learning algorithms (Koehrsen, n.d.). In its core, RF lays on the foundation of *decision tree (DT)* – decision making process, sophisticated version of *question-and-answer* decision taking flowchart, technique applied in our daily routine. Example here would be coming up with suggestion about tomorrow's maximum temperature within the city of Vienna. Perhaps unknowingly, we would ask ourselves a series of questions to narrow down the prediction. Based on common knowledge, generally the temperature could be anything between –50 and +50 degree Celsius. Knowing that the city is Vienna, one could easily expect the range to move within –25°C to +45°C. The season (summer), and the month (July) would help to pinpoint the likely maximum temperature more accurately at 33°C. These are three highly relevant questions, and by answering them relying on nothing but personal experiences, one could end up with rather valid guesstimate. In most such cases this end result would be satisfactory. Of course, there is always the possibility to gather further information, such as the monthly average temperature, to take under account past days humidity, rain volume or account for the global warming trend (Koehrsen, 2017).

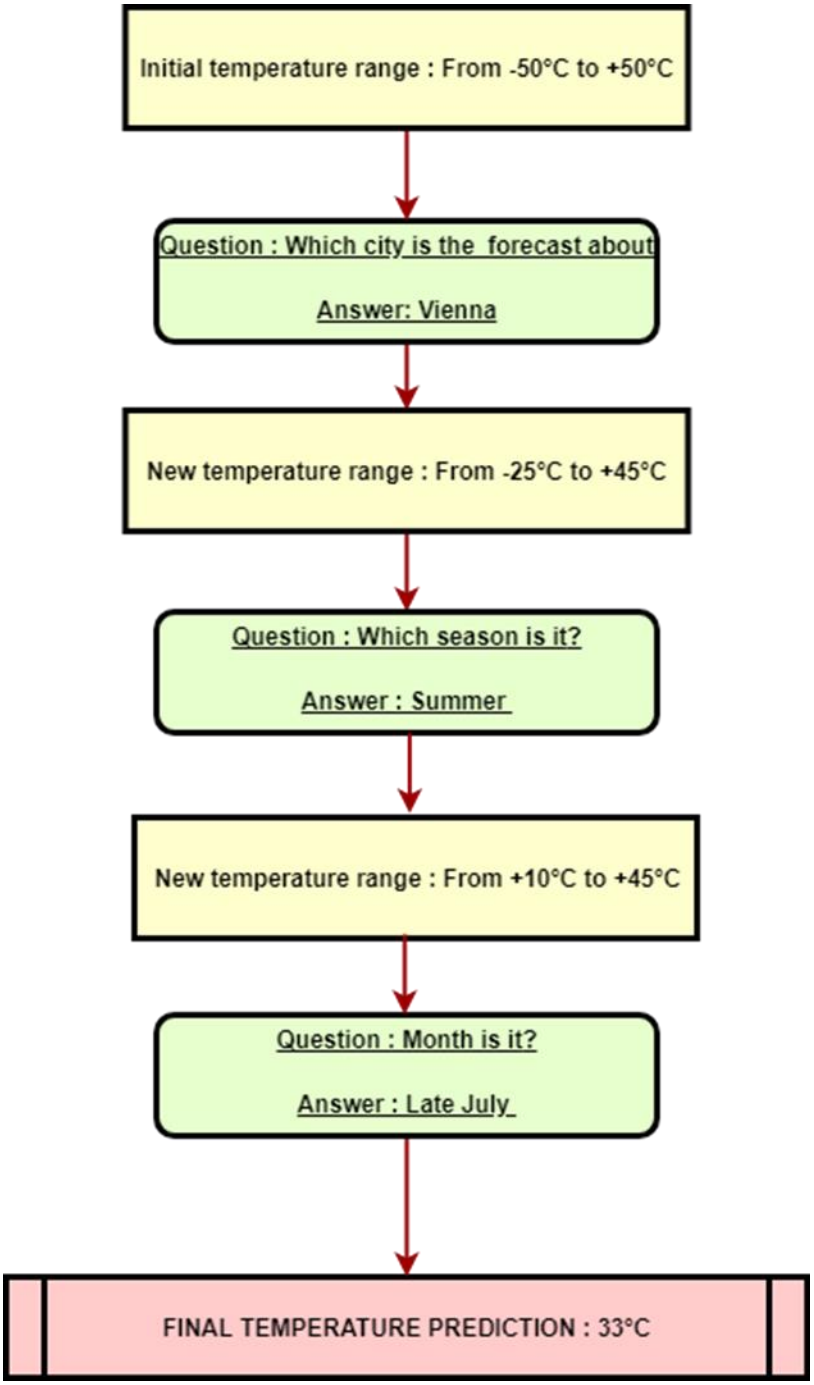

Initial temperature range : From -50°C to +50°C
Question : Which city is the forecast about
Answer: Vienna
New temperature range : From -25°C to +45°C
Question : Which season is it?
Answer : Summer
New temperature range : From +10°C to +45°C
Question : Month is it?
Answer : Late July
FINAL TEMPERATURE PREDICTION : 33°C

Figure 8. Decision taking flowchart

One could say, that decision tree implemented in machine learning, functions in a similar fashion. However, unlike us humans, the method adopts binary response format – *"Yes"* or *"No"*, and will consider all possible branches, and figuratively said, their twigs, that develop when answering the questions. For example, by adding the Scandinavian city of Malmo, the scenarios would look quite differently. If the season was winter instead of summer, that would change the forecast suggestions quite a bit too. The algorithm will not disregard any new branches growing as the symbolic tree thrives, and its output [of the statistical method's] will include prognosis for every chain of questions and answers (Koehrsen, 2017).

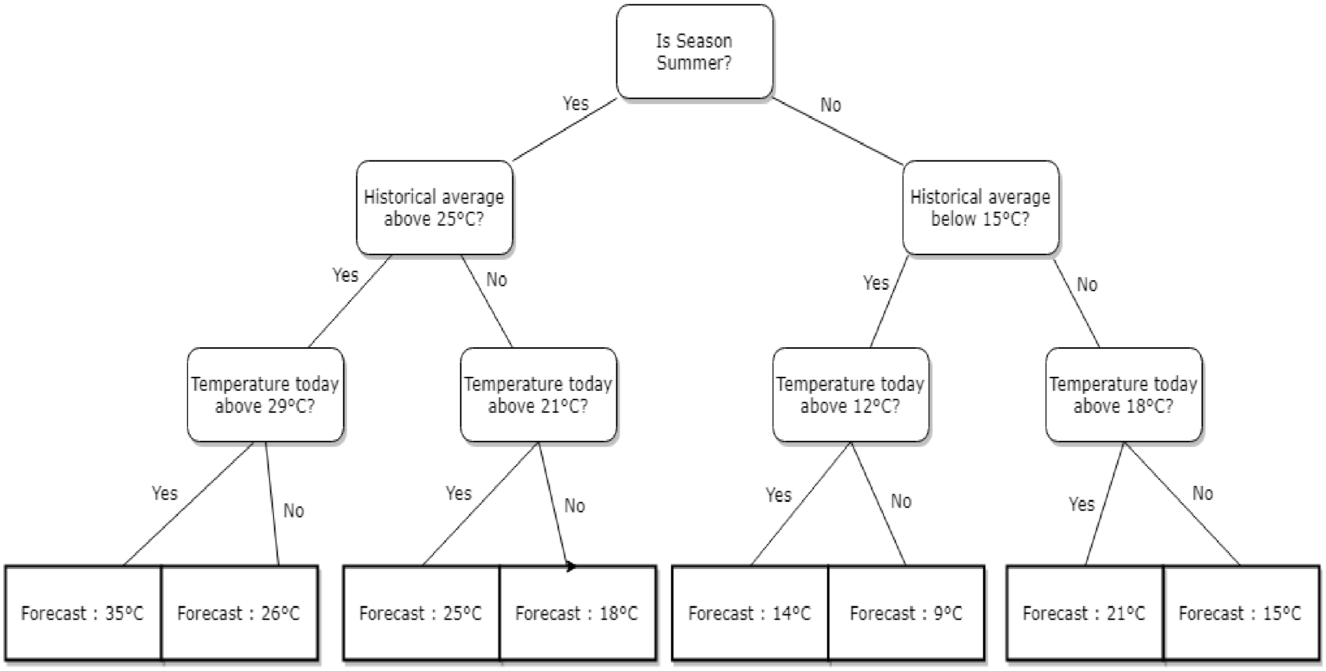

Is Season Summer?
Yes
No
Historical average above 25°C?
Historical average below 15°C?
Yes
No
Yes
No
Temperature today above 29°C?
Temperature today above 21°C?
Temperature today above 12°C?
Temperature today above 18°C?
Yes
No
Yes
No
Yes
No
Yes
No
Forecast : 35°C
Forecast : 26°C
Forecast : 25°C
Forecast : 18°C
Forecast : 14°C
Forecast : 9°C
Forecast : 21°C
Forecast : 15°C

Figure 9. Representation of a decision tree model (DT)

Important note here is that while we humans, with the help of common knowledge, agree for summer to feature higher temperatures than winter, or that Malmo is generally colder than Vienna, a decision tree does not have previously gained knowledge or experiences. DT, or any similar model, learns to build such relationships through the data it is provided with. This process is called training, where the algorithm *assimilates* the connections between individual variables. Perhaps it is worth mentioning, that not at any point, even after successful fitting, the model *does not have domain understanding.* All types of data appear the same to it – would it be weather forecast, sales volume or staffing, it is up to actual people to make sense of the statistical results (Koehrsen, 2017).

What random forest does, is to combine multiple decision trees in its equation. Back to the Viennese maximum temperature prognosis, instead of relying on the prediction of 33°C made by one single person, it would be a lot be better to source the forecasts of multiple individuals and simply take the average. The reasoning behind such an approach is quite simple – different people might take varying factors affecting their calculation. While the season and month gave us 33 degrees Celsius, someone who considers humidity and cloud cover density is likely to come with another prediction. Random forest considers the outputs of a number of decision trees, with each DT taking into consideration a random set of variables, factors when forming the questions and completing the unfolding scenarios (Koehrsen, 2017). Rightfully so, random forest could be compared to *computer – assisted, wisdom – of – crowds decision – making process.* Technicalities of how the algorithms function and the way information is being processed, have been included in the works by Tom Mitchell (1997) and Breiman (2001) for decision tree and random forest respectively, however these would end up being confusing for most.

Since the method has found its place in the financial literature too (Elagamy, Stanier & Sharp, 2018; Maragoudakis and Serpanos, 2010), the confident decision was taken to apply random forest for feature selection process within this current research. While *feature selection* might not sound as related to what has been explained in the previous paragraphs, it is intrinsically the same process. The maximum temperature example might not be appropriate for a depiction clear enough, however, *the choice whether to play tennis tomorrow after work,* is. Generally speaking, such a decision is likely to depend on a few factors – (1) hours of sleep the night before, (2) amount of food consumed during the day, (3) temperature of the air, (4) temperature on the field and even (5) family obligations during the evening or (6) mood. After proper training, where the algorithm connects the dots between the different variables, random forest model would develop multiple scenarios, decision trees, aimed at figuring out *which out of these 6 key factors, has the greatest importance when taking the decision.*

With the help of the programming language python, random tree forest statistical algorithm was trained and fitted using the data collected for this study, the same applied earlier within the correlation matrix and the regression calculations. Once fully tuned, the point behind each individual trial is to rank the features as per their importance demonstrated towards a predefined variable. On the following pages selected results have been shown. Entire list with multiple experiments is to be found in the appendix.

Figure №10 depicts the outcome *from three separate* top ten feature importance selection calculations for the following price – action related variables : (*1) absolute change, (2) percentage change and (3) change tag*.

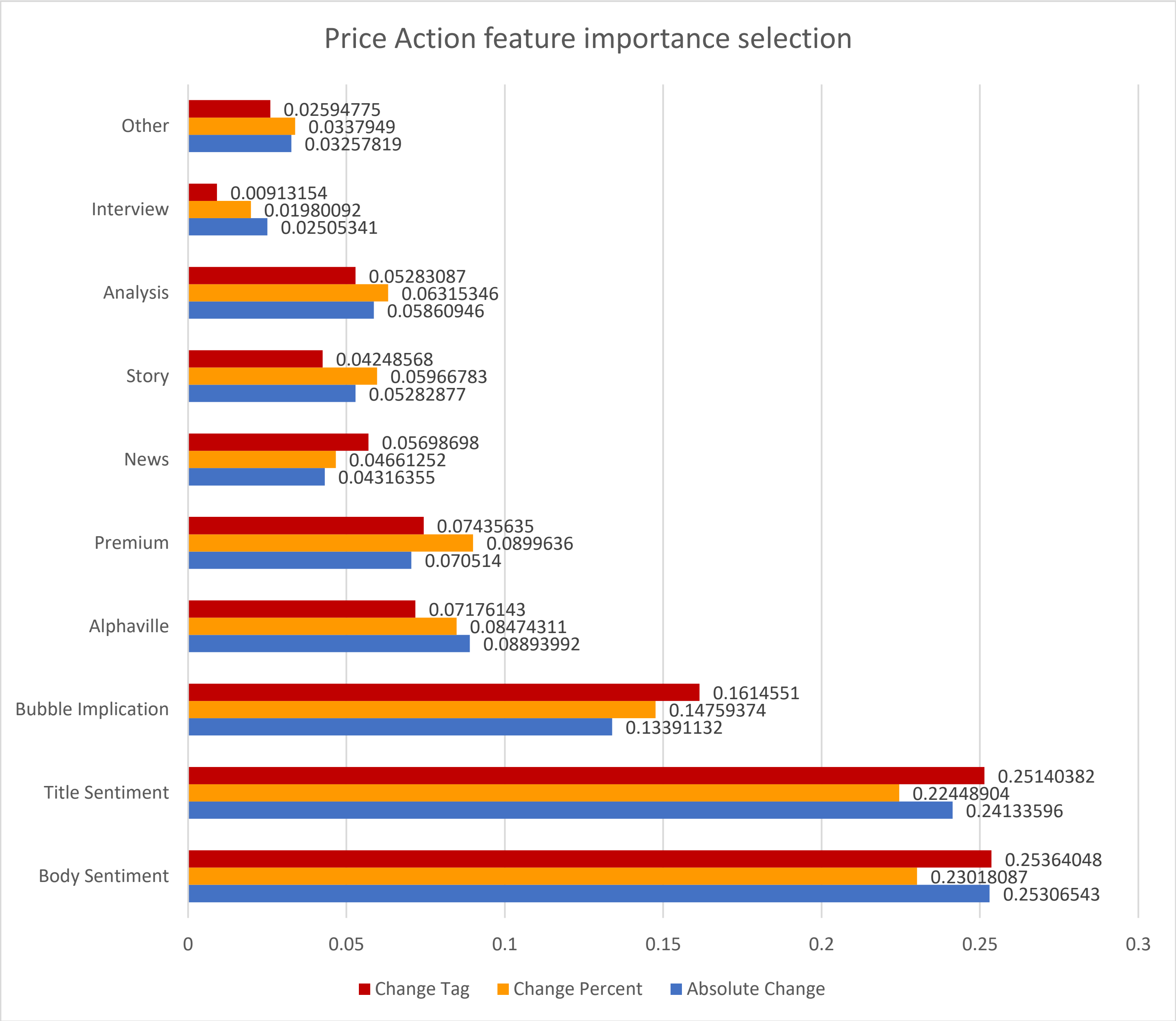


Figure 10. Feature importance selection for the three price fluctuation variables

Interesting observation is that for all three responders *body sentiment, title sentiment and bubble implication*, in the same order, return the highest score of significance. Both of the sentiment categories fluctuate within very small range – less than 3%, with them being slightly less pronounced for the *absolute change of price*, compared to the other two price – related respondents. The bubble implication factor follows closely on the 3rd position for all of them, while *interview as article genre is generally of a least importance*. Surprisingly, *the publication being premium content or part of the Alphaville blog has greater weight than the article's category – news, story or analysis.*

The second histogram, Figure №11 consists of two separate tests. One examines the most important determinants for the *length of articles*, while the second one – these for *the length of negative publications*. Perhaps surprisingly, the RF algorithm has sorted *the number of characters and the number of words* as the most important features for both of the dependent variables. The title sentiment variable plays only a secondary role, with the *body sentiment being of no importance at all for the length of negative articles.*

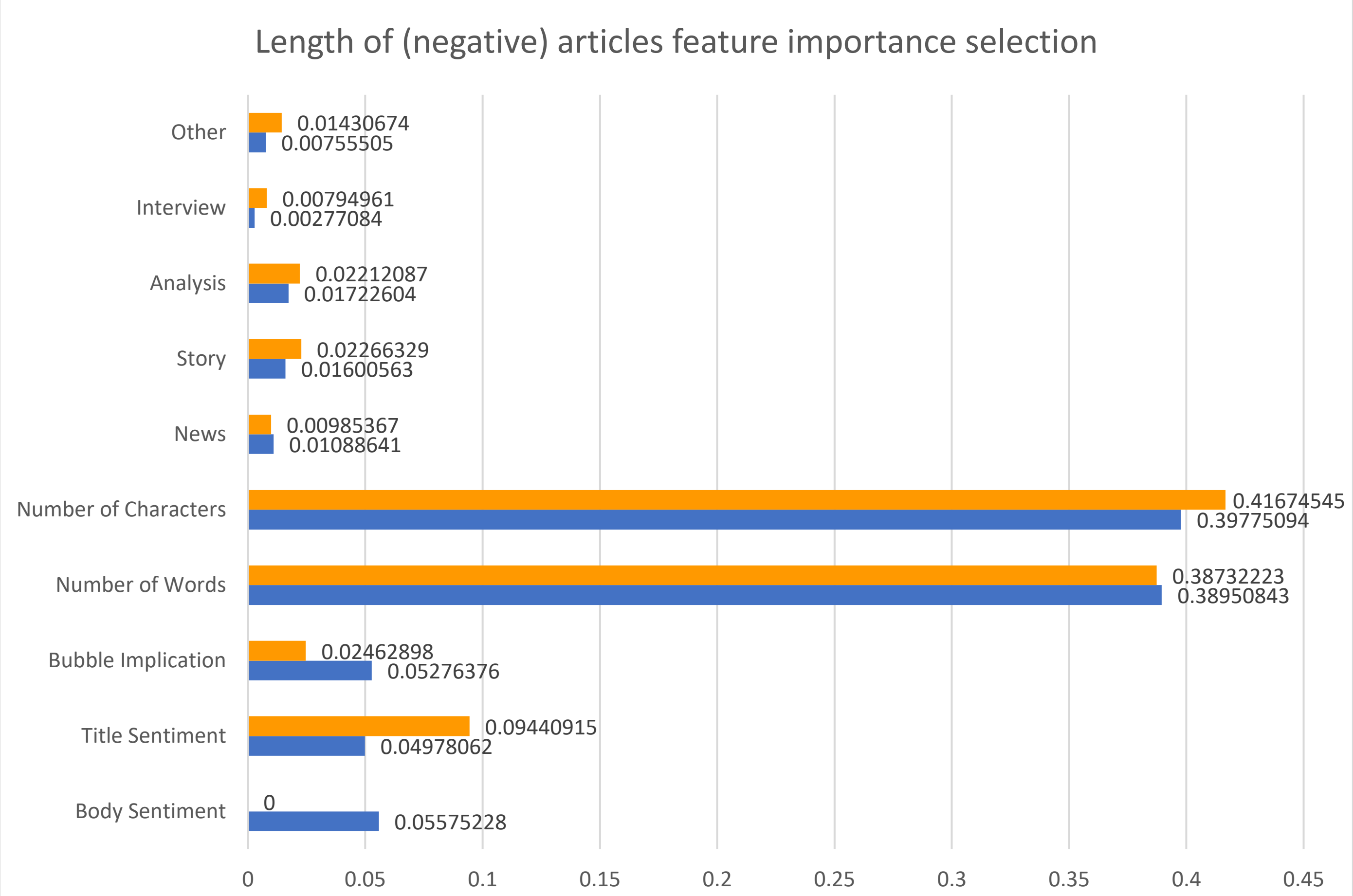


Figure 11. Feature importance selection for the length of (negative) articles

In this last experiment, the top determinants for articles, suggesting that BTC might be in a bubble were selected. In that scenario, the machine learning algorithm managed to come up with 8 significant features. Body sentiment and title sentiment variables again turned out to be of a biggest importance, with close to 55% when combined.

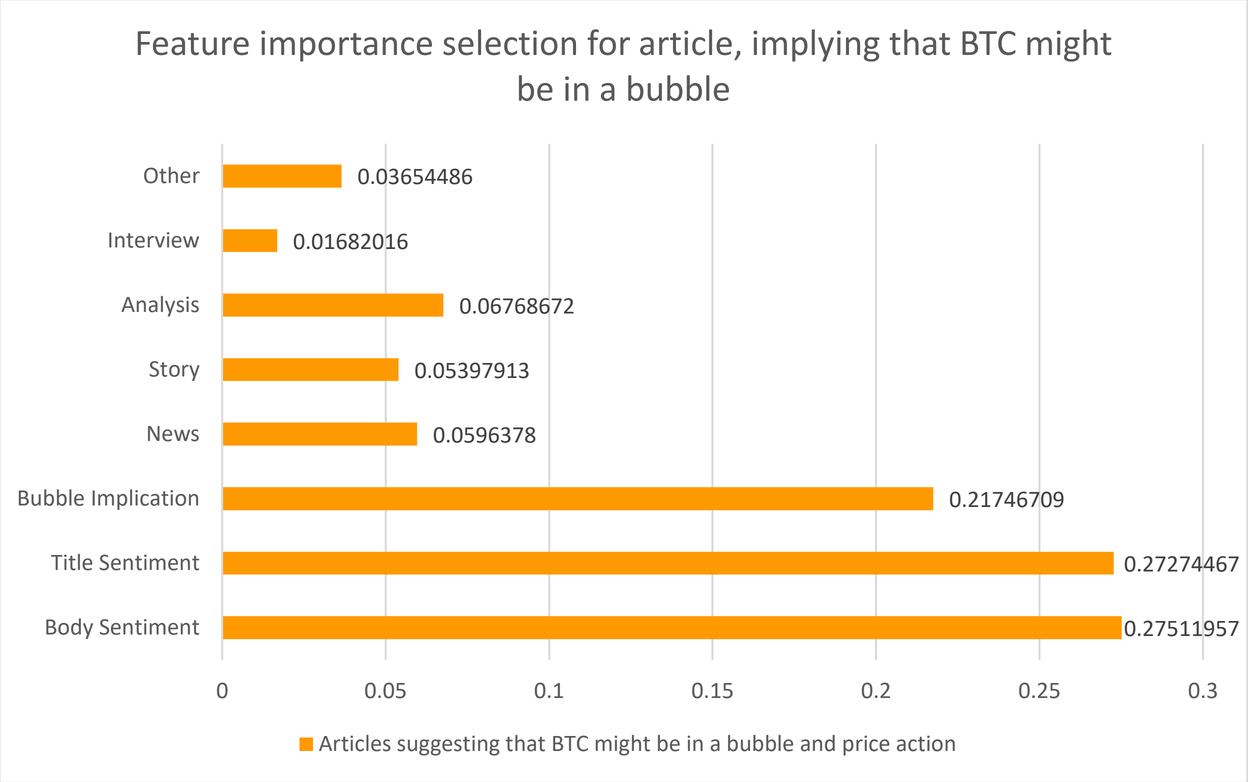


Figure 12. Feature importance selection for articles, suggesting that Bitcoin might be in a bubble

DISCUSSON, LIMITATIONS AND FURTHER RESEARCH

Regarding the first research question, and namely, *the general frequency characteristics of Bitcoin – related news for the period of study*, there are some noteworthy observations to be made. The month of Bitcoin's bottom, January 2015 (3 articles in total), does not feature the same publication density as the month of its top – December 2017, with 49 articles. From this fact alone, one can conclude that media, or at least in the case of the selected research medium within this study, The Financial Times, **has granted the top of the bubble with significantly more attention than the discontinuation of decline from the preceding implosion.** The same could be observed on a yearly basis too, with 49 articles within 2015 versus 155 for year 2017, however, year 2016, which one can pinpoint as the period of build-up (accumulation) before the explosive move remained least active, and accounts for less than 17% from all filtered publications within the 3 years of research.

There are no set rules to what would be considered as *[in]frequent Bitcoin publications*, however a quick look through the dataset shows that there are often a*nywhere from 5 to 20 days between individual articles, with both 2015 and 2016 featuring a period of 37 days with no relevant publications.* With the gaps being irregular to that extend, and the inability to come up reasonably spaced intervals (minimum would be 37 days) the common statistical methods, partially forecast ones, *cannot figure out the extend with which negative / positive sentiment loaded Bitcoin – related articles could predict Bitcoin's price action.* This research question was bound on the assumption of validity of 1st and 2nd level of agenda – setting, and the way the data was designed would generally allow for such observation, which outcomes would have been discussed with the generous amount of studies and theories review introduced earlier.

The hurdles of bringing the fourth research question to an end lie within the limited data. After the selection process, out of 1056 publications 244 were found to be relevant, randomly scattered across 155 days, which equals to *less than 15% of the period of study featuring some meaningful coverage.* Ideally, there should be data for every day during the period of study, but model could have been constructed with equal 4 day intervals, which equals to roughly 55% of days featuring relevant publications, instead of the current 15%. ***This sparsity of the final data set, evidenced by the irregular publication intervals, is also the study's biggest limitation.*** In hindsight, the selection and annotation processes are the ones that resulted in such turn of events. As covered earlier in the chapter *"Method"*, computer – assisted article filtration and their sentiment categorization, have the advantage of being considerably faster. For the current research, both of these procedures were completed manually, a labour – intensive, time – consuming task, which also stretched out beyond the deadline initially set. In retrospect, despite the delay, this was the correct approach. While it is true that Fouad, Gharib and Mashat (2018) managed to achieve over 90% congruence to previously human annotated polarity of a Twitter corpus, coming up with such tool is equally lengthy process that requires a whole bucket of additional skills and knowledge. ***What could have been done better and should be considered in future studies sharing similar research interests, is to carefully plan for a bigger group of annotators, knowledgeable on the topic of study.*** Kalogeropoulos et. al., (2014) for example *employed* full-time three annotators to classify their articles. That would allow for *(1) intercoder reliability and (2) potentially a bigger number of outlets to be explored*. If the same type of data was present not only for The Financial Times (www.ft.com) but for The Wall Street Journal (www.wsj.com), The New York Times (www.nyt.com) and CNN as well, there is a higher chance of filling the gaps or bin valid intervals with proportionate data. That would of course significantly extend the time required for completion, with Bhattacharya et.al., (2009) being an extreme example with them needing 2 years.

The sentiment characteristics of Bitcoin – related publications (RQ2), show that both their headlines and bodies stay mostly neutral and least positive. The negative titles are 27%, but the amount of actually negative articles lies at 35% (Table 4, pg. 48). Looking at the various article genres, what stands out are the *news,* which headlines are either favorable or unfavorable towards Bitcoin, with less than 12% taking no side. At the same time, their content is largely neutral – 65% from all articles from that same category [news]. *Further similar discrepancy is within the analyses, the second biggest group of publications, where 59 out 96 headlines are neutral, however only 36 of them stay so in their bodies too. In their contents, more than 60% (58 out of 96) of the analyses share negative sentiment towards Bitcoin.* Having in mind that there are 244 relevant publications in total, and 86 of them are negative, negatively loaded analyses occupy some 67% of all unfavorable and close to 24% of all articles.

All three polarity classifications are represented evenly during the year of the bottom, 2015. Both in 2016, the year of momentum build – up, and 2017, positive articles account for only 15%. And while in 2016 there were 17 negative and 17 neutral publications, 2017 was dominated by the neutral ones – 79 out of 155. The most active month overall, December 2017, has similar overall distribution, with close to 50% of the titles and 57% of the contents conveying no clear sentiment direction. It is true, that there are some minor disparities, however, the sentiment of the title and the body of text sentiment show positive, and the second biggest among all variables correlation, with coefficient of r = 0.54.

Table 8. Distribution of Negative / Neutral / Positive articles as a total number / % from all relevant publications within the year

Year	Negative	Neutral	Positive
2015	15 / 30.61 %	16 / 32.65%	18 / 36.73%
2016	17 / 42.5%	17 / 42.5%	6 / 15%
2017	54 / 34.84%	79 / 50.97%	22 / 14.19%

Looking closer at the 3rd research question, and the bubble – implication variable within the reporting, year 2017 again seems to be the most active, when 41 out of all 49 articles, suggesting that Bitcoin might be overvalued have been published and with 15 of them released in December 2017. 27 out of these 49 publications, or some 55% are analyses.

Table 9. Distribution of articles suggesting that Bitcoin could be in a bubble as a total number / % from all relevant publications within the year

Year	No	Yes
2015	43 / 87.76 %	6 / 12.24%
2016	38 / 95%	2 / 5%
2017	114 / 73.55%	41 / 26.45%

Last but not least, the interest point of this study, is to explore the degree to which sentiment plays role in Bitcoin's price formation (RQ5). It would have been of an immense benefit if the fourth research question was answered too and combine their interpretation. The fifth research question gave way to the Random Forest model – a recent supervised machine – learning method, that simulates wisdom – of – crowds knowledge and has been shown to be *vastly flexible and accurate with smaller data sets too* (Biau and Scornet, 2015, p.2; Floares et al., 2017).

The statistical model clearly signified, that from all the coded variables, body sentiment is the most important feature defining Bitcoin's valuation, in all of its dimensions. Having in mind the Change Tag variable has been coded in such a way, to represent price fluctuations with core percentage (0 ≤ -5%; 1 ≤ 0%; 2 ≤ 5%; 3 ≥ 5%), that would lead to the interpretation that body sentiment, followed by title sentiment and bubble implication characteristics, are the three most important determinants of Bitcoin's value fluctuation. On the other side, if further assumptions are to be made, there is the theoretical possibility to make a relation between the RF results and the ones from the correlation matrix (please look at the chapter *"Random Forest Outputs"*), coming to the conclusion that *the body sentiment being favorable, leads to BTC's price fluctuations being in the positive too.* While such ratiocinations regarding the direction of affect could be made for the other two dependent price variables too, all of these would be rather highly speculative, since the correlation results were not even weak, with their coefficient being positive but in the levels of only up to r = +0.20. On top of that, the linear and multiple regressions ran, let only a marginal percentage of the valuations variations be explained with the help of the sentiment variables.

Random forest however lets us conclude, that indeed the sentiment of the article and its headline are important features within the price formation, while the article genres are likely just noise. Both of the sentiment determinants are only of a secondary importance when it comes to the length of (dominantly negative) publications. This is valuable conclusion for future similar studies, by pinpointing the most influential features regarding the price – action. That would allow researchers to explore only the meaningful variables, and potentially focus on smaller determinants within them. Last but not least, it should be noted that in the current case, Random Forest does not imply the existence or build – up of lag, since the calculations performed are valid for the intraday movement, i.e. daily price fluctuations. This leads to the more concrete inference, that ***the sentiment of news articles together with their respective titles are the variables featuring the greatest importance within Bitcoin daily price action, inevitable part of longer bubble-formations.***

CONCLUSION

The payment system nowadays, would it be a bank or card transaction, relies on financial institutions between the parties involved in order for it to function. Bitcoin on the other side, is a decentralized system, a form of digital cash, that allows for payments to be processed without the need of a trusted vendor. It relies on cryptographic proof, to ensure transactions which are irreversible, faster, cheaper and with significantly higher level of anonymity.

While in the beginning the majority was skeptical towards the concept, maybe due to the fact that it is hard to grasp, mere 8 years after the inception of Bitcoin, the cryptocurrency started to appear for the broader public, investors, institutions and regulators. Created by the mystical Satoshi Nakamoto, unknown individual or a group of people, there are clear signals for Bitcoin to be enjoying growing adoption.

The novel technology is here to stay, and its potential to become part of everyone's life, requires Bitcoin to be explored from various disciplines. It is certainly phenomenon of a high interest for economics, since it grew from $0 to $20 000 within less than 10 years, accompanied by multiple booms and bursts.

There a number of theories in the financial literature able to explain such formations, however Shiller (2000, p.95) points at media to be the*"[...] fundamental propagators of speculative price movements [...]"*. Indeed, communication science has been for long studying the way media affects its audience and their perceptions. 1st and 2nd level of agenda setting, as well as their extensions enjoy high popularity, and their validity has been confirmed in various contexts. The concept implies for media to dictate what recipients to think about, and how to think about it, *"embedding in their minds"* the selected facts about the objects of discussion, as well as the polarity they have been presented with – negative, neutral or positive.

Literature is abound with studies investigating the relationship between media reporting and the financial markets. There is enough evidence, for news to impact valuations of various assets and cause severe price deformations.

Since its inception, Bitcoin went through 3 major bubble cycles. The last one, which lasted from 14.01.2015 till 17.12.2017, raised the cryptocurrency's valuation from just under $165 to almost $20 000 before the inevitable burst. Namely this recent formation is of main interest for the current master thesis, which explored the role media might have played.

The ressources allocated, allowed only for the analysis of website of the internationally recognized financial outlet *The Financial Times* – (www.ft.com). During the period of study, which is a total of 1068 days, 1058 publications matched with the keyword *"bitcoin"* as a search criteria. After careful manual selection process, only 244 articles were found to be relevant to Bitcoin or have a meaningful matter as main topic. Subsequently, the title sentiment, body of text sentiment as well as the bubble implication characteristics among many others were manually annotated.

These 244 relevant articles were published in 155 unique days. Having in mind the sparsity of the data, no exact statistical calculations could be performed to figure out the response of price on the articles over a longer period of time. Among the more interesting observations are, that majority of the publications stayed neutral and least positive. Analyses, the second largest genre type, occupying close to 40% of all articles, conveyed mostly negative sentiment towards Bitcoin. Almost 50% of the articles, hinting that the cryptocurrency might be overvalued fall under the category of analyses too.

Applying Random Forest, a supervised machine – learning method, and relying on it for feature importance selection, body sentiment and title sentiment, resulted to be the most influential variables for intraday price fluctuations. While the correlation coefficients are rather low, their speculative interpretation hints at the sentiment being positive to lead to higher daily price action, which theoretically could result into the formation of a bubble.

The results are curious and call for more research towards the relationship of media and Bitcoin, ideally with more data at hand.

REFERENCES

Agrawal, H. (2018). *How To Keep Your Recovery Seed Safe And Secure*. Retrieved from https://coinsutra.com/keep-recovery-seed-safe-secure/

Albuquerque, B., & Callado, M. (2015). Understanding Bitcoins: Facts and Questions. *Revista Brasileira de Economia, 69*(1), 2-16.

Alex, D. (2018). *Phishing for cryptocurrencies: How bitcoins are stolen*. Retrieved from https://www.kaspersky.com/blog/crypto-phishing/20765/

Alkorta, J., Gojenola, K., Iruskieta, M., & Taboada, M. (2018). Using lexical level information in discourse structures for Basque sentiment analysis. (2016), 39-47.

Alonso, K. (2018). Zero to Monero: First Edition a technical guide to a private digital currency; for beginners, amateurs, and experts *v1.0.0.*

Ankalkoti, P., & Santhosh, S. (2017). A Relative Study on Bitcoin Mining. *Imperial Journal of Interdisciplinary Research (IJIR), 3*(5), 1757-1761.

Arthur, C. (2011). *Bitcoin value crashes below cost of production as broader use stutters | Technology | theguardian.com*. Retrieved from https://www.theguardian.com/technology/2011/oct/18/bitcoin-value-crash-cryptocurrency

Asghar, M., Kundi, F., Khan, A., & Ahmad, S. (2014). Lexicon-based sentiment analysis on the social web. *Journal of Basic and Applied Scientific Research, 4*(6), 238-248.

Ayyoub, M., Essa, S., & Alsmadi, I. (2015). Lexicon-based sentiment analysis of Arabic tweets. *International Journal of Social Network Mining, 2*(2), 101.

Barber, B., & Odean, T. (2008). All that Glitters: The effect of Attention and news on the Buying Behavior of Individual and Institutional Investors. *Christopher, D., Jonathan, a, & Nicholas, S. (2014). CFS Working Paper No. 465. Barber, B. M., & Odean, T. (2008). All that Glitters: The effect of Attention and news on the Buying Behavior of Individual and Institutional Investors. The Handbook of News A*, 173-210.

Bariviera, A., Basgall, M., Hasperué, W., & Naiouf, M. (2017). *Some stylized facts of the Bitcoin market.*

Barlevy, G. (2007). Economic Theory and Asset Bubbles. *Federal Reserve Bank of Chicago Economic Perspectives*(Q3), 44-59.

Barzilai-Nahon, K. (2011). Gatekeeping: A critical review. *Annual Review of Information Science and Technology, 43*(1), 1-79.

Baur, D. (2013). *Gold - Fundamental Drivers and Asset Allocation.*

Berkowitz, D. (1992). *Public Opinion, the Press, and Public Policy - Google Books.*

Bernard, Z. (2018). *Everything you need to know about Bitcoin, its origins, and its creator - Business Insider Deutschland*. Retrieved from https://www.businessinsider.de/bitcoin-history-cryptocurrency-satoshi-nakamoto-2017-12?r=US&IR=T

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test, 25*(2), 197-227.

Biddle, S. (2018). *The NSA Worked to "Track Down" Bitcoin Users*. Retrieved from https://theintercept.com/2018/03/20/the-nsa-worked-to-track-down-bitcoin-users-snowden-documents-reveal/

BitcoinFees. (2018). *Bitcoin Transaction Fees*. Retrieved from https://bitcoinfees.info/

Bitnodes. (2019). *Global Bitcoin Nodes Distribution - Bitnodes*. Retrieved from https://bitnodes.earn.com/

BitPay. (2019). *BitPay.Com*. Retrieved from https://bitpay.com/

Blood, W. (1982). Agenda Setting: A Review of the Theory. *Media Information Australia, 26*(1), 3-12.

Bordoloi, M., & Biswas, S. (2018). Sentiment Analysis of Product using Machine Learning Technique: A Comparison among NB, SVM and MaxEnt. *International Journal of Pure and Applied Mathematics, 118*(July), 71-83.

Bowe, B., Fahmy, S., & Wanta, W. (2013). Missing religion. *International Communication Gazette, 75*(7), 636-652.

Boydstun, A., Gross, J., Resnik, P., & Smith, N. (2013). Identifying Media Frames and Frame Dynamics Within and Across Policy Issues. *New Directions in Analyzing Text as Data Workshop*, 1-13.

Brave New Coin. (0). *BNC Bitcoin Liquid Index*. Retrieved from https://bravenewcoin.com/enterprise-solutions/indices-program/blx

Brave New Coin. (2019). *BNC Weekly Bitcoin Blockchain Statistics 7th January 2019 » Brave New Coin*. Retrieved from https://bravenewcoin.com/insights/bnc-weekly-bitcoin-blockchain-statistics-7th-january-2019

Breiman, L. (2001). Random Forests. 1-33.

Buchholz, M., Delaney, J., & Warren, J. (2012). Bits and Bets.

Buck, J. (2017). *Chase Bank Buys Bitcoin Even as Jamie Dimon Rejects It | Cointelegraph*. Retrieved from https://cointelegraph.com/news/chase-bank-buys-bitcoin-even-as-jamie-dimon-rejects-it

Bulat, R. (2018). *Accept Crypto with Coinbase Commerce: Intro & Integration Guide*. Retrieved from https://medium.com/@rossbulat/accept-crypto-with-coinbase-commerce-intro-integration-guide-8fc4dc7df10f

Burgt, J. (2018). Making Sense of Bitcoin Price Levels. (April), 1-4.

Bushee, B., & Miller, G. (2012). Investor relations, firm visibility, and investor following. *Accounting Review, 87*(3), 867-897.

Campbell, R. (2019). *Today Marks The 10th Anniversary Of The First Bitcoin Transaction*. Retrieved from https://www.forbes.com/sites/rebeccacampbell1/2019/01/12/today-marks-the-10th-anniversary-of-the-first-bitcoin-transaction/#6ca7c41e415a

Canellis, D. (2018). *Criminals used Bitcoin to launder $2.5B in dirty money, data shows*. Retrieved from https://thenextweb.com/hardfork/2018/10/10/bitcoin-money-laundering/

Carrel, P. (2015). *U.S. spy agency tapped German chancellery for decades: WikiLeaks*. Retrieved from https://www.reuters.com/article/us-germany-usa-spying/u-s-spy-agency-tapped-german-chancellery-for-decades-wikileaks-idUSKCN0PI2AD20150708

Carroll, C., & McCombs, M. (2003). Agenda-setting Effects of Business News on the Public's Images and Opinions about Major Corporations. *Corporate Reputation Review, 6*(1), 36-46.

Cecchetti, S. (1992). *Prices during the Great Depression: Was the Deflation of 1930-1932 Really Unanticipated?*

Cecchetti, S. (1997). Understanding the Great Depression: Lessons for Current Policy.

Chainanalysis. (2018). *Bitcoin's $30 billion sell-off*. Retrieved from https://blog.chainalysis.com/reports/money-supply

Chinco, A., Bordalo, P., Da, Z., Gabaix, X., Hartzmark, S., Jin, L., . . . Pollet, J. (2018). The Madness Of Crowds And The Likelihood Of Bubbles *. 1-47.

Chong, N. (2018). *China Banned Everything Bitcoin, Video Games Seem To Be Next | NewsBTC*. Retrieved from https://www.newsbtc.com/2018/08/31/china-banned-everything-bitcoin-video-games-seem-to-be-next/

Christopher, D., Jonathan, a., & Nicholas, S. (2014). CFS Working Paper No. 465. *Christopher, D., Jonathan, a, & Nicholas, S. (2014). CFS Working Paper No. 465. Barber, B. M., & Odean, T. (2008). All that Glitters: The effect of Attention and news on the Buying Behavior of Individual and Institutional Investors. The Handbook of News A*.

Clayton, J. (2017). *SEC.gov | Statement on Cryptocurrencies and Initial Coin Offerings*. Retrieved from https://www.sec.gov/news/public-statement/statement-clayton-2017-12-11

COHEN, B. (1963). *THE PRESS AND FOREIGN POLICY.* Princeton University Press.

CoinmarketCap. (2019). *Cryptocurrency Market Capitalizations | CoinMarketCap*. Retrieved from https://coinmarketcap.com/

Coleman, R., & Wu, D. (2010). PROPOSING EMOTION AS A DIMENSION OF AFFECTIVE AGENDA SETTING:. *Journalism and Mass Communication Quarterly*(2), 315-327.

Comben, C. (2018). *Venezuela Sees Biggest Increase in Bitcoin Volume to Date - Bitcoinist.com*. Retrieved from https://bitcoinist.com/venezuela-buying-bitcoin-record-jump/

Conerly, B. (2013). *What Is A Bubble?* Retrieved from https://www.forbes.com/sites/billconerly/2013/07/24/what-is-a-bubble/#4d12b75ee648

De Graaf, R., & Van Der Vossen, R. (2013). Bits versus brains in content analysis. Comparing the advantages and disadvantages of manual and automated methods for content analysis. *Communications, 38*(4), 433-443.

De Heij, H. (2012). *Designing Banknote Identity DNB Occasional Studies.*

Dearing, J., & Rogers, E. (1996). *Agenda - Setting.* Thousand Oaks.

DeCuir-Gunby, J., Marshall, P., & McCulloch, A. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods, 23*(2), 136-155.

Dickinson, B. (2018). *How to pick the right online crypto wallet*. Retrieved from https://thenextweb.com/cryptocurrency/2018/02/26/why-you-should-not-store-your-cryptos-on-an-exchange-but-use-a-wallet-instead/

Dou, Z.-Y. (2018). Capturing User and Product Information for Document Level Sentiment Analysis with Deep Memory Network. 521-526.

Duarte, N., Llansó, E., & Loup, A. (2018). Mixed Messages? The Limits of Automated Social Media Content Analysis 1 Presented at the 2018 Conference on Fairness, Accountability, and Transparency. (4).

Durden, T. (2013). *Citi: Bitcoin Could Look Attractive To Reserve Managers As A Complement To Gold | Zero Hedge*. Retrieved from https://www.zerohedge.com/news/2013-12-05/citi-bitcoin-could-look-attractive-reserve-managers-complement-gold

Durden, T. (2017). *Is The People's Bank Of China Manipulating The Bitcoin Price? | Zero Hedge*. Retrieved from https://www.zerohedge.com/news/2017-06-15/peoples-bank-china-manipulating-bitcoin-price

Eckner, A. (2014). A framework for the analysis of unevenly spaced time series data. (1991), 1-45.

Edison, M., & Aloysius, A. (2017). Polarity detection of lexicon based sentiment analysis with negation handling. *Journal of Advanced Research in Dynamical and Control Systems, 9*(Special Issue 13), 44-54.

ENTMAN, R. (1993). Framing - Toward Clarification of a Fractured Paradigm. *Journal of Communication, 43*(4), 51-58.

Erickson, B., Novilla, L., Barnes, M., Meacham, A., Hanson, C., & McIntyre, E. (2008). Analysis of Media Agenda Setting During and After Hurricane Katrina: Implications for Emergency Preparedness, Disaster Response, and Disaster Policy. *American Journal of Public Health, 98*(4), 604-610.

Er-Rajy, L., El Kiram, A., El Ghazouani, M., & Achbarou, O. (2017). Blockchain: Bitcoin Wallet Cryptography Security, Challenges and Countermeasures. *Journal of Internet Banking and Commerce, 22*(3), 1-29.

European Commision. (2018). *European Commission - PRESS RELEASES - Press release - Remarks by Vice-President Dombrovskis at the informal ECOFIN press conference in Vienna*. Retrieved from http://europa.eu/rapid/press-release_SPEECH-18-5716_en.htm

Evanoff, D., Kaufman, G., & Malliaris, A. (2012). Asset price bubbles: What are the causes, consequences, and public policy options? Chicag o Fed Letter ESSAYS ON ISSUES THE FEDERAL RESERVE BANK NOVEMBER 2012 OF CHICAGO NUMBER 304. (November).

Fernández-Cabana, M., Jiménez-Féliz, J., Alves-Pérez, M., Mateos, R., Gómez-Reino Rodríguez, I., & García-Caballero, A. (2015). Linguistic analysis of suicide notes in Spain. *The European Journal of Psychiatry, 29*(2), 145-155.

Feroz, Z. (2019). *Spend your crypto instantly with Coinbase Card*. Retrieved from https://blog.coinbase.com/spend-your-crypto-instantly-with-coinbase-card-4c840e59a8d8

Field, A., Kliger, D., Wintner, S., Pan, J., Jurafsky, D., & Tsvetkov, Y. (2018). Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies. 3570-3580.

Filho, R., Almeida, J., & Pappa, G. (2015). Twitter Population Sample Bias and its impact on predictive outcomes. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1254-1261.

Floares, A., Applications, A., View, I., & Floares, A. (2017). The Smallest Sample Size for the Desired Diagnosis Accuracy The Smallest Sample Size for the Desired Diagnosis Accuracy. *2*(September).

Foley, S., Karlsen, J., Putniņš, T., Goldstein, I., Jiang, W., Karolyi, A., . . . Easley, D. (2018). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? .

Foote, C., Gerardi, K., & Willen, P. (2015). Why Did so Many People Make so Many Ex Post Bad Decisions? The Causes of the Foreclosure Crisis. *Ssrn*.

Fouad, M., Gharib, T., & Mashat, A. (2018). The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018). *723*(January).

French, D. (2006). THE DUTCH MONETARY ENVIRONMENT DURING TULIPMANIA. *9*(1), 3-14.

Frey, C., & Cook, J. (2004). *How Amazon.com survived, thrived and turned a profit*. Retrieved from Seattle Post Intelligence: http://seattlepi.nwsource.com/business/158315_amazon28.html

Friedman, H., & Hirakubo, N. (2002). Dot-Bombs: lessons from the dot-com debacle.

Funkhouser, G. (1973). THE ISSUES OF THE SIXTIES: AN EXPLORATORY STUDY IN THE DYNAMICS OF PUBLIC OPINION. *The Public Opinion Quarterly, 37*.

Gamberini, R., Lolli, F., Rimini, B., & Sgarbossa, F. (2010). Forecasting of sporadic demand patterns with seasonality and trend components: An empirical comparison between holt-winters and (s)ARIMA methods. *Mathematical Problems in Engineering, 2010*(July 2010).

Garbarino, J. (2011). TULIPMANIA: THE ECONOMIC BUBBLE OF THE SEVENTEENTH CENTURY. *Natural Selections*(76), 1-2.

Garber, P. (2000). *Famous First Bubbles; The Fundamentals of Early Manias.* The MIT Press.

Garcia Swartz, D., Hahn, R., & Layne-Farrar, A. (2004). J O I N T The Economics of a Cashless Society : An Analysis of the Costs and Benefits of Payment Instruments The Economics of a Cashless Society : An Analysis of the Costs and Benefits of Payment Instruments. (January 2006).

Garcia-Alfaro, J., Herrera-Joancomartí, J., Lupu, E., Posegga, J., Aldini, A., Martinelli, F., & Suri, N. (2015). Data privacy management, autonomous spontaneous security, and security assurance. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8872*(September).

Gauer, M. (2017). Bitcoin miners true energy consumption. (December), 9.

Gavora, P. (2015). The State-of-the-Art of Content Analysis. *Research Centre, Faculty of Humanities, Tomas Bata University*, 6-18.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266-6282.

Gingrich, P. (2004). Chapter: Regression. *Introductory Statistics for the Social Sciences*, 834-884.

Girdzijauskas, S., Štreimikiene, D., Čepinskis, J., Moskaliova, V., Jurkonyte, E., & Mackevičius, R. (2009). Formation of economic bubbles: Causes and possible preventions. *Technological and Economic Development of Economy, 15*(2), 267-280.

Gisler, M. (2012). Tulip Mania ? : The Dutch Tulip Bulb Episode (1636-1637) Revisited Tulip Mania ? *Schweizerische Gesellschaft für Wirtschafts- und Sozialgeschichte, 27*, 79 - 96.

Giusti, G., Jiang, J., & Xu, Y. (2014). Interest on Cash, Fundamental Value Process, and Bubble Formation on Experimental Asset Markets. *MPRA*(54970).

Goetzmann, W. (2015). BUBBLE INVESTING : LEARNING FROM HISTORY. *NATIONAL BUREAU OF ECONOMIC RESEARCH, 49*(23–6).

Golan, G., Kiousis, S., & McDaniel, M. (2007). Second-level agenda setting and political advertising: Investigating the transfer of issue and attribute saliency during the 2004 us presidential election. *Journalism Studies, 8*(3), 432-443.

Goldfeder, S., Kalodner, H., Reisman, D., & Narayanan, A. (n.d.). *When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies.*

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2014). Comparing and Combining Sentiment Analysis Methods.

Griffin, J., & Shams, A. (2018). Is Bitcoin Really Un-Tethered? *SSRN Electronic Journal*, 10-15.

Guegan, D. (2018). *The Digital World: I-Bitcoin: from history to real live.*

Han, H., Zhang, Y., Zhang, J., Yang, J., & Zou, X. (2018). Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. *PLoS ONE, 13*(8), 1-11.

Hatzivassiloglou, V., & McKeown, K. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 35th annual meeting on Association for Computational Linguistics* -, 174-181.

Helfer, L., & Aelst, P. (2016). What Makes Party Messages Fit for Reporting? An Experimental Study of Journalistic News Selection. *Political Communication, 33*(1), 59-77.

Helms, K. (n.d.). *Thailand Unveils Details of Crypto Regulations, Legalizing 7 Cryptocurrencies - Bitcoin News*. Retrieved from https://news.bitcoin.com/thailand-crypto-regulations-legalizing-cryptocurrencies/

Hern, A. (2013). *Bitcoin hype worse than 'tulip mania', says Dutch central banker | Technology | The Guardian*. Retrieved from https://www.theguardian.com/technology/2013/dec/04/bitcoin-bubble-tulip-dutch-banker

Hileman, G., & Rauchs, M. (2017). *GLOBAL CRYPTOCURRENCY BENCHMARKING STUDY 2 3 Global Cryptocurrency Benchmarking Study.*

Hill, D. (1985). Viewer Characteristics and Agenda Setting by Television News. *The Public Opinion Quarterly, 49*(3), 340 - 350.

Hodge, M. (2018). *Who is Satoshi Nakamoto? Bitcoin creator whose identity is unknown but could be one of the richest people in the world*. Retrieved from https://www.thesun.co.uk/news/5037060/satoshi-nakamoto-bitcoin-inventor-richest-world/

Horesh, N. (2012). *From Chengdu to Stockholm: A Comparative Study of the Emergence of Paper Money in East and West.*

Hott, C., & Monnin, P. (2006). Fundamental real estate prices: An empirical estimation with international data. *Journal of Real Estate Finance and Economics, 36*(4), 427-450.

Howcroft, D., Richardson, H., & Wilson, M. (2001). *NOW YOU SEE IT… NOW YOU DON'T" MYTHS OF THE DOT.COM MARKET.*

Huillet, M. (2018). *Confirmed: Nasdaq's Bitcoin Futures Will Launch in 'First Half' of 2019*. Retrieved from https://cointelegraph.com/news/confirmed-nasdaqs-bitcoin-futures-will-launch-in-first-half-of-2019

Imbert, F. (2017). *JPMorgan's Dimon: Bitcoin is a fraud that will eventually blow up*. Retrieved from https://www.cnbc.com/2017/09/12/jpmorgan-ceo-jamie-dimon-raises-flag-on-trading-revenue-sees-20-percent-fall-for-the-third-quarter.html

Iyengar, S., Peters, M., & Kinder, D. (1982). Experimental Demonstrations of the {\textquotedblleft}Not-So-Minimal{\textquotedblright} Consequences of Television News Programs. *American Political Science Review, 76*(4), 848-858.

Jones, B. (2014). *Identifying Speculative Bubbles: A Two-Pillar Surveillance Framework Identifying Speculative Bubbles: A Two-Pillar Surveillance Framework 1 The author is grateful for comments from Tamim Bayoumi.*

Joosten, N. (2012). *The effects of the Dot Com bubble and the Credit Crisis on leverage ratios of US non-financial firms.* Tilburg School of Economic and Management .

Juhász, P., Stéger, J., Kondor, D., & Vattay, G. (2016). A Bayesian Approach to Identify Bitcoin Users. *PLoS ONE*(January 2017).

Kaminska, I. (2013). *The Hubble bubble theory of the continuous expansion of the financial universe | FT Alphaville*. Retrieved from https://ftalphaville.ft.com/2013/12/06/1715892/the-hubble-bubble-theory-of-the-continuous-expansion-of-the-financial-universe/

Kaminski, J., Demchik, M., Timilsina, N., Genuer, R., Poggi, J.-m., & Tuleau-malot, C. (2019). Variable selection using Random Forests To cite this version :. *Pattern Recognition Letters, 117*(3), 256-266.

Kaur, A., & Gupta, V. (2013). A survey on sentiment analysis and opinion mining techniques. *Journal of Emerging Technologies in Web Intelligence, 5*(4), 367-371.

Kearns, J. (2013). *Greenspan Says Bitcoin a Bubble Without Intrinsic Currency Value*. Retrieved from https://www.bloomberg.com/technology

Kenyon-Dean, K., Ahmed, E., Fujimoto, S., Georges-Filteau, J., Glasz, C., Kaur, B., . . . Ruths, D. (2018). Sentiment Analysis: It's Complicated! 1886-1895.

Kharpal, A. (2018). *Tether: What you need to know about the cryptocurrency worrying markets*. Retrieved from https://www.cnbc.com/2018/02/02/tether-what-you-need-to-know-about-the-cryptocurrency-worrying-markets.html

Khatri, Y. (2018). *Nearly $1 Billion Stolen In Crypto Hacks So Far This Year: Research - CoinDesk*. Retrieved from https://www.coindesk.com/nearly-1-billion-stolen-in-crypto-hacks-so-far-this-year-research

Khoo, C., Nourbakhsh, A., & Na, J. (2012). Sentiment analysis of online news text: A case study of appraisal theory. *Online Information Review, 36*(6), 858-878.

Kiousis, S., Min Park1, J., Kim, Y., & Go, E. (2013). Exploring the Role of Agenda-Building Efforts in Media Coverage and Policymaking Activity of Healthcare Reform. *Journalism & Mass Communication Quarterly, 90*(4), 652 - 672.

Koehrsen, W. (0). *Will Koehrsen's LinkedIn Profile*. Retrieved from https://www.linkedin.com/in/william-koehrsen-48a643a5

Koehrsen, W. (2017). *Random Forest Simple Explanation*. Retrieved from https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

Koehrsen, W. (2019). *Weill Koehrsen - Personal Website*. Retrieved from https://willk.online/

Kraslawski, A. (2017). The Prospects Of Bitcoin As A Driver Of Economic Changes. (April).

Krippendorff, K. (2003). *Content Analysis: An Introduction to Its Methodology Ch2 and 4.*

Kubicová, I., & Komárek, L. (2011). The classification and identification of asset price bubbles. *Finance a Uver - Czech Journal of Economics and Finance, 61*(1), 34-48.

Kumar, A., & Sebastian, T. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications, 4*(10), 1-14.

Labonte, M. (2011). Inflation: Causes, Costs, and Current Status Specialist in Macroeconomic Policy. *Congressional Research Service , 7*(5700).

Lánský, J. (2017). Bitcoin System. *Acta Informatica Pragensia, 6*(1), 20-31.

Ledger Wallet. (2019). *Ledger Documentation Hub Release 2.*

Lee, J., & Hahn, K. (2014). Factors influencing the agenda-setting effects of newspapers on their subscribers a multi-level analysis. *12*(1), 192-233.

Lee, T. (2011). *The Bitcoin bubble*. Retrieved from http://timothyblee.com/2011/04/18/the-bitcoin-bubble/

Lee, T. (2011). *The Bitcoin Crash*. Retrieved from https://www.forbes.com/sites/timothylee/2011/08/07/the-bitcoin-crash/#5aaa05096f4d

Lewis, A. (2015). *A Gentle Introduction To Bitcoin Mining.*

Lima, M., Nascimento, T., Labidi, S., Timbó, N., Batista, M., Neto, G., . . . Sonia, R. (2016). U Sing S Entiment a Nalysis for S Tock. *7*(1), 59-67.

Liu, B. (2010). NLP-handbook-sentiment-analysis.pdf. 1-38.

Liu, B. (2015). *Introduction & The Problem of Sentiment Analysis.*

Lo, A. (2007). Efficient markets hypothesis. *The New Palgrave: A Dictionary of Economics*(2), 1-28.

Lo, S., & Wang, J. (2018). Bitcoin as Money? Motivation. (14), 1-28.

Loughran, T., & McDonald, B. (2008). Internet appendix for "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". *Journal of Finance, 66*(1), 35-65.

Lucey, B., & O'Connor, F. (2013). Bubbles in Gold? *The Alchemist*(69), p. 23.

Luke Parker. (2015). *The Decline in Bitcoin Full Nodes » Brave New Coin*. Retrieved from https://bravenewcoin.com/insights/the-decline-in-bitcoins-full-nodes

Macdonell, A. (2014). Popping the Bitcoin Bubble: An application of log--periodic power law modeling to digital currency. (December 2013), 1-33.

Mack, E. (2017). *As Bitcoin Flirts With $20,000, Let's Revisit Its Earlier Crashes*. Retrieved from https://www.forbes.com/sites/ericmack/2017/12/16/bitcoin-cryptocurrency-crash-cash-price-fork-futures/#4c06d05b1bd4

Madore, P. (2018). *Quantitative Hedge Funds: Wall Street's Own Mt. Gox Willy Bot(s)?* Retrieved from https://www.ccn.com/quantitative-hedge-funds-mt-gox-willy-bots

Malhotra, P., & Malhotra, S. (2012). The Impact of Mass Media Communication on Stock Trading Decisions : An Empirical Study. *4*(6), 57-63.

Mäntylä¹, M., Graziotin, D., Kuutila, M., & Mäntylä, M. (2018). The Evolution of Sentiment Analysis. *27*(February), 16-32.

Maragoudakis, M., & Serpanos, D. (2010). Towards stock market data mining using enriched random forests from textual resources and technical indicators. *IFIP Advances in Information and Communication Technology, 339 AICT*, 278-286.

McCombs, M., & Shaw, D. (1972). The agenda-setting function of mass media. *American Association for Public Opinion Research, 36*, 176-187.

McCombs, M., Llamas, J., Lopez-Escobar, E., & Rey, F. (1997). Candidate images in Spanish elections: Second-level agenda-setting effects. *Journalism and Mass Communication Quaterly, 74*(4), 703-717.

Mccombs, M., Lopez-Escobar, E., & Llamas, J. (2000). Setting the Agenda of Attributes in the 1996 Spanish General Election. 77-92.

Medipally, S. (2018). *Herding and the Dotcom Bubble*. Retrieved from Herding and the Dotcom Bubble: https://nickledanddimed.wordpress.com/2018/08/01/herding-and-the-dotcom-bubble/

Mercille, J. (2014). The Role of the Media in Sustaining Ireland's Housing Bubble. *New Political Economy, 19*(2), 282-301.

Mitchell, T. (1997). Decision Tree Learning. In T. Mitchell, *Machine Learning* (pp. 52 - 80). McGraw Hill.

Mizrahi, A. (2018). *Thomson Reuters Eikon to Display Data on 50 Cryptocurrencies From Cryptocompare - Bitcoin News*. Retrieved from https://news.bitcoin.com/thomson-reuters-eikon-to-display-data-on-50-cryptocurrencies-from-cryptocompare/

Montag, A. (2018). *Nobel prize-winning economist Joseph Stiglitz criticizes bitcoin*. Retrieved from https://www.cnbc.com/2018/07/09/nobel-prize-winning-economist-joseph-stiglitz-criticizes-bitcoin.html

Moreo, A., Romero, M., Castro, J., & Zurita, J. (2012). Lexicon-based Comments-oriented News Sentiment Analyzer system. *Expert Systems with Applications, 39*(10), 9166-9180.

Morris, D. (2017). *Bitcoin Hits a New Record High, But Stops Short of $20,000 | Fortune*. Retrieved from http://fortune.com/2017/12/17/bitcoin-record-high-short-of-20000/

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System.*

Nakamoto, S. (2009). *Bitcoin open source implementation of P2P currency - P2P Foundation*. Retrieved from http://p2pfoundation.ning.com/forum/topics/bitcoin-open-source

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. (January 2003).

Neuman, R., & Guggenheim, L. (2011). The Evolution of Media Effects Theory: A Six-Stage Model of Cumulative Research. *Communication Theory, 21*, 169-196.

Norry, A. (2018). *The History of the Mt Gox Hack: Bitcoin's Biggest Heist*. Retrieved from https://blockonomi.com/mt-gox-hack/

O'Brien, K. (2018). *Now You Can Buy A Texas Mansion With Bitcoin - Bitcoinist.com*. Retrieved from https://bitcoinist.com/buy-texas-mansion-real-estate-bitcoin/

Ofek, E., & Richardson, M. (2003). DotCom Mania: The Rise and Fall of Internet Stock Prices. *Journal of Finance, 58*(3 (June)), 1113--1137.

Ogundeji, O. (2016). *Antonopoulos: Your Keys, Your Bitcoin. Not Your Keys, Not Your Bitcoin*. Retrieved from https://cointelegraph.com/news/antonopoulos-your-keys-your-bitcoin-not-your-keys-not-your-bitcoin

Okhuese, A. (2017). Introducing cryptocurrency. *READS Capital, Schemas Group*(October), 1-2.

Oliphant, V. (2017). *Bitcoin 20000: Bubble crash fears as cryptocurrency hits record high value | City &amp; Business | Finance | Express.co.uk*. Retrieved from https://www.express.co.uk/finance/city/893471/Bitcoin-20000-bubble-crash-value-price-latest-cryptocurrency-finance

Oliver, J. (2018). *Cryptocurrencies: Last Week Tonight with John Oliver (HBO) - YouTube*. Retrieved from https://www.youtube.com/watch?v=g6iDZspbRMg

Olszewicz, J. (2019). *Bitcoin Price Analysis - A sustained and significant rise in transactions » Brave New Coin*. Retrieved from https://bravenewcoin.com/insights/bitcoin-price-analysis-a-sustained-and-significant-rise-in-transactions

Page, B. (2002). Pricking the bubble: financial scandal and the media. *British Journalism Review, 13*(3), 49-57.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *22*(45), 1 - 135.

Parmelee, J. (2014). The agenda-building function of political tweets. *New Media and Society, 16*(3), 434-450.

Peter D. DeVries. (2016). An Analysis of Cryptocurrency, Bitcoin, and the Future (PDF Download Available). *3*(University of Houston - Downtown), 899.

Pflugfelder, E. (2017). Reddit's “Explain Like I'm Five”: Technical Descriptions in the Wild. *Technical Communication Quarterly, 26*(1), 25-41.

Pham-Gia, T., & Choulakian, V. (2014). Distribution of the Sample Correlation Matrix and Applications. *Open Journal of Statistics, 04*(05), 330-344.

Rambocas, M., & Gama, J. (2013). Marketing Research : The Role of Sentiment Analysis. *Universidade do Porto, Faculdade de Economia do Porto*(April), 1-24.

Rapoza, K. (2017). *What China Ban? Cryptocurrency Market Cap Rebounding*. Retrieved from https://www.forbes.com/sites/kenrapoza/2017/09/28/china-ico-ban-bitcoin-crypto-currency-market-cap-returns/#7d03cb056c21

Reeb, B. (2010). FT.com (Financial Times). *Journal of Business and Finance Librarianship, 15*(1), 31-36.

Rogers, E., & Dearing, J. (1988). Agenda-Setting Research: Where Has It Been, Where Is It Going? *Annals of the International Communication Association, 11*(1), 555-594.

Romer, C. (2003). Great Depression. *Encyclopedia Britannica*, 0.

Roskos-Ewoldsen, D., & Roskos-Ewoldsen, B. Carpentier, F. (2009). Media priming: A synthesis. *Media Effects: Advances in Theory and Research,*(February), 74 - 93.

Rothbard, M. (2010). *America's Great Depression.* (Vol. 5).

Ruscheinsky, J., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment and Finance, 36*(5), 410-428.

Scheufele, B., Haas, A., & Brosius, H. (2011). Mirror or Molder? A Study of Media Coverage, Stock Prices, and Trading Volumes in Germany. *Journal of Communication, 61*(1), 48-70.

Scheufele, D., & Tewksbury, D. (2007). Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models.

Schmidt, T., & Burghardt, M. (2018). An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. (2005), 139-149.

Schulder, M., Wiegand, M., Ruppenhofer, J., & Roth, B. (2017). Towards Bootstrapping a Polarity Shifter Lexicon using Linguistic Features. 624-633.

Scott A. Wolla, P. (2018). Bitcoin: Money or Financial Investment? *PAGE ONE Economics*(November), 2-4.

Segendorf, B. (2014). What is Bitcoin? *XRDS: Crossroads, The ACM Magazine for Students, 20*(1), 40.

Semetko, H., & Valkenburg, P. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication, 50*(2), 93-109.

Shiller, R. (2000). *Irrational Exhuberance.* y Princeton University Press.

Shin, L. (2017). *Will This Battle For The Soul Of Bitcoin Destroy It?* Retrieved from https://www.forbes.com/sites/laurashin/2017/10/23/will-this-battle-for-the-soul-of-bitcoin-destroy-it/#4a5388603d3c

Siddharth, S., Darsini, R., & Sujithra, M. (2018). Sentiment Analysis on Twitter Data Using Machine Learning Algorithms in Python. *ISSN (Online) 2394-2320 International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), 5*(2).

Smith, M., & Smith, G. (2007). Bubble, Bubble, Where's the Housing Bubble? *Brookings Papers on Economic Activity, 2006*(1), 1-67.

Soldevilla Estrada, J. (2017). Analyzing Bitcoin Price Volatility. *University of California, Berkley*, 1-49.

Sommerlad, J. (2018). *John Oliver attacks cryptocurrency craze: 'You're not investing. You're gambling' | The Independent*. Retrieved from https://www.independent.co.uk/life-style/gadgets-and-tech/news/bitcoin-price-latest-john-oliver-cryptocurrency-craze-investing-gambling-last-week-tonight-hbo-a8253326.html

Soroka, S., Farnsworth, S., Lawlor, A., & Young, L. (2015). Mass media and policy-making. In S. Soroka, S. Farnsworth, A. Lawlor, & L. Young, *Routledge Handbook of Public Policy.*

Southurst, J. (2013). *194,993 BTC transaction worth $147 Million sparks mystery and speculation - CoinDesk*. Retrieved from https://www.coindesk.com/194993-btc-transaction-147m-mystery-and-speculation

Squires, C. (2012). Coloring in the Bubble: Perspectives from Black-Oriented Media on the (Latest) Economic Disaster. *American Quarterly, 64*(3), 543-570.

Staiano, J., & Guerini, M. (2014). DepecheMood: a Lexicon for Emotion Analysis from Crowd-Annotated News. 427-433.

STATISTA. (2019). • *Bitcoin price index monthly 2016-2018 | Statistic*. Retrieved from https://www.statista.com/statistics/326707/bitcoin-price-index/

Statistics, M. (2014). DOCTORAL THESIS Mgr . Tom ´ aˇ s Hanz ´ ak Methods for periodic and irregular time series.

Strauß, N., Vliegenthart, R., & Verhoeven, P. (2016). Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands. *Public Relations Review, 42*(4), 548-555.

Strömbäck, J., & Kiousis, S. (2010). A New look at Agenda-setting effects - comparing the predictive power of overall political news consumption and specific news media consumption across different media channels and media types. *Journal of Communication, 60*(2), 271-292.

Suberg, W. (2017). *Bitcoin Hits $20,000 Per Coin, Capping Year of Enormous Growth*. Retrieved from https://cointelegraph.com/news/bitcoin-hits-20000-per-coin-capping-year-of-enormous-growth

Tetlock, P., Saar-Tsechansky, M., & Macskassy, S. (2008). American Finance Association More than Words: Quantifying Language to Measure Firms' Fundamentals More Than Words: Quantifying Language to Measure Firms' Fundamentals. *Source: The Journal of Finance THE JOURNAL OF FINANCE, 63*(3), 1437-1467.

Thelwall, M., & Prabowo, R. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics, 3*(2), 143-157.

Torabian, B. (2016). Sentiment Classification with Case-Based Approach Sentiment Classification with Case-Based Approach.

Townsend, R. (1980). Models with spatially separated agents.pdf.

Tsygankov, A. (2017). The dark double: The American media perception of Russia as a neo-soviet autocracy, 2008–2014. *Politics, 37*(1), 19-35.

Tully, S. (n.d.). *The NYSE's Owner Is Launching a Startup Exchange for Bitcoin | Fortune*. Retrieved from 2018: http://fortune.com/longform/nyse-owner-bitcoin-exchange-startup/

Turk, J., & Franklin, B. (1987). Information subsidies: Agenda-setting traditions. *Public Relations Review, 13*(4), 29 - 41.

Turney, P., & Littman, M. (2003). Measuring Praise and Criticism: Inference of Semantic Orientation from Association. 1-37.

Underwood, B. (2018). *Office of the New York State Attorney General VIRTUAL MARKETS INTEGRITY INITIATIVE REPORT.*

van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., & Schlobach, S. (2008). Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology and Politics, 5*(1), 73-94.

van der Veen, A. (2012). The Dutch Tulip Mania: The Social Foundations of a Financial Bubble. *University of Georgia, March, 3*, 1-46.

Vigna, P. (2018). *Pay Taxes With Bitcoin? Ohio Says Sure - WSJ*. Retrieved from https://www.wsj.com/articles/pay-taxes-with-bitcoin-ohio-says-sure-1543161720

Vijay, N., Singh, S., & Malhotra, G. (2018). Sentiment Analysis: Gauging the Effect of News on Stock Prices in Indian Stock Market. *International Journal of Trade, Economics and Finance, 9*(4), 148-152.

Vohra, M., & Teraiya, P. (2013). Journal of Information, Knowledge and Research in Computer Engineering a Comparative Study of Sentiment Analysis Techniques. *Journal of Information,Knowledge and Research in Computer Engineering*, 313-317.

Vujičić, D., Jagodić, D., & Randić, S. (2018). Blockchain technology, bitcoin, and Ethereum: A brief overview. *2018 17th International Symposium on INFOTEH-JAHORINA, INFOTEH 2018 - Proceedings, 2018-Janua*(August), 1-6.

Walgrave, S., & De Bruycker, I. (2013). How a New Issue Becomes an Owned Issue. Media Coverage and the Financial Crisis in Belgium (2008-2009). *International Journal of Public Opinion Research, 26*(1), 86-97.

Walker, C. (2014). Housing booms and media coverage. *Applied Economics, 46*(32), 3954-3967.

Wanta, W., Golan, G., & Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism and Mass Communication Quarterly, 81*(2), 364-377.

Wearden, G. (2017). *Bitcoin bubble warnings grow louder as futures trading begins on CME – as it happened | Business | The Guardian*. Retrieved from https://www.theguardian.com/business/live/2017/dec/18/bitcoin-bubble-ubs-futures-trading-20000-cme-stock-markets-tax-business-live

Wei Dai. (1998). *Wei Dai's Home Page*. Retrieved from http://www.weidai.com/

Wheelock, D., & Reserve Bank of St Louis, F. (n.d.). *The Great Depression: An Overview.*

Williams, M. (2013). *FINANCE PROFESSOR: Bitcoin Will Crash To $10 By Mid-2014*. Retrieved from https://www.businessinsider.com/williams-bitcoin-meltdown-10-2013-12?IR=T

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics, 35*(3), 399-433.

Wong, J. (2018). *Bitcoin Pizza Day 2018: Eight years ago, someone bought two pizzas with bitcoins now worth $82 million — Quartz*. Retrieved from https://qz.com/1285209/bitcoin-pizza-day-2018-eight-years-ago-someone-bought-two-pizzas-with-bitcoins-now-worth-82-million/

Worstall, T. (2013). *Central Banker On Bitcoin, At Least With Tulipmania You Got a Tulip At The End*. Retrieved from https://www.forbes.com/sites/timworstall/2013/12/05/central-banker-on-bitcoin-at-least-with-tulipmania-you-got-a-tulip-at-the-end/#6fe26dba297f

Yang, W., Lin, D., & Yi, Z. (2017). Impacts of the mass media effect on investor sentiment. *Finance Research Letters, 22*, 1-4.

Yao, W., & Liu, G. (2018). Study on the relationship between investor sentiment and stock bubble. *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, 1160-1165.

Yazdavar, A., & Sheth, A. (2017). On the Challenges of Sentiment Analysis for Dynamic Events, IEEE Intelligent Systems, 2017 (in print). *2017*.

Yoshinaga, C., & Junior, F. (2010). The Relationship between Market Sentiment Index and Brazilian Stock Rates of Return: a GMM Panel Data Analysis. *The 2010 Annual Meeting of …*(June), 189-210.

Young, J. (2018). *$194 Million was Moved Using Bitcoin With $0.1 Fee, Potential of Crypto*. Retrieved from https://www.ccn.com/194-million-was-moved-using-bitcoin-with-0-1-fee-true-potential-of-crypto/

Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication, 29*(2), 205-231.

Younis, E. (2005). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study Article. *International Journal of Computer Applications, 25*(2), 335-352.

Zoldan, A. (2013). *Bitcoin: Society's Boon or Bane? | HuffPost*. Retrieved from https://www.huffingtonpost.com/ari-zoldan/bitcoin-societys-boon-or-_b_3715125.html

CODE BOOK

Research Information

The thesis aims to gather insight into the possible interdependencies between online news media and the price movement of the largest and most prominent virtual cryptocurrency Bitcoin (BTC).

Articles to code

The news articles to be coded have been manually predefined. They are selected from the website of the daily newspaper *The Financial Times*, www.ft.com, and have been published between 14.01.2015 and 17.12.2017 – the time period of importance for this research. Relevant is every publication to be found on the website for the selected timeframe, by searching for the keyword "bitcoin". This results in 1058 articles, which include the term somewhere in their body of text. Since often articles have been sorted under multiple categories on FT.com, for example *"Cryptocurrencies", "Markets", "Bitcoin", Fintech*, and there is not clear boundary between the various categories, all publications underwent through a manual selection process, filtering out these which do not refer to Bitcoin or to a meaningful related matter. As a result, 244 Bitcoin – relevant articles were selected. In order to access unlimited number of articles and exclusive premium content, monthly subscription is required. All of the preselected publications are to be manually annotated for the following variables:

DATE [DATE] – The date of publishing should be noted in the format *mm/dd/yyyy*. Financial Times are medium of high quality, follow strict formatting rules, and article without exact date is highly unlikely to be published. Date is vital for figuring out density and regularity of news.

TITLE [TITLE] – The full title of the article (without the subtitle) is another variable. It will function as a mark for identification, since publications on FT.com are not provided with unique ID or other way for their later identification.

TITLE SENTIMENT [TITLE_SENT] – Titles are the first reference towards the article itself. Often headlines are metaphorical or do not seem to be relevant, but a second look shows otherwise. The sentiment towards Bitcoin, expressed through the title should be classified in one of the three categories: *Negative, Neutral or Positive.* No clearly defined polarity is to be sorted as *Neutral*. In case the title fails to undoubtedly suggest what the main object of the publication is, but clearly shows negative or positive attitude towards it, the headline is to be classified that same (*negative or positive*) way.

ARTICLE SENTIMENT [BODY_SENT] – The annotator must define if the article expresses *Negative, Neutral or Positive* sentiment towards Bitcoin. The person reviewing the publications should have deep understanding of both cryptocurrencies and the underlying technology. It might be challenging to differentiate the polarity of articles, however a negative one would be such that aims to emphasize of BTC being useful for nothing else but drug trafficking and child porn, an exaggeration of the electricity costs associated with the network running smoothly or that the virtual currency is just another Ponzi scheme. A positive article would outline the benefits of Bitcoin compared to the traditional FIAT currencies or alternative cryptocurrencies. Decentralization, lack of single point of failure, high speed, low transaction costs, irreversibility and non-dependency on trusted vendors are a few favorable points. Of course, it could be that a publication objectively mentions both pros and cons of the cryptocurrency, as done in the Bitcoin introductory chapters of that thesis. Depending on the choice of words and their formulation, the article in review might project any of the three attitudes – it is the annotator's task to make use of his knowledge, best interpretation skills and pinpoint the correct one.

BUBBLE IMPLICATION [BUBBLE] - Just as with the attitude, there are various ways to hint at a significant correction. It is once again up to the annotator to recognize if the author indirectly says Bitcoin is to go through violent implosion, or simply normal volatility is being discussed. The possible codes for this category are *NO* and *YES*, where under the former one are to be sorted any publications that (1) do not discuss / suggest the asset being in a bubble, (2) explicitly mention BTC is not in a bubble. Under the latter category are to be sorted articles that suggest, hint or directly say Bitcoin is overvalued.

NEWS ARTICLE GENRE [NAG] – the articles should be categorized according to their type. There is a total of 5 major categories;

1. **NEWS** – articles that shares newsworthy objective facts, for example price highlights, volatility or reportages. Only moderate amount of personal commentary is allowed, generally very short.

2. **STORY** – articles, that in a similar fashion to news, share recent happenings or facts. However, the story is usually more detailed, often referring to the past or other related events. Again only moderate amount of personal commentary is present, and are longer than news.
3. **ANALYSIS** – more detailed articles, which would take a deeper look into Bitcoin, its ecosystem or related matters. Publications usually include opinions of the media or the author, who can apply various ways to defend his point of view. It could be the case, where an analysis presents various opinions.
4. **INTERVIEW** – article that includes and mainly develops around an interview of one or more people.
5. **OTHER** – under this category are to be sorted publications, which cannot be categorized among one of the previous ones like caricatures, collages and short Q & As for example.

ALPHAVILLE CATEGORY [ALPHAVILLE] – publications, that are part of Alphaville, FT's financial team blog. Easy to distinguish due to the branded ALPHAVILLE tag on top of the page.

PREMIUM PUBLICATION [PREMIUM] – premium articles, accessible only to members, who are on one of FT's paid subscription plans. Marked by a small black "PREMIUM" badge in the articles as well among the search results.

LENGTH IN WORDS [WORDS] – length of the article in words, excluding the title and image captions. Microsoft Office Word is to be used.

LENGTH IN CHARACTERS [CHARS] – length of the publication in characters without / with empty spaces. Use Microsoft Office Word.

Bitcoin Price Parameters

Various price features of Bitcoin are also to be extracted. For that, BLX Index is to be used, available on TradingView.com The following parameters are to be noted, for every day, when a relevant article was published:

OPENING PRICE [DAILY OPEN] – the opening price of Bitcoin for the day, in USD. **CLOSING PRICE [DAILY CLOSE]** – the closing price of Bitcoin for the day, in USD.

DAILY HIGHEST PRICE [DAILY HIGH] – the highest price that Bitcoin hit for the day, in USD.

DAILY LOWEST PRICE [DAILY LOW] – the lowest price that Bitcoin touched for the day, in USD.

DAILY AVERAGE PRICE [DAILY AVG] – the average price of Bitcoin for the day, in USD. It is the calculation of the arithmetic mean between daily low and daily high.

These parameters are sufficient for subsequent calculations regarding price changes and fluctuations. Historical daily price for the whole period is to be automatically extracted via Google Finance, however, the tool only provides the closing price of BTC.

ABSTRACT IN ENGLISH

Bitcoin, the fintech phenomenon, incepted by the anonymous Satoshi Nakamoto back in 2009, hit a valuation short of $20, 000 in late 2017. There are a number of suggestions what caused this latest continuous price rise of roughly 12,000% in less than three years, resulting in yet another bubble burst – from unfair exchange play to anonymous actors in the market, that stay in the shadows.

This thesis explores the role media might have played. Effects of media on its audience have been for long studied in various settings, and there are convincing evidences that it can indeed steer and alter opinions. It is certainly of high interest, if the irrational exuberance of investors in the field of blockchain and cryptocurrencies is to be blamed on media.

This research presents several cases where media directly affected the pricing of assets or caused bubbles. Further, relying completely on manual means, the study explores how *The Financial Times (http://www.ft.com)* reported on Bitcoin, as well as the sentiment expressed in the articles towards the cryptocurrency and its related matters, during the timeline of the bubble formation – from its starting point on 14.01.2015 till its top on 17.12.2017. Due to the sparsity of data, lagging between Bitcoin's valuation and polarity of the publications could not be explored. However, relying on random forest, a supervised machine – learning method, it was confirmed that sentiment of the articles as well as sentiment of their titles are the most important features defining intraday price actions.

Keywords : Bitcoin, Cryptocurrency, Bitcoin bubble, Financial bubble, Sentiment analysis, Media effects, Random forest

ABSTRACT IN GERMAN

Bitcoin, das fintech Phänomen, das 2009 von dem anonymen Satoshi Nakamoto eingeführt wurde, erreichte Ende 2017 einen Wert von knapp 20.000 $. Es gibt eine Reihe von Vorschlägen, was diesen letzten kontinuierlichen Preisanstieg von rund 12.000 % in weniger als drei Jahren verursacht hat, was zu einem weiteren Blasenbruch führte - vom unfairen Börsenspiel bis hin zu anonymen Akteuren auf dem Markt, die im Schatten stehen.

Diese Arbeit untersucht die Rolle, die Medien gespielt haben könnten. Die Auswirkungen der Medien auf ihr Publikum werden seit langem in verschiedenen Kontexten untersucht, und es gibt aussagekräftige Beweise dafür, dass sie tatsächlich Meinungen steuern und verändern können. Es ist sicherlich von großem Interesse, wenn der irrationale Exuberanz von Investoren im Bereich Blockchain und Kryptowährungen den Medien zugeschrieben werden soll.

Diese Studie stellt mehrere Fälle vor, in denen Medien direkt die Preise von Anlagen beeinflussten oder Bubbles verursachten. Darüber hinaus untersucht die Studie, die sich vollständig auf manuelle Methoden stützt, wie *The Financial Times (http://www.ft.com)* über Bitcoin berichtete, sowie die in den Artikeln geäußerte Sentiment in Bezug auf die Kryptowährung und die damit verbundenen Sachverhalte während der Timeline der Blasenbildung - von ihrem Startpunkt am 14.01.2015 bis zu ihrer Spitze am 17.12.2017. Aufgrund der Datenknappheit konnten Verzögerungen zwischen Bitcoin Preis und der Polarität der Publikationen nicht untersucht werden. Unter Verwendung von Random Forest, einer maschinellen Lernmethode, wurde jedoch bestätigt, dass die Sentiment der Artikel sowie die Sentiment der Titel die wichtigsten Elemente sind, die tägliche Preisaktionen definieren.

Keywords: Bitcoin, Kryptowährung, Bitcoinblase, Finanzblase, Sentimentanalyse, Medienwirkung, Random forest.

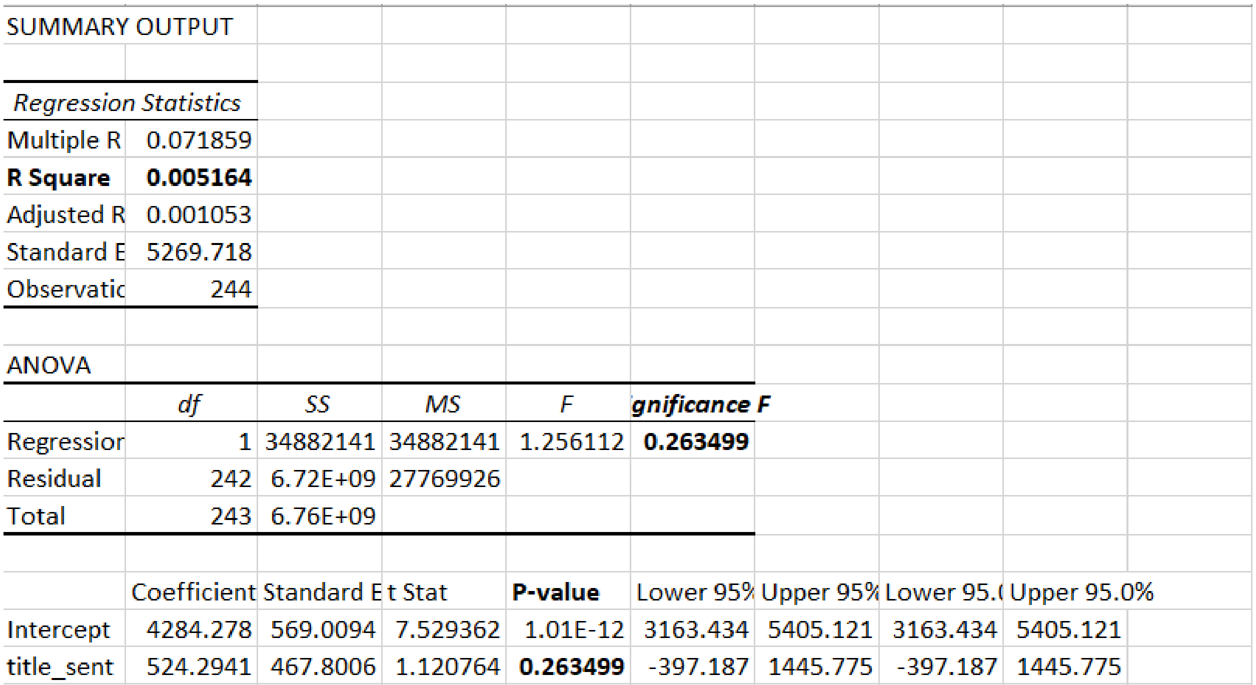
APPENDIX

Numerous linear (with one independent variable) and multiple regression (with at least two independent variables) analyses were run. Five price codes served as response variables - [1] daily high, [2] daily average, [3] absolute change, [4] percentage change and the change tag one [5]. In separate linear regression models, the response of each and every one of these was measured against each of the following independent variables [a] title sentiment, [b] body sentiment, [c] bubble implication, [d] length in words and [e] news article genre. Multiple regression plots were run with [I] all independent variables combined, from [a] to [e], and [II] the purely sentiment – loaded predictors, [a] title sentiment and [b] body sentiment.

The results are shown below. They are titled with the following format : ***dependent variable & independent variable(s)***

LINEAR REGRESSION PLOTS

[1] Daily High & [a] Title Sentiment



SUMMARY OUTPUT

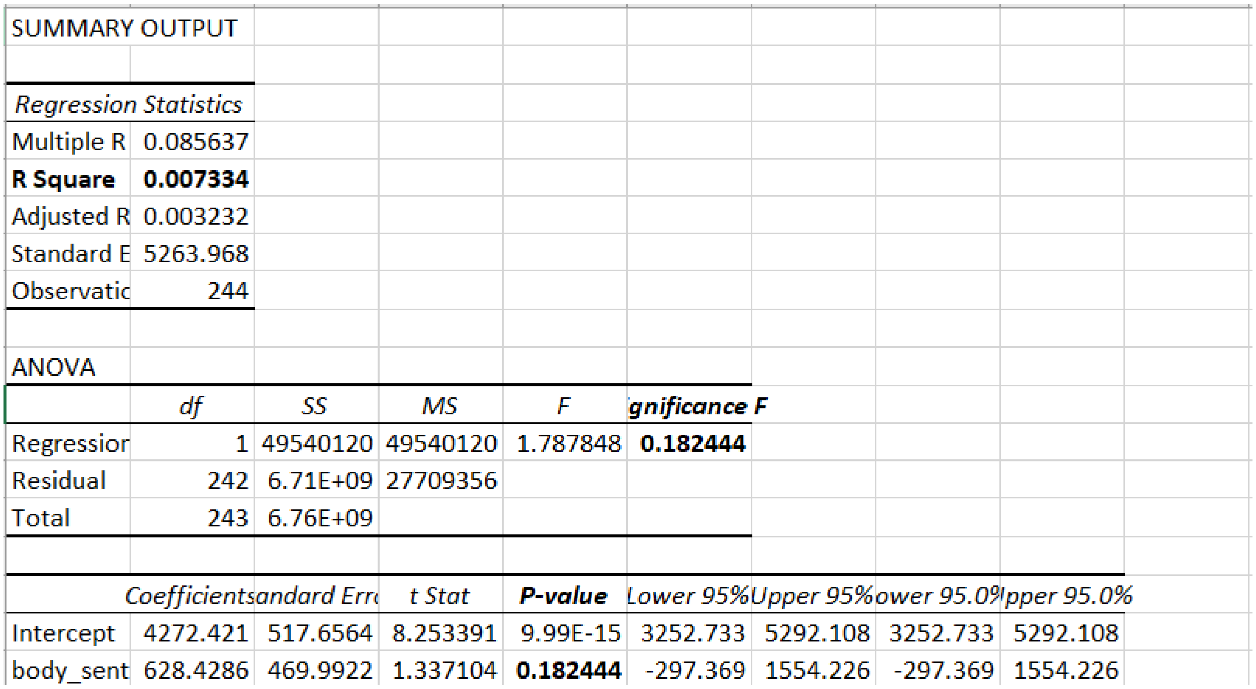
Regression Statistics	
Multiple R	0.071859
R Square	**0.005164**
Adjusted R	0.001053
Standard E	5269.718
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	34882141	34882141	1.256112	**0.263499**
Residual	242	6.72E+09	27769926		
Total	243	6.76E+09			

	Coefficient	Standard E	t Stat	**P-value**	Lower 95%	Upper 95%	Lower 95.(	Upper 95.0%
Intercept	4284.278	569.0094	7.529362	1.01E-12	3163.434	5405.121	3163.434	5405.121
title_sent	524.2941	467.8006	1.120764	**0.263499**	-397.187	1445.775	-397.187	1445.775

[1] Daily High & [b] Body Sentiment



SUMMARY OUTPUT

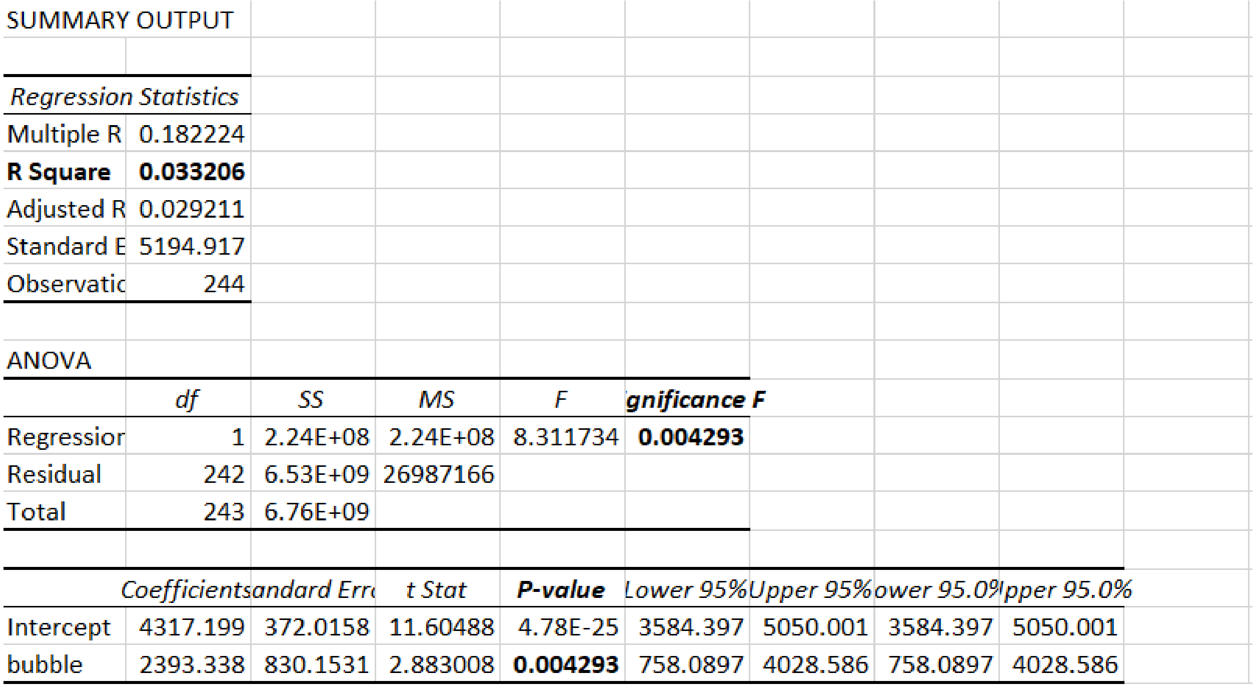
Regression Statistics	
Multiple R	0.085637
R Square	**0.007334**
Adjusted R	0.003232
Standard E	5263.968
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	49540120	49540120	1.787848	**0.182444**
Residual	242	6.71E+09	27709356		
Total	243	6.76E+09			

	Coefficients	*andard Errc*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4272.421	517.6564	8.253391	9.99E-15	3252.733	5292.108	3252.733	5292.108
body_sent	628.4286	469.9922	1.337104	**0.182444**	-297.369	1554.226	-297.369	1554.226

[1] Daily High & [c] Bubble Implication



SUMMARY OUTPUT

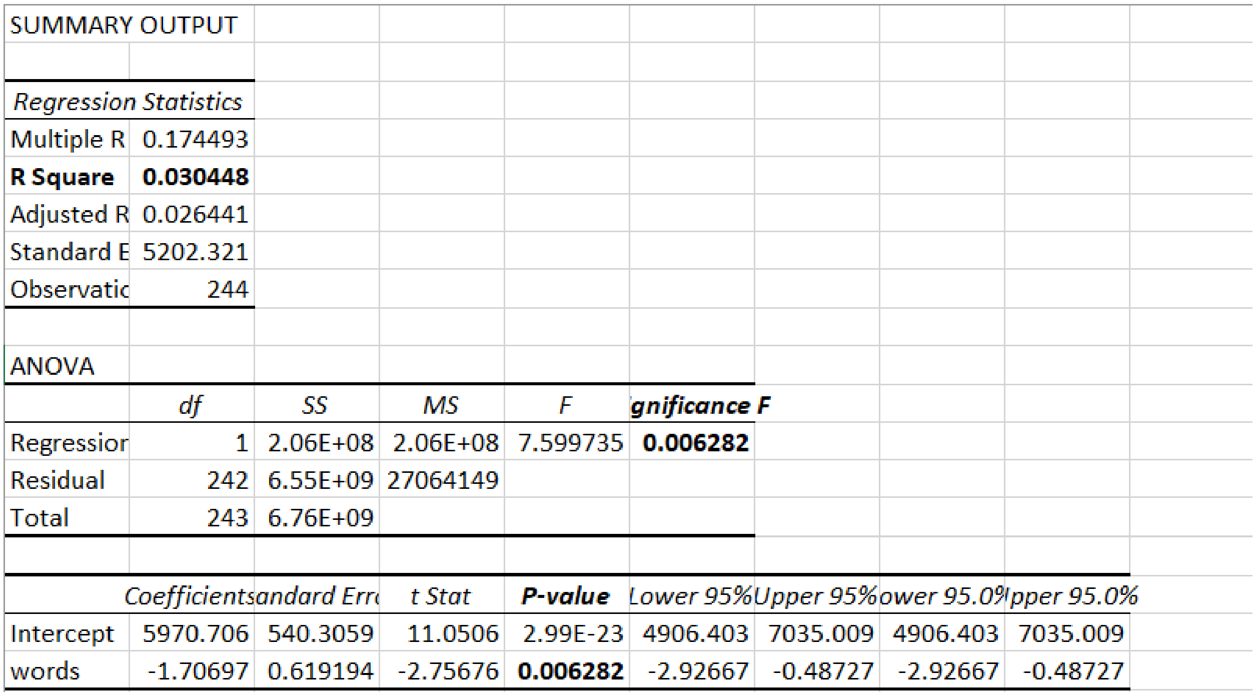
Regression Statistics	
Multiple R	0.182224
R Square	**0.033206**
Adjusted R	0.029211
Standard E	5194.917
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	2.24E+08	2.24E+08	8.311734	**0.004293**
Residual	242	6.53E+09	26987166		
Total	243	6.76E+09			

	Coefficients	*andard Errc*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4317.199	372.0158	11.60488	4.78E-25	3584.397	5050.001	3584.397	5050.001
bubble	2393.338	830.1531	2.883008	**0.004293**	758.0897	4028.586	758.0897	4028.586

[1] Daily High & [d] Length In Words



SUMMARY OUTPUT

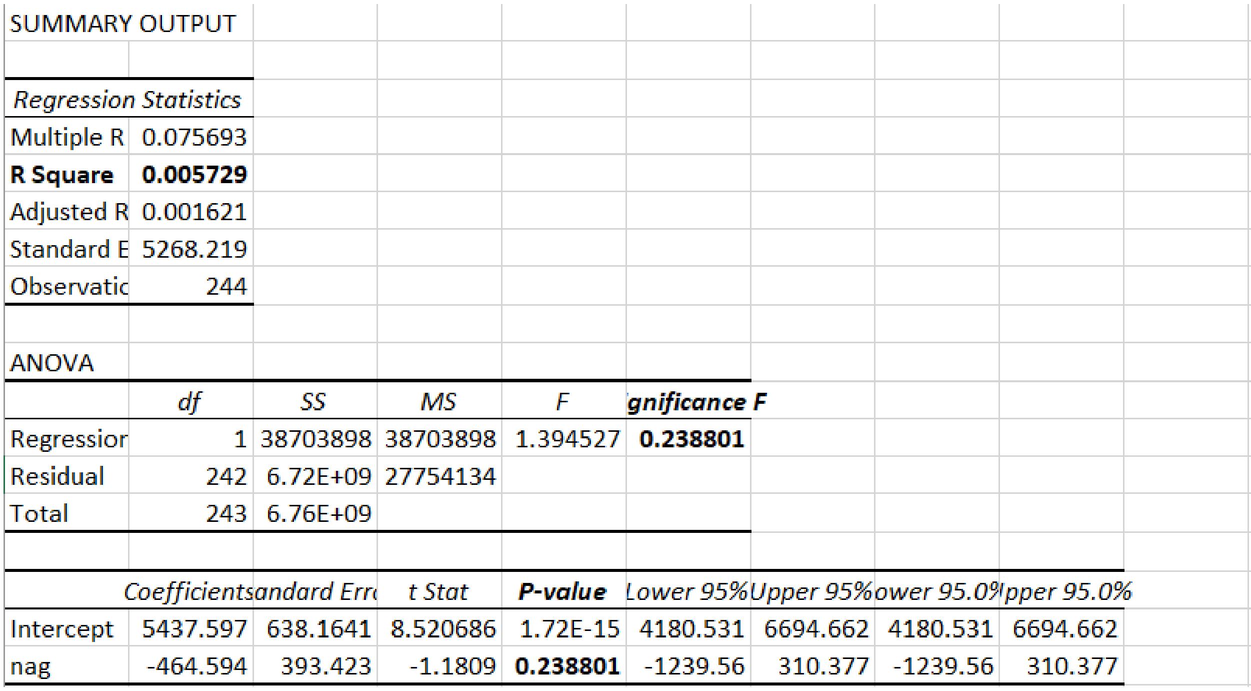
Regression Statistics	
Multiple R	0.174493
R Square	**0.030448**
Adjusted R	0.026441
Standard E	5202.321
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	2.06E+08	2.06E+08	7.599735	**0.006282**
Residual	242	6.55E+09	27064149		
Total	243	6.76E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	5970.706	540.3059	11.0506	2.99E-23	4906.403	7035.009	4906.403	7035.009
words	-1.70697	0.619194	-2.75676	**0.006282**	-2.92667	-0.48727	-2.92667	-0.48727

[1] Daily High & [e] News Article Genre



SUMMARY OUTPUT

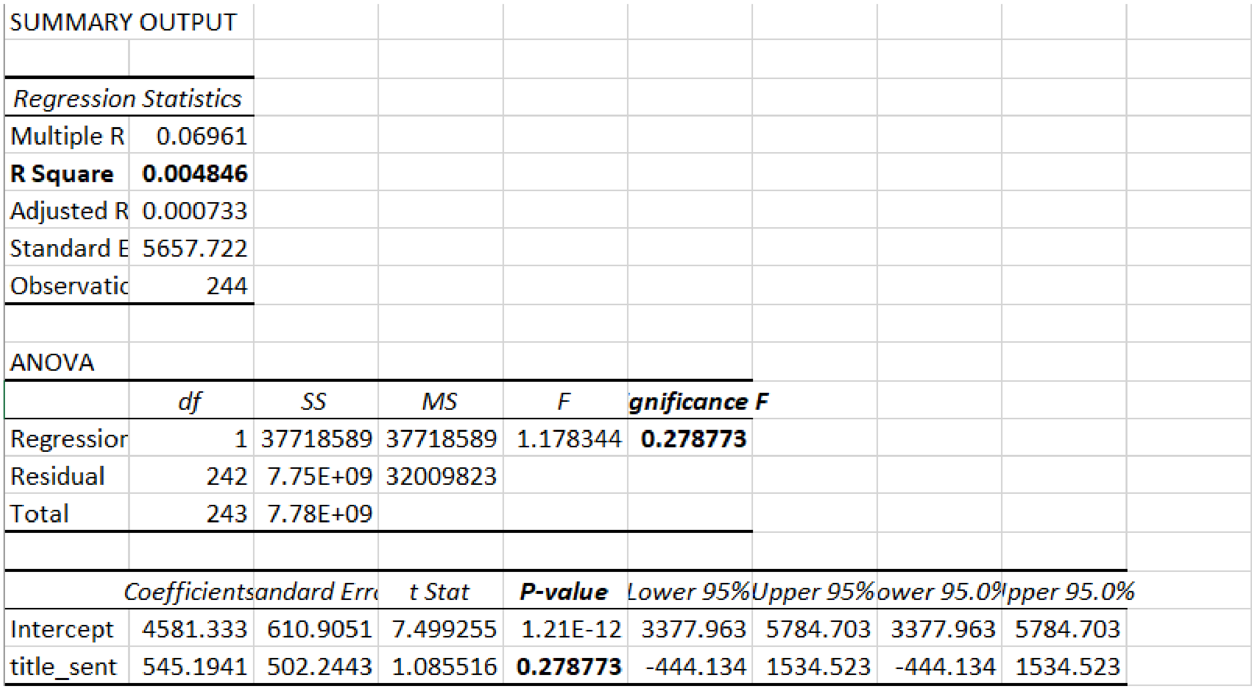
Regression Statistics	
Multiple R	0.075693
R Square	**0.005729**
Adjusted R	0.001621
Standard E	5268.219
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	38703898	38703898	1.394527	**0.238801**
Residual	242	6.72E+09	27754134		
Total	243	6.76E+09			

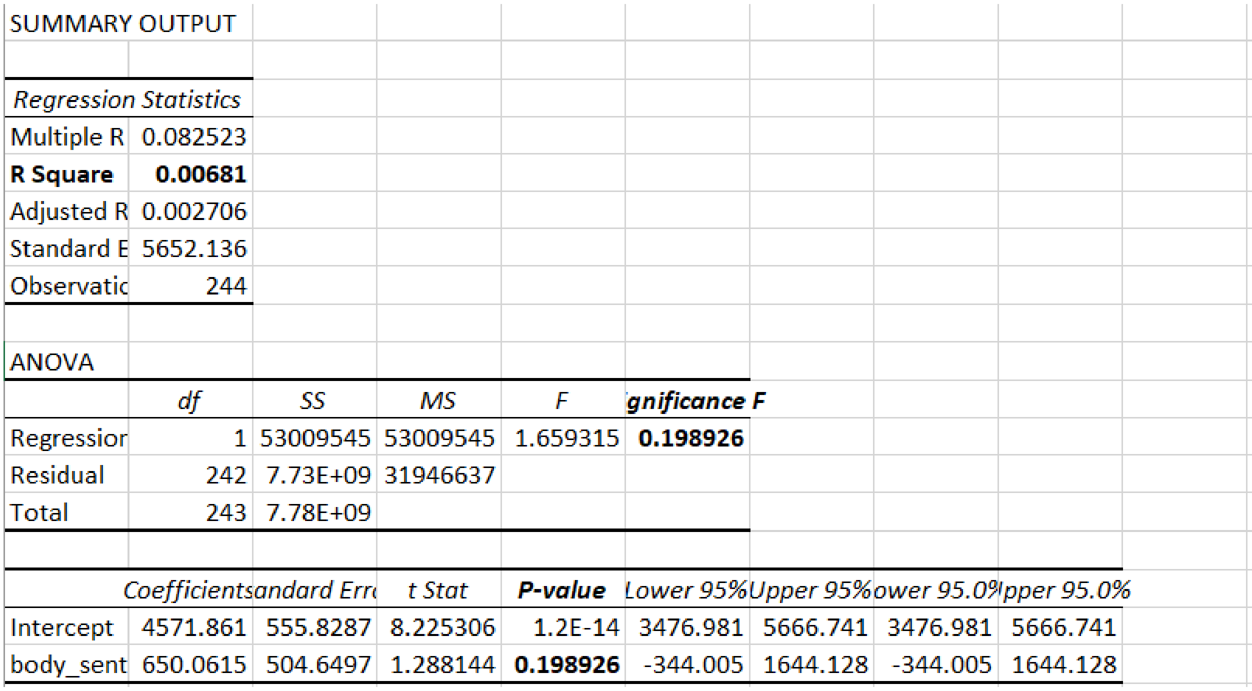
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	5437.597	638.1641	8.520686	1.72E-15	4180.531	6694.662	4180.531	6694.662
nag	-464.594	393.423	-1.1809	**0.238801**	-1239.56	310.377	-1239.56	310.377

[2] Daily Average & [a] Title Sentiment



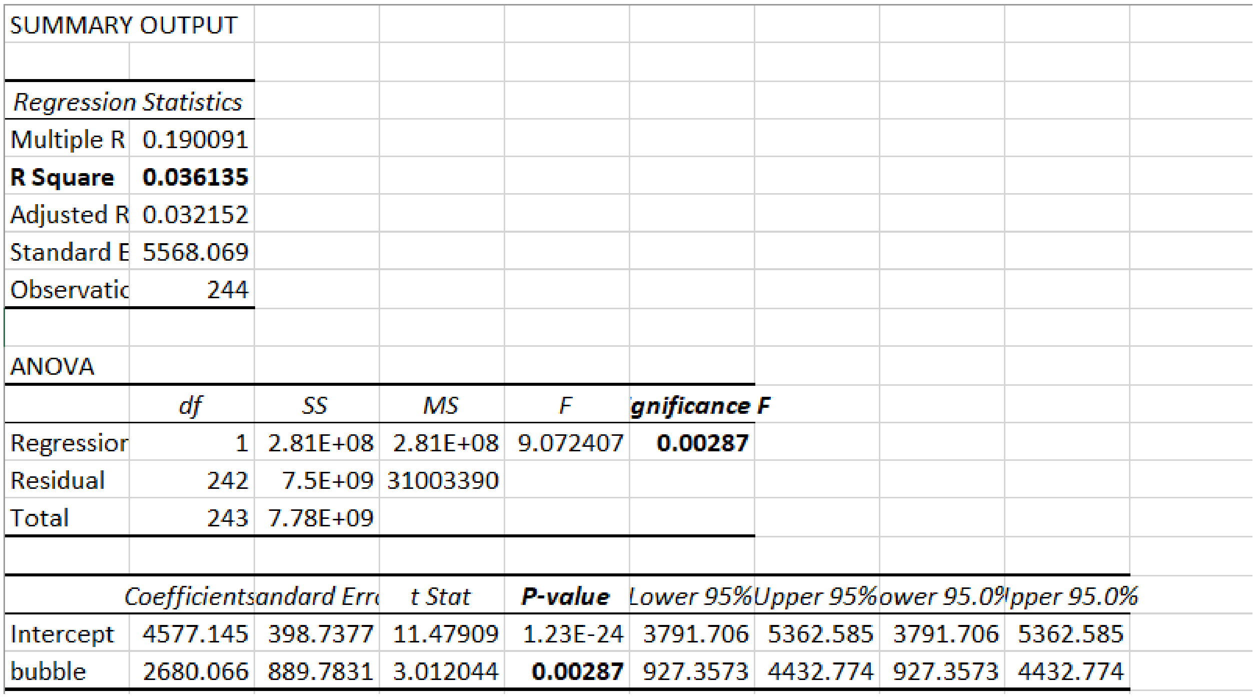
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.06961							
R Square	**0.004846**							
Adjusted R	0.000733							
Standard E	5657.722							
Observatic	244							
ANOVA								
	df	*SS*	*MS*	*F*	***gnificance F***			
Regressior	1	37718589	37718589	1.178344	**0.278773**			
Residual	242	7.75E+09	32009823					
Total	243	7.78E+09						
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4581.333	610.9051	7.499255	1.21E-12	3377.963	5784.703	3377.963	5784.703
title_sent	545.1941	502.2443	1.085516	**0.278773**	-444.134	1534.523	-444.134	1534.523

[2] Daily Average & [b] Body Sentiment



SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.082523							
R Square	**0.00681**							
Adjusted R	0.002706							
Standard E	5652.136							
Observatic	244							
ANOVA								
	df	*SS*	*MS*	*F*	***gnificance F***			
Regressior	1	53009545	53009545	1.659315	**0.198926**			
Residual	242	7.73E+09	31946637					
Total	243	7.78E+09						
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4571.861	555.8287	8.225306	1.2E-14	3476.981	5666.741	3476.981	5666.741
body_sent	650.0615	504.6497	1.288144	**0.198926**	-344.005	1644.128	-344.005	1644.128

[2] Daily Average & [c] Bubble Implication



SUMMARY OUTPUT

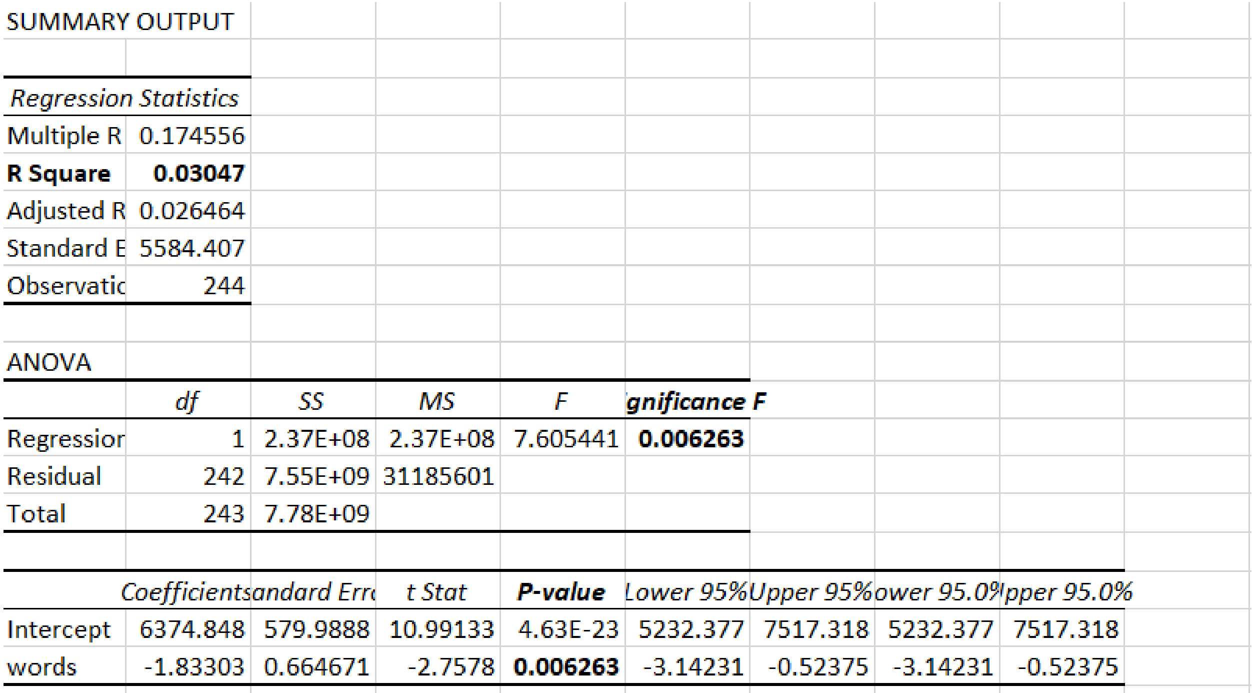
Regression Statistics	
Multiple R	0.190091
R Square	**0.036135**
Adjusted R	0.032152
Standard E	5568.069
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	2.81E+08	2.81E+08	9.072407	**0.00287**
Residual	242	7.5E+09	31003390		
Total	243	7.78E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4577.145	398.7377	11.47909	1.23E-24	3791.706	5362.585	3791.706	5362.585
bubble	2680.066	889.7831	3.012044	**0.00287**	927.3573	4432.774	927.3573	4432.774

[2] Daily Average & [d] Length In Words



SUMMARY OUTPUT

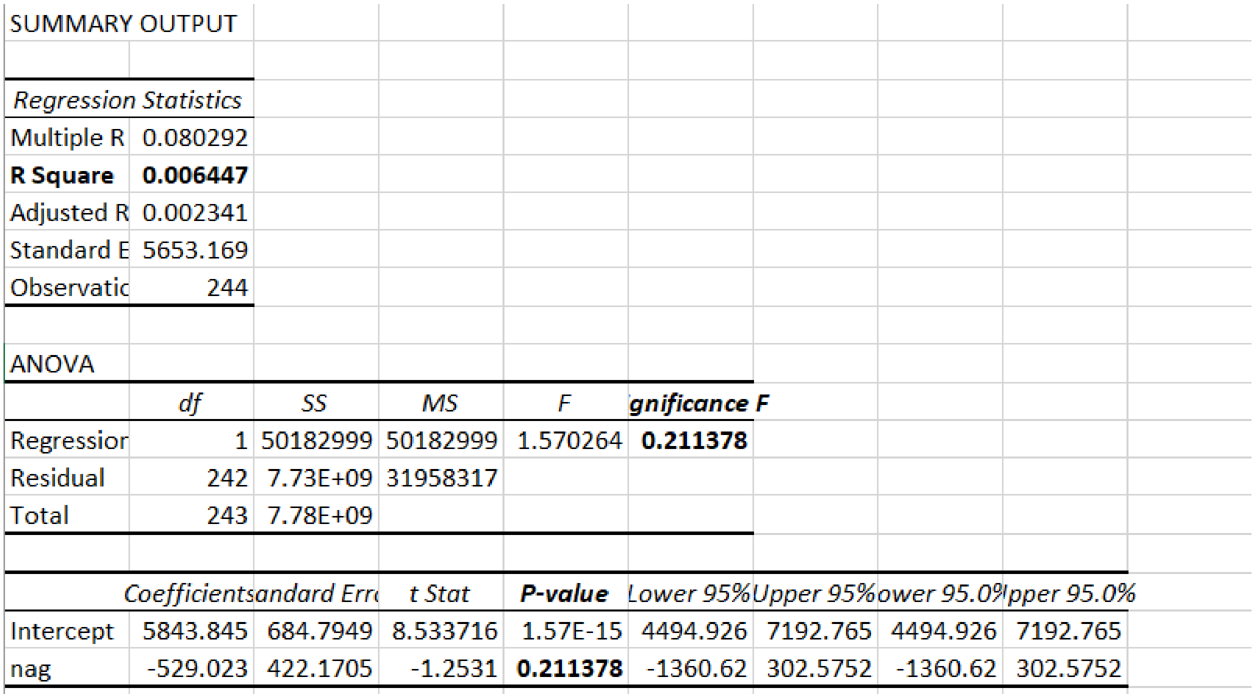
Regression Statistics	
Multiple R	0.174556
R Square	**0.03047**
Adjusted R	0.026464
Standard E	5584.407
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	2.37E+08	2.37E+08	7.605441	**0.006263**
Residual	242	7.55E+09	31185601		
Total	243	7.78E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	6374.848	579.9888	10.99133	4.63E-23	5232.377	7517.318	5232.377	7517.318
words	-1.83303	0.664671	-2.7578	**0.006263**	-3.14231	-0.52375	-3.14231	-0.52375

[2] Daily Average & [e] News Article Genre



SUMMARY OUTPUT

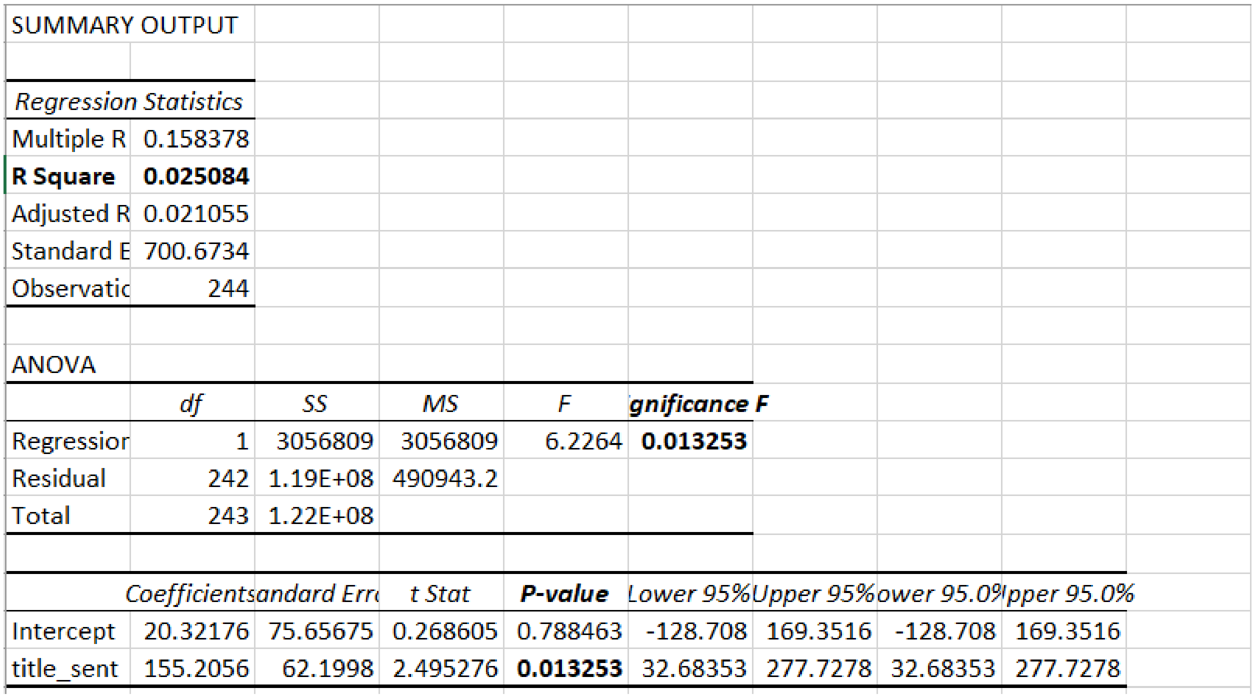
Regression Statistics	
Multiple R	0.080292
R Square	**0.006447**
Adjusted R	0.002341
Standard E	5653.169
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	50182999	50182999	1.570264	**0.211378**
Residual	242	7.73E+09	31958317		
Total	243	7.78E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	5843.845	684.7949	8.533716	1.57E-15	4494.926	7192.765	4494.926	7192.765
nag	-529.023	422.1705	-1.2531	**0.211378**	-1360.62	302.5752	-1360.62	302.5752

[3] Absolute Change & [a] Title Sentiment



SUMMARY OUTPUT

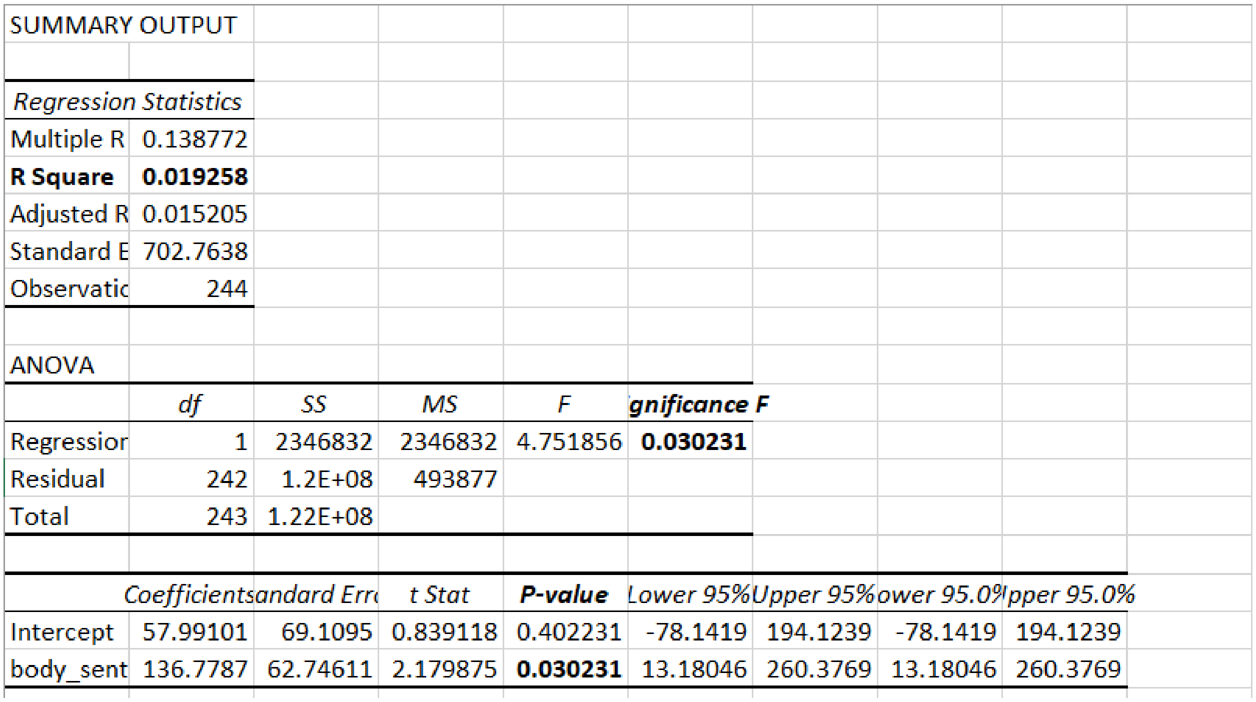
Regression Statistics	
Multiple R	0.158378
R Square	**0.025084**
Adjusted R	0.021055
Standard E	700.6734
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	3056809	3056809	6.2264	**0.013253**
Residual	242	1.19E+08	490943.2		
Total	243	1.22E+08			

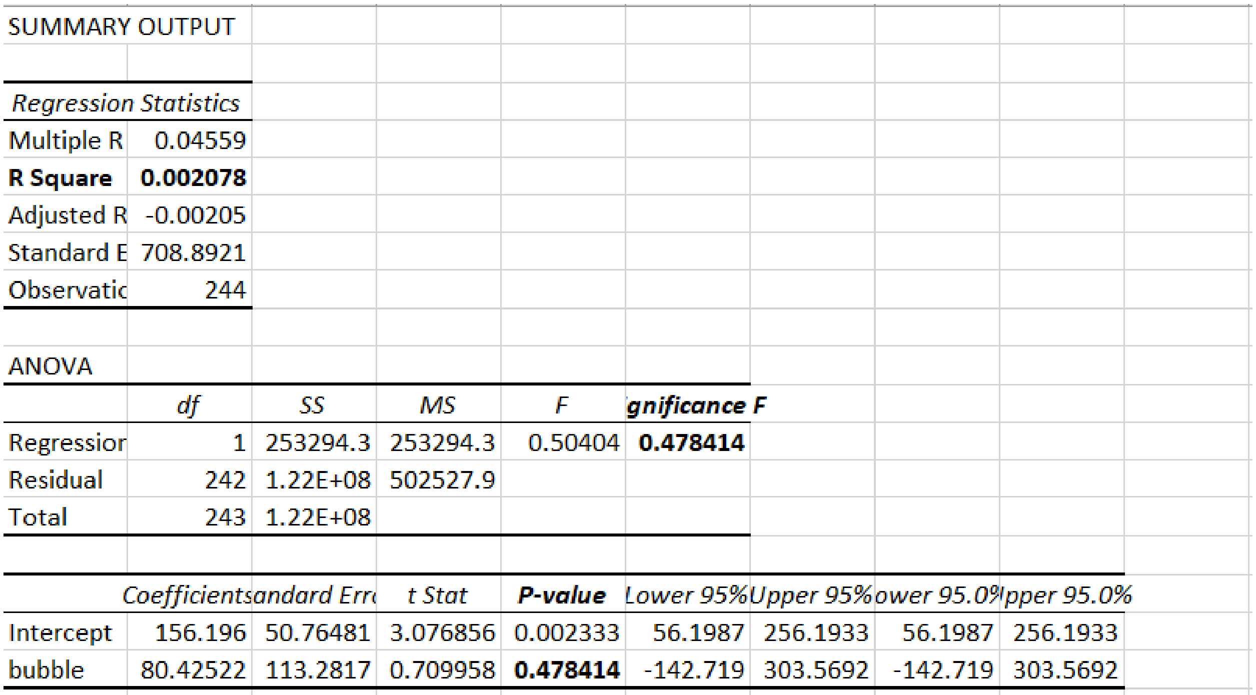
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	20.32176	75.65675	0.268605	0.788463	-128.708	169.3516	-128.708	169.3516
title_sent	155.2056	62.1998	2.495276	**0.013253**	32.68353	277.7278	32.68353	277.7278

[3] Absolute Change & [b] Body Sentiment



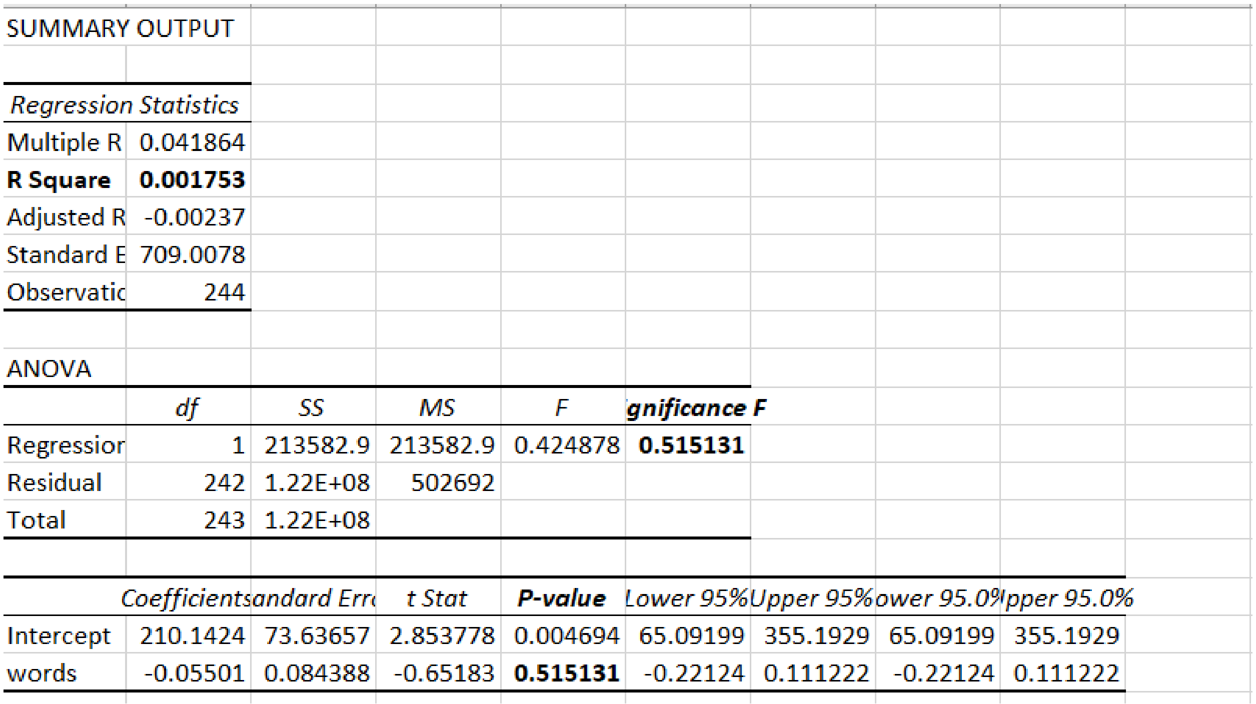
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.138772							
R Square	**0.019258**							
Adjusted R	0.015205							
Standard E	702.7638							
Observatic	244							
ANOVA								
	df	*SS*	*MS*	*F*	***gnificance F***			
Regressior	1	2346832	2346832	4.751856	**0.030231**			
Residual	242	1.2E+08	493877					
Total	243	1.22E+08						
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	57.99101	69.1095	0.839118	0.402231	-78.1419	194.1239	-78.1419	194.1239
body_sent	136.7787	62.74611	2.179875	**0.030231**	13.18046	260.3769	13.18046	260.3769

[3] Absolute Change & [c] Bubble Implication



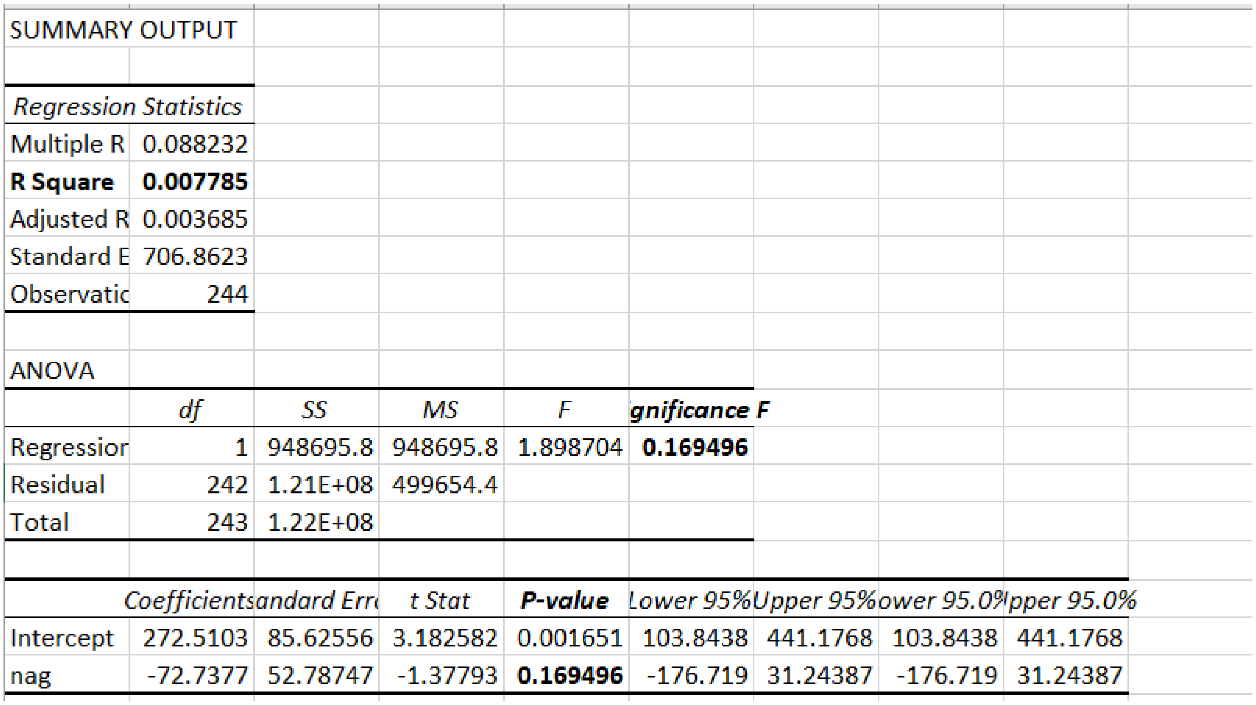
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.04559							
R Square	**0.002078**							
Adjusted R	-0.00205							
Standard E	708.8921							
Observatic	244							
ANOVA								
	df	*SS*	*MS*	*F*	***gnificance F***			
Regressior	1	253294.3	253294.3	0.50404	**0.478414**			
Residual	242	1.22E+08	502527.9					
Total	243	1.22E+08						
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	156.196	50.76481	3.076856	0.002333	56.1987	256.1933	56.1987	256.1933
bubble	80.42522	113.2817	0.709958	**0.478414**	-142.719	303.5692	-142.719	303.5692

[3] Absolute Change & [d] Length In Words



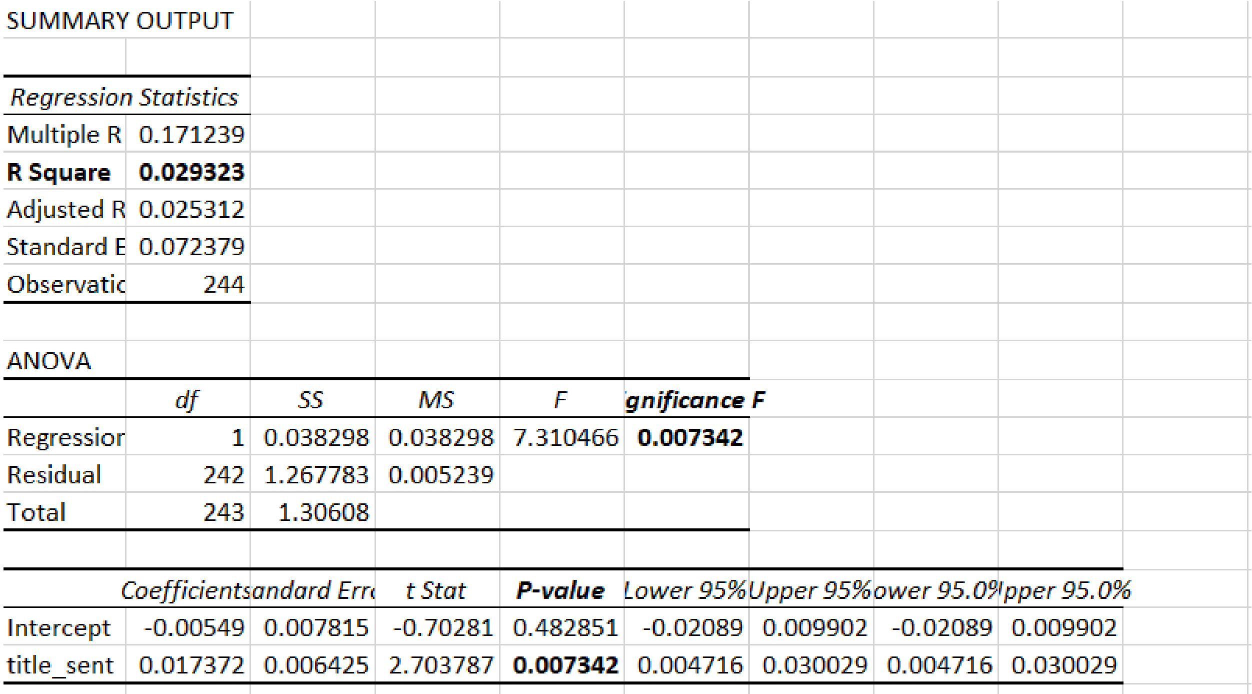
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.041864							
R Square	**0.001753**							
Adjusted R	-0.00237							
Standard E	709.0078							
Observatic	244							
ANOVA								
	df	*SS*	*MS*	*F*	***gnificance F***			
Regressior	1	213582.9	213582.9	0.424878	**0.515131**			
Residual	242	1.22E+08	502692					
Total	243	1.22E+08						
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	210.1424	73.63657	2.853778	0.004694	65.09199	355.1929	65.09199	355.1929
words	-0.05501	0.084388	-0.65183	**0.515131**	-0.22124	0.111222	-0.22124	0.111222

[3] Absolute Change & [e] News Article Genre



SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.088232							
R Square	**0.007785**							
Adjusted R	0.003685							
Standard E	706.8623							
Observatic	244							
ANOVA								
	df	*SS*	*MS*	*F*	***gnificance F***			
Regressior	1	948695.8	948695.8	1.898704	**0.169496**			
Residual	242	1.21E+08	499654.4					
Total	243	1.22E+08						
	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	272.5103	85.62556	3.182582	0.001651	103.8438	441.1768	103.8438	441.1768
nag	-72.7377	52.78747	-1.37793	**0.169496**	-176.719	31.24387	-176.719	31.24387

[4] Percentage Change & [a] Title Sentiment



SUMMARY OUTPUT

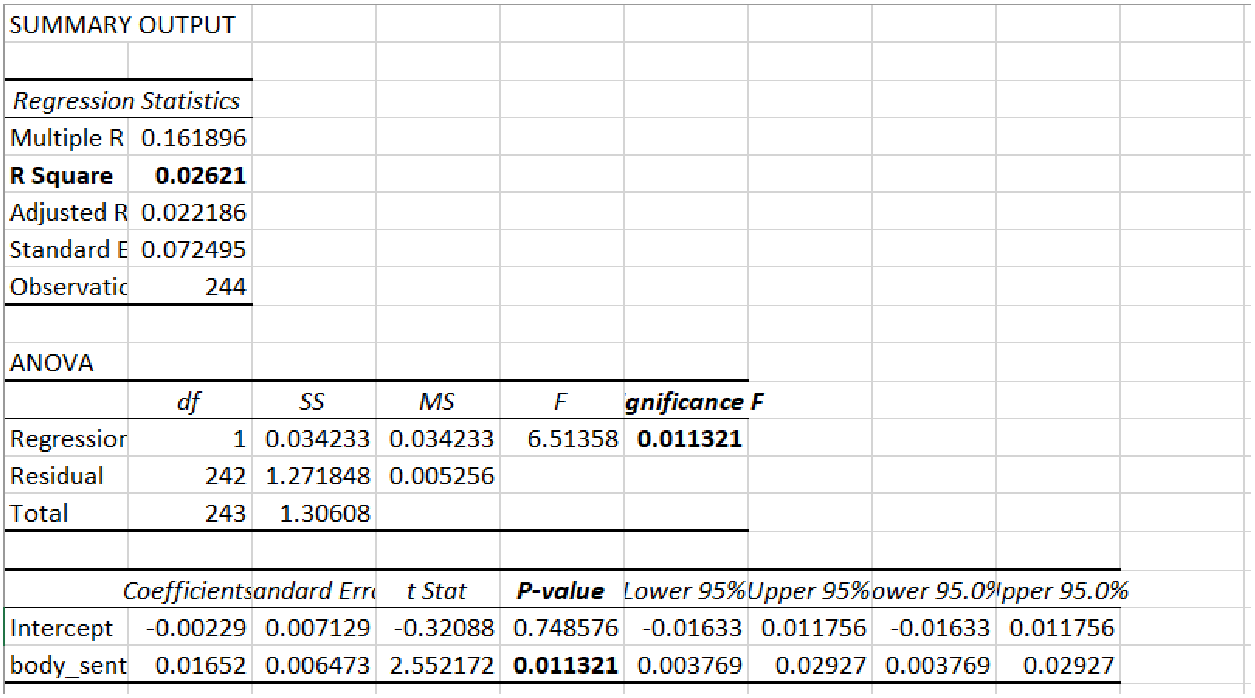
Regression Statistics	
Multiple R	0.171239
R Square	**0.029323**
Adjusted R	0.025312
Standard E	0.072379
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.038298	0.038298	7.310466	**0.007342**
Residual	242	1.267783	0.005239		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-0.00549	0.007815	-0.70281	0.482851	-0.02089	0.009902	-0.02089	0.009902
title_sent	0.017372	0.006425	2.703787	**0.007342**	0.004716	0.030029	0.004716	0.030029

[4] Percentage Change & [b] Body Sentiment



SUMMARY OUTPUT

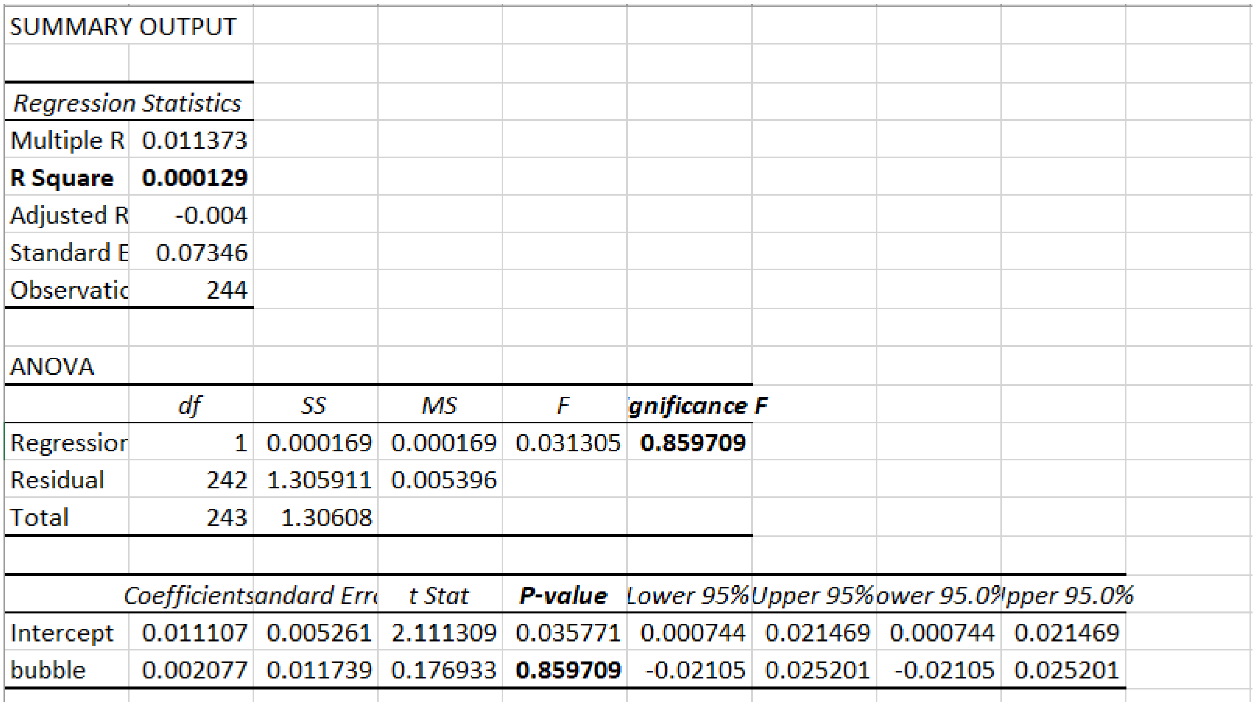
Regression Statistics	
Multiple R	0.161896
R Square	**0.02621**
Adjusted R	0.022186
Standard E	0.072495
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.034233	0.034233	6.51358	**0.011321**
Residual	242	1.271848	0.005256		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-0.00229	0.007129	-0.32088	0.748576	-0.01633	0.011756	-0.01633	0.011756
body_sent	0.01652	0.006473	2.552172	**0.011321**	0.003769	0.02927	0.003769	0.02927

[4] Percentage Change & [c] Bubble Implication



SUMMARY OUTPUT

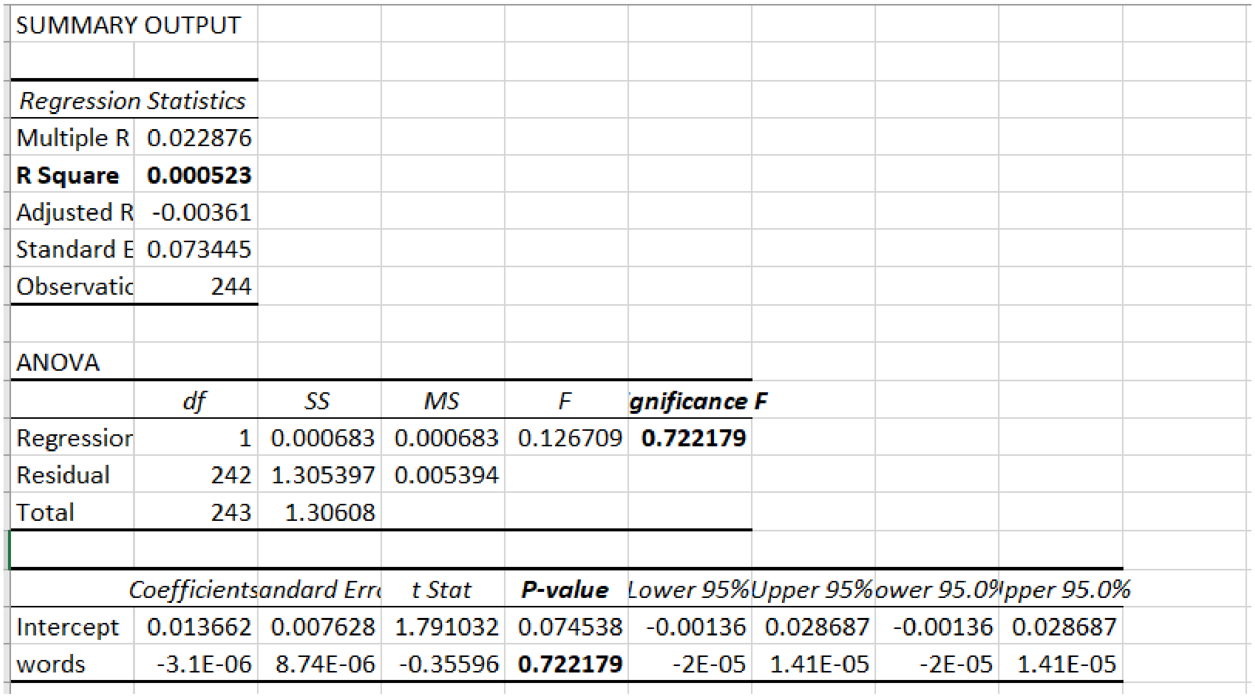
Regression Statistics	
Multiple R	0.011373
R Square	**0.000129**
Adjusted R	-0.004
Standard E	0.07346
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.000169	0.000169	0.031305	**0.859709**
Residual	242	1.305911	0.005396		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.011107	0.005261	2.111309	0.035771	0.000744	0.021469	0.000744	0.021469
bubble	0.002077	0.011739	0.176933	**0.859709**	-0.02105	0.025201	-0.02105	0.025201

[4] Percentage Change & [d] Length In Words



SUMMARY OUTPUT

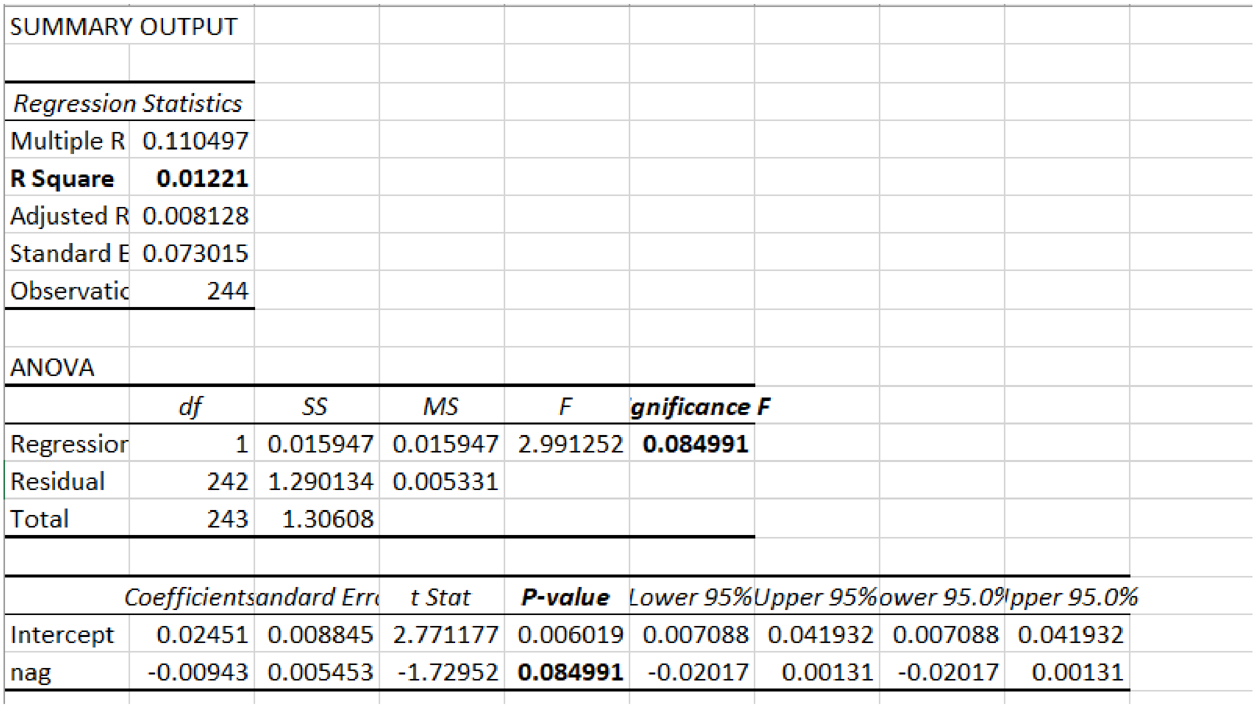
Regression Statistics	
Multiple R	0.022876
R Square	**0.000523**
Adjusted R	-0.00361
Standard E	0.073445
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.000683	0.000683	0.126709	**0.722179**
Residual	242	1.305397	0.005394		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.013662	0.007628	1.791032	0.074538	-0.00136	0.028687	-0.00136	0.028687
words	-3.1E-06	8.74E-06	-0.35596	**0.722179**	-2E-05	1.41E-05	-2E-05	1.41E-05

[4] Percentage Change & [e] News Article Genre



SUMMARY OUTPUT

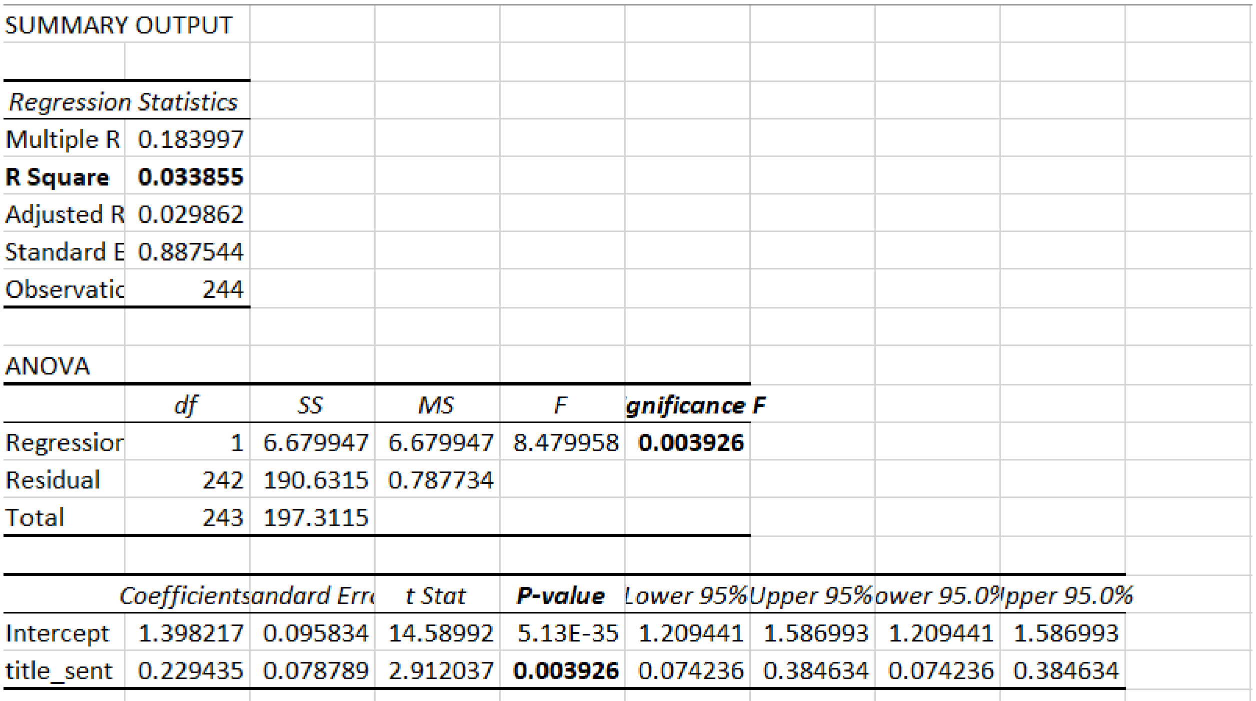
Regression Statistics	
Multiple R	0.110497
R Square	**0.01221**
Adjusted R	0.008128
Standard E	0.073015
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.015947	0.015947	2.991252	**0.084991**
Residual	242	1.290134	0.005331		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.02451	0.008845	2.771177	0.006019	0.007088	0.041932	0.007088	0.041932
nag	-0.00943	0.005453	-1.72952	**0.084991**	-0.02017	0.00131	-0.02017	0.00131

[5] Change Tag & [a] Title Sentiment



SUMMARY OUTPUT

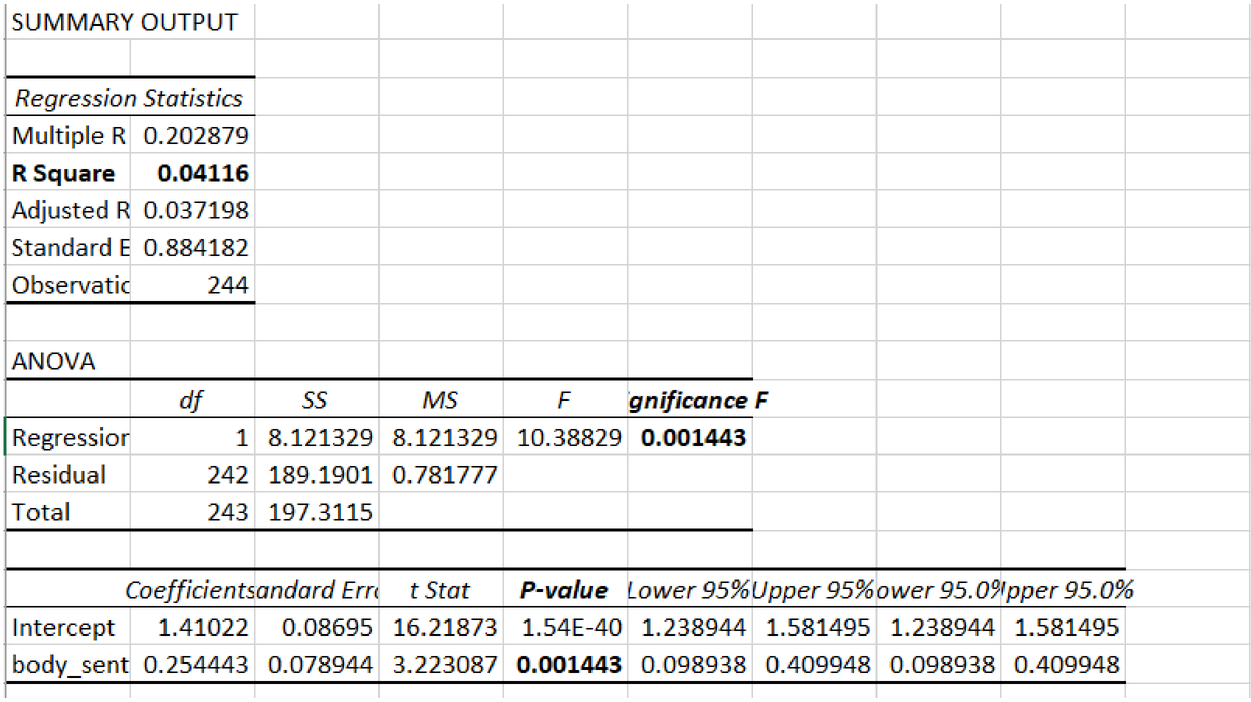
Regression Statistics	
Multiple R	0.183997
R Square	**0.033855**
Adjusted R	0.029862
Standard E	0.887544
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	6.679947	6.679947	8.479958	**0.003926**
Residual	242	190.6315	0.787734		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.398217	0.095834	14.58992	5.13E-35	1.209441	1.586993	1.209441	1.586993
title_sent	0.229435	0.078789	2.912037	**0.003926**	0.074236	0.384634	0.074236	0.384634

[5] Change Tag & [b] Body Sentiment



SUMMARY OUTPUT

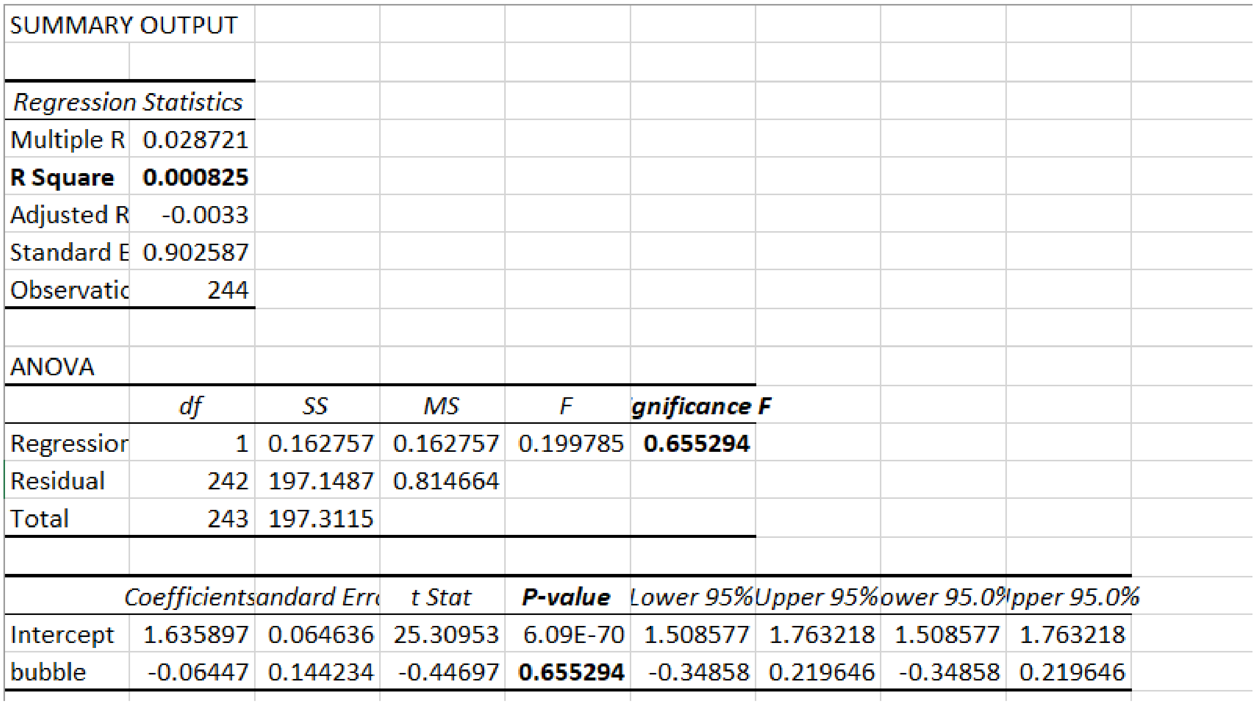
Regression Statistics	
Multiple R	0.202879
R Square	**0.04116**
Adjusted R	0.037198
Standard E	0.884182
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	8.121329	8.121329	10.38829	**0.001443**
Residual	242	189.1901	0.781777		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.41022	0.08695	16.21873	1.54E-40	1.238944	1.581495	1.238944	1.581495
body_sent	0.254443	0.078944	3.223087	**0.001443**	0.098938	0.409948	0.098938	0.409948

[5] Change Tag & [c] Bubble Implication



SUMMARY OUTPUT

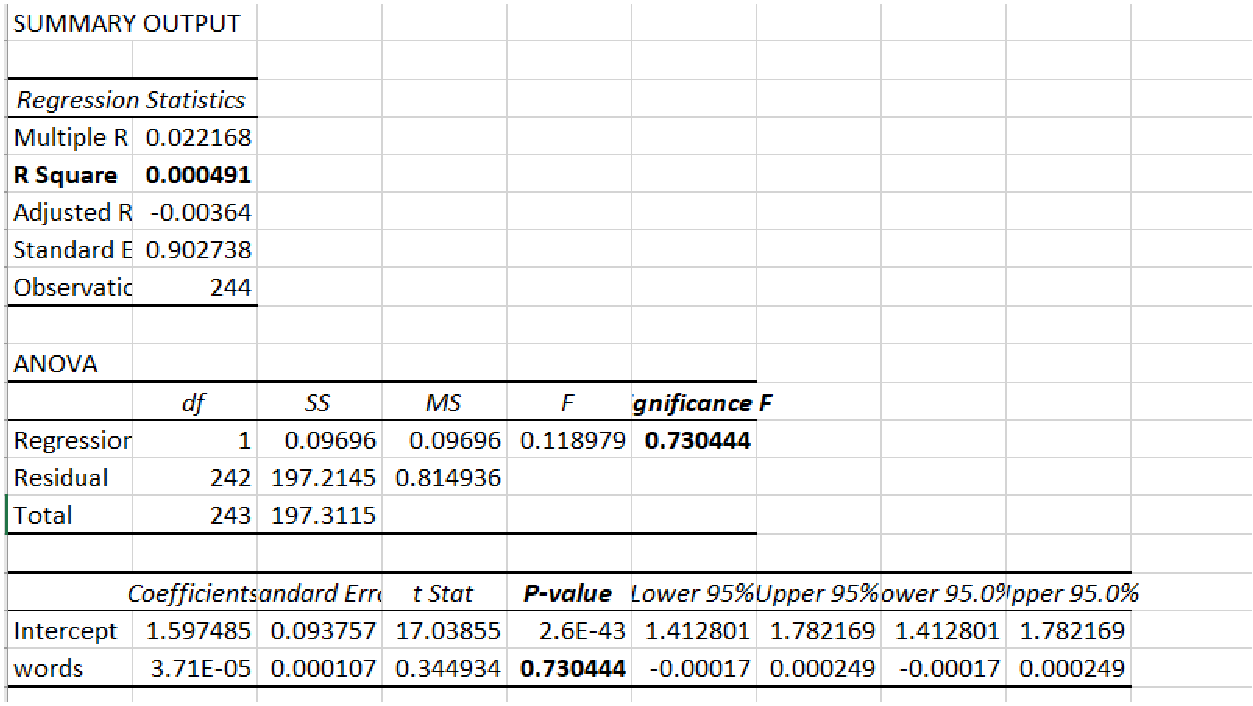
Regression Statistics	
Multiple R	0.028721
R Square	**0.000825**
Adjusted R	-0.0033
Standard E	0.902587
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.162757	0.162757	0.199785	**0.655294**
Residual	242	197.1487	0.814664		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.635897	0.064636	25.30953	6.09E-70	1.508577	1.763218	1.508577	1.763218
bubble	-0.06447	0.144234	-0.44697	**0.655294**	-0.34858	0.219646	-0.34858	0.219646

[5] Change Tag & [d] Length In Words



SUMMARY OUTPUT

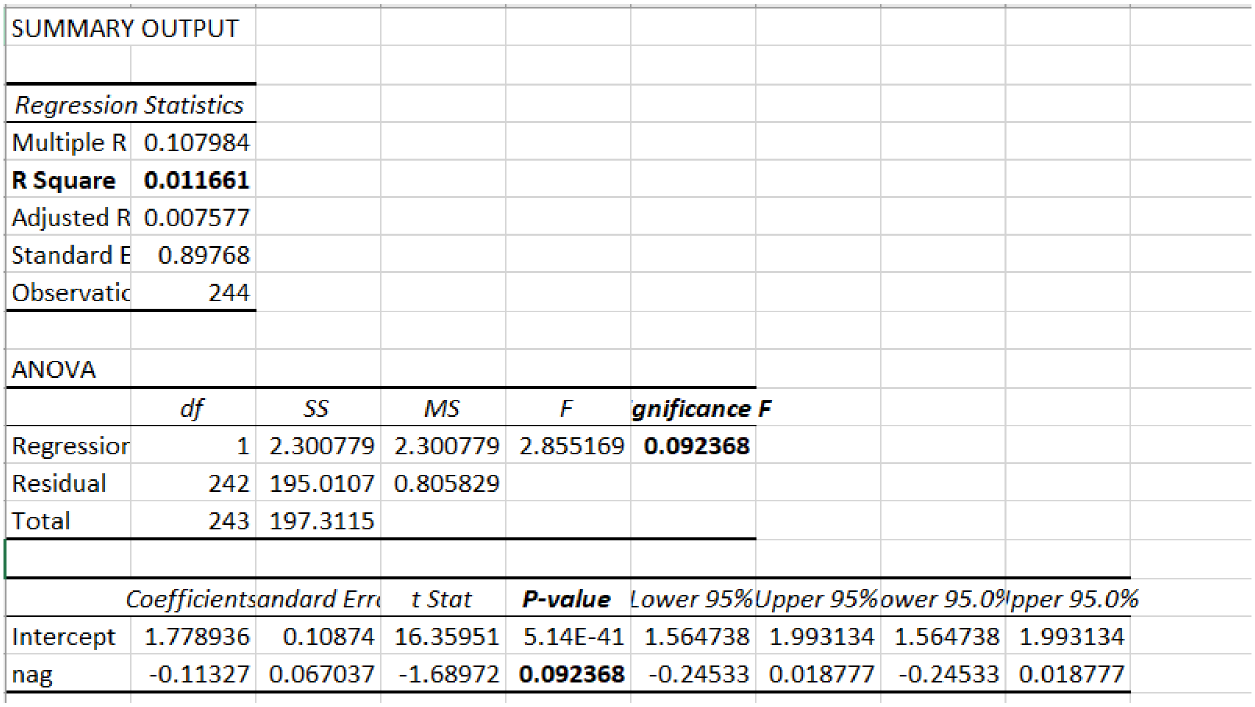
Regression Statistics	
Multiple R	0.022168
R Square	**0.000491**
Adjusted R	-0.00364
Standard E	0.902738
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	0.09696	0.09696	0.118979	**0.730444**
Residual	242	197.2145	0.814936		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.597485	0.093757	17.03855	2.6E-43	1.412801	1.782169	1.412801	1.782169
words	3.71E-05	0.000107	0.344934	**0.730444**	-0.00017	0.000249	-0.00017	0.000249

[5] Change Tag & [e] News Article Genre



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.107984
R Square	**0.011661**
Adjusted R	0.007577
Standard E	0.89768
Observatic	244

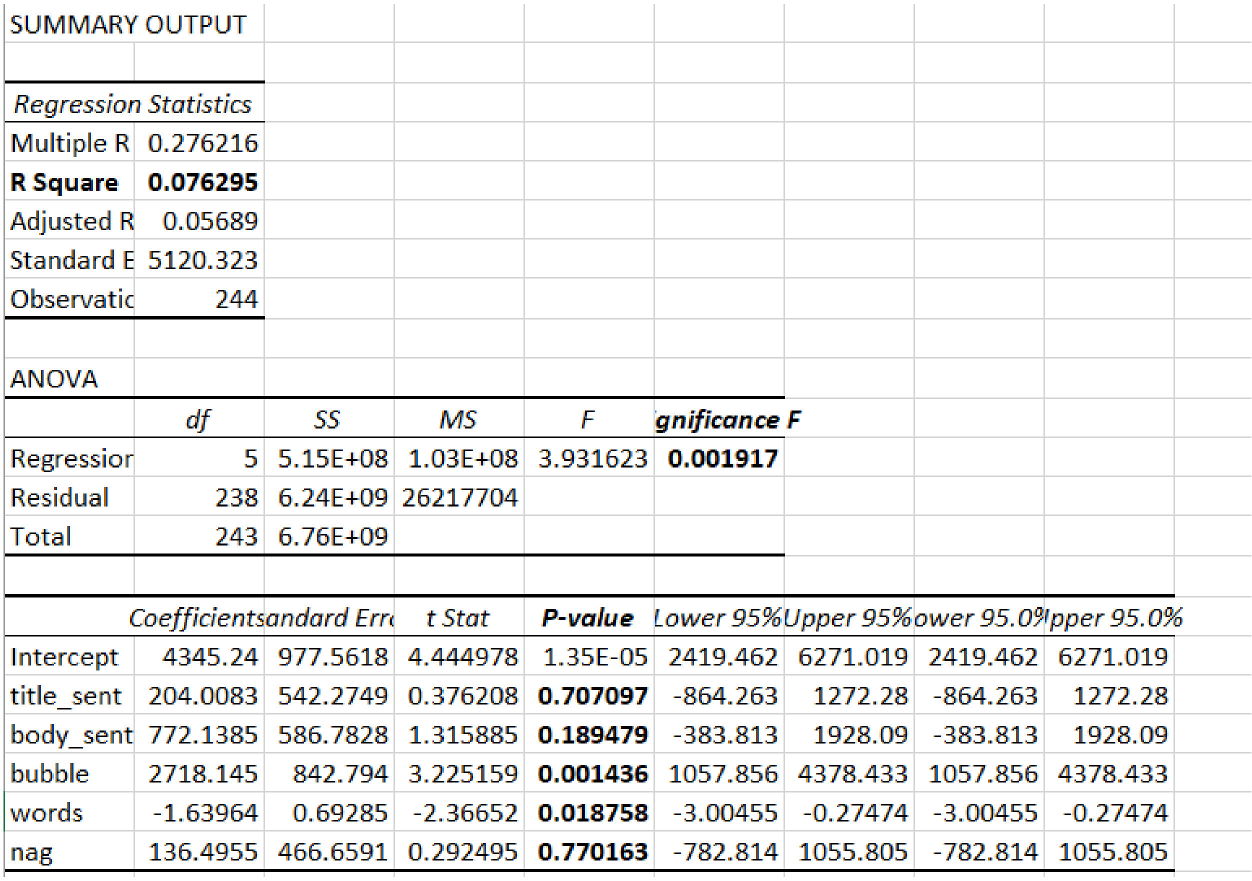
ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	1	2.300779	2.300779	2.855169	**0.092368**
Residual	242	195.0107	0.805829		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.778936	0.10874	16.35951	5.14E-41	1.564738	1.993134	1.564738	1.993134
nag	-0.11327	0.067037	-1.68972	**0.092368**	-0.24533	0.018777	-0.24533	0.018777

MULTIPLE REGRESSION PLOTS

[1] Daily High & [I] All Independent Variables



SUMMARY OUTPUT

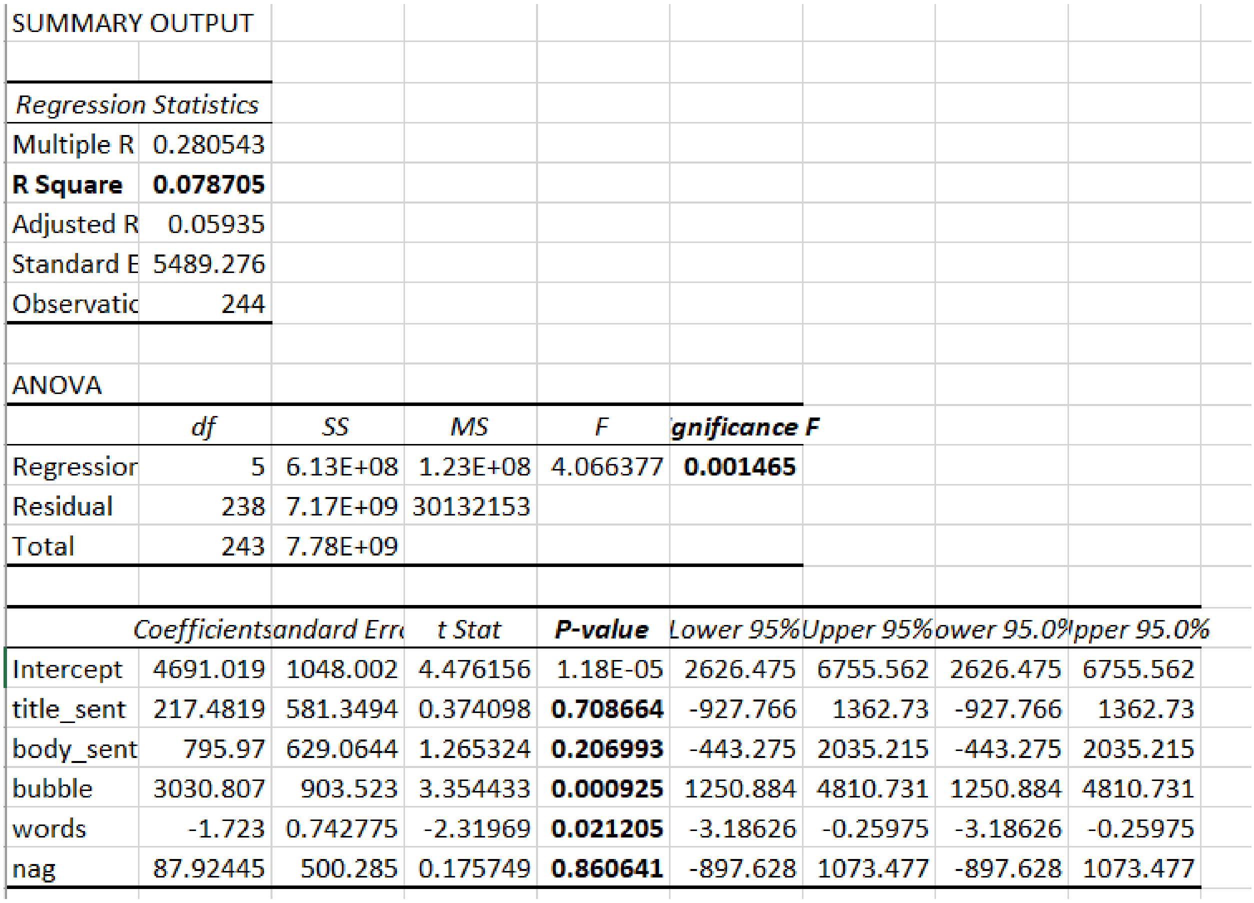
Regression Statistics	
Multiple R	0.276216
R Square	**0.076295**
Adjusted R	0.05689
Standard E	5120.323
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	5	5.15E+08	1.03E+08	3.931623	**0.001917**
Residual	238	6.24E+09	26217704		
Total	243	6.76E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4345.24	977.5618	4.444978	1.35E-05	2419.462	6271.019	2419.462	6271.019
title_sent	204.0083	542.2749	0.376208	**0.707097**	-864.263	1272.28	-864.263	1272.28
body_sent	772.1385	586.7828	1.315885	**0.189479**	-383.813	1928.09	-383.813	1928.09
bubble	2718.145	842.794	3.225159	**0.001436**	1057.856	4378.433	1057.856	4378.433
words	-1.63964	0.69285	-2.36652	**0.018758**	-3.00455	-0.27474	-3.00455	-0.27474
nag	136.4955	466.6591	0.292495	**0.770163**	-782.814	1055.805	-782.814	1055.805

[2] Daily Average & [I] All Independent Variables



SUMMARY OUTPUT

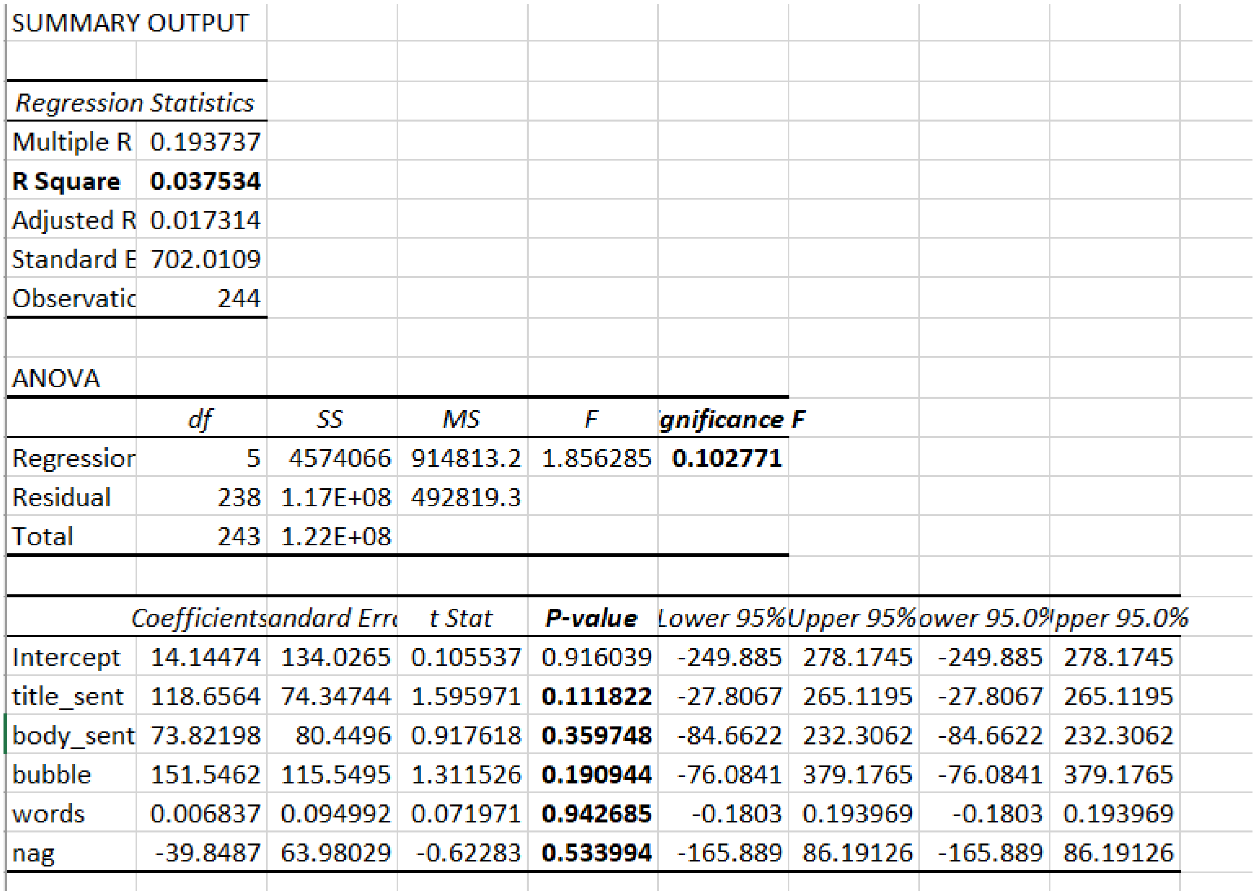
Regression Statistics	
Multiple R	0.280543
R Square	**0.078705**
Adjusted R	0.05935
Standard E	5489.276
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	5	6.13E+08	1.23E+08	4.066377	**0.001465**
Residual	238	7.17E+09	30132153		
Total	243	7.78E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4691.019	1048.002	4.476156	1.18E-05	2626.475	6755.562	2626.475	6755.562
title_sent	217.4819	581.3494	0.374098	**0.708664**	-927.766	1362.73	-927.766	1362.73
body_sent	795.97	629.0644	1.265324	**0.206993**	-443.275	2035.215	-443.275	2035.215
bubble	3030.807	903.523	3.354433	**0.000925**	1250.884	4810.731	1250.884	4810.731
words	-1.723	0.742775	-2.31969	**0.021205**	-3.18626	-0.25975	-3.18626	-0.25975
nag	87.92445	500.285	0.175749	**0.860641**	-897.628	1073.477	-897.628	1073.477

[3] Absolute Change & [I] All Independent Variables



SUMMARY OUTPUT

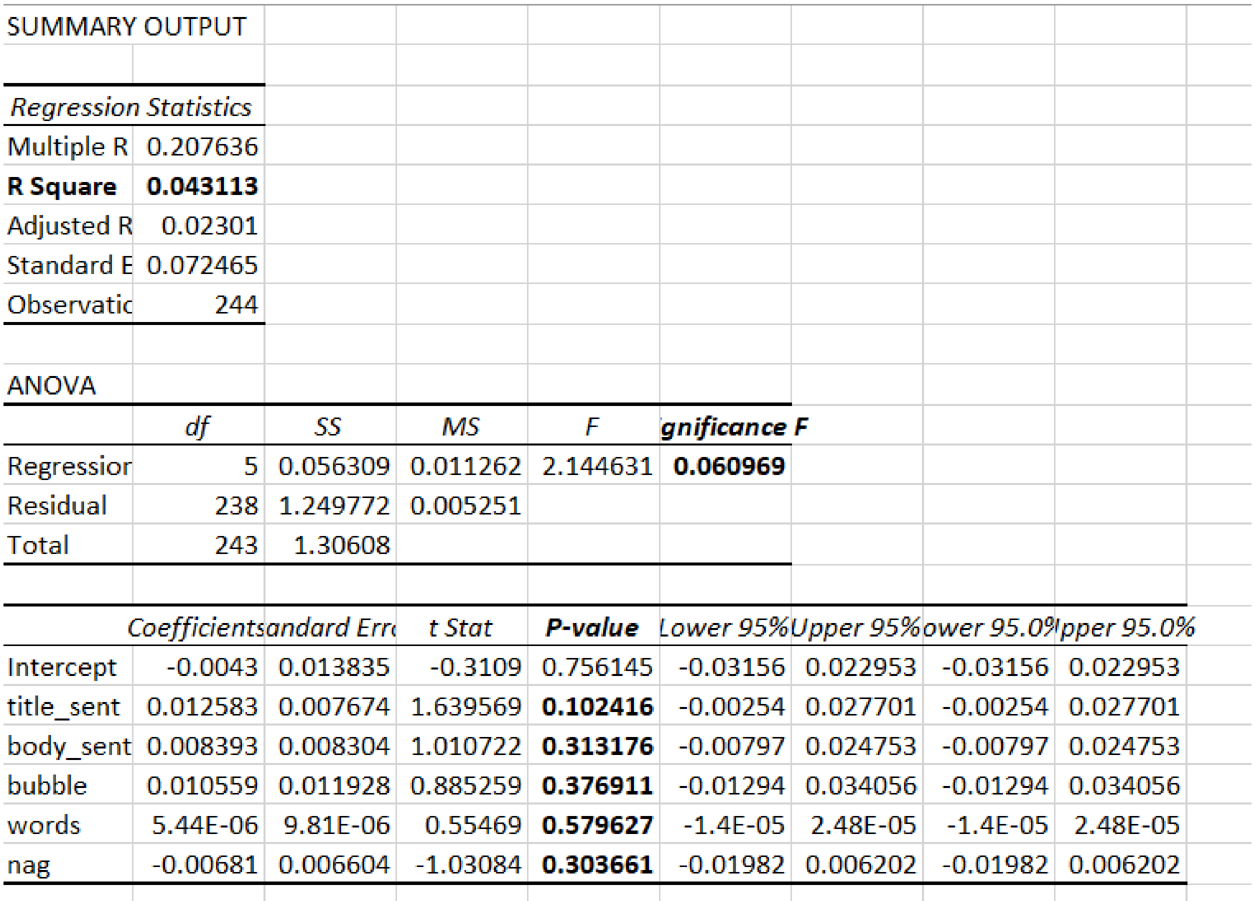
Regression Statistics	
Multiple R	0.193737
R Square	**0.037534**
Adjusted R	0.017314
Standard E	702.0109
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	5	4574066	914813.2	1.856285	**0.102771**
Residual	238	1.17E+08	492819.3		
Total	243	1.22E+08			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	14.14474	134.0265	0.105537	0.916039	-249.885	278.1745	-249.885	278.1745
title_sent	118.6564	74.34744	1.595971	**0.111822**	-27.8067	265.1195	-27.8067	265.1195
body_sent	73.82198	80.4496	0.917618	**0.359748**	-84.6622	232.3062	-84.6622	232.3062
bubble	151.5462	115.5495	1.311526	**0.190944**	-76.0841	379.1765	-76.0841	379.1765
words	0.006837	0.094992	0.071971	**0.942685**	-0.1803	0.193969	-0.1803	0.193969
nag	-39.8487	63.98029	-0.62283	**0.533994**	-165.889	86.19126	-165.889	86.19126

[4] Percentage Change & [I] All Independent Variables



SUMMARY OUTPUT

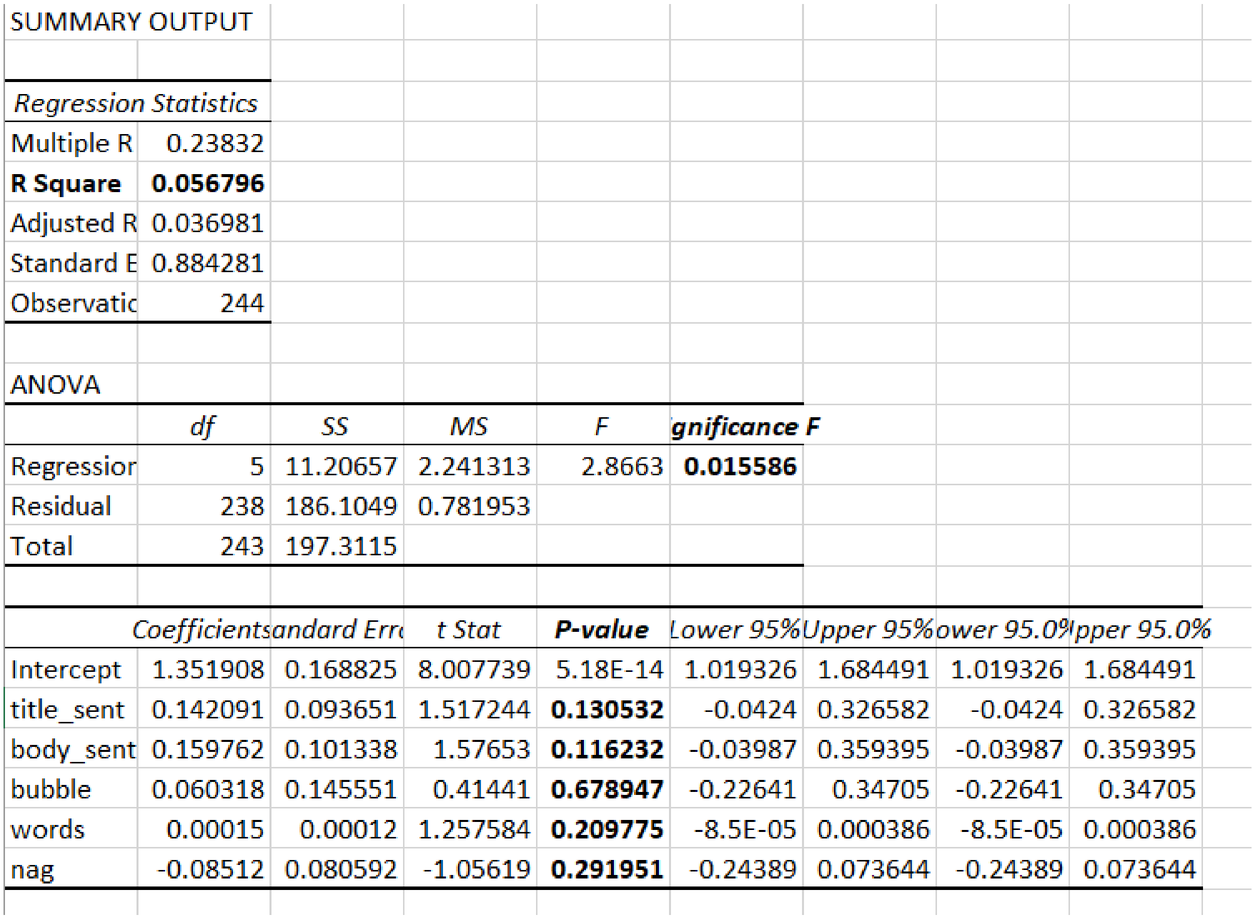
Regression Statistics	
Multiple R	0.207636
R Square	**0.043113**
Adjusted R	0.02301
Standard E	0.072465
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	5	0.056309	0.011262	2.144631	**0.060969**
Residual	238	1.249772	0.005251		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-0.0043	0.013835	-0.3109	0.756145	-0.03156	0.022953	-0.03156	0.022953
title_sent	0.012583	0.007674	1.639569	**0.102416**	-0.00254	0.027701	-0.00254	0.027701
body_sent	0.008393	0.008304	1.010722	**0.313176**	-0.00797	0.024753	-0.00797	0.024753
bubble	0.010559	0.011928	0.885259	**0.376911**	-0.01294	0.034056	-0.01294	0.034056
words	5.44E-06	9.81E-06	0.55469	**0.579627**	-1.4E-05	2.48E-05	-1.4E-05	2.48E-05
nag	-0.00681	0.006604	-1.03084	**0.303661**	-0.01982	0.006202	-0.01982	0.006202

[5] Change Tag & [I] All Independent Variables



SUMMARY OUTPUT

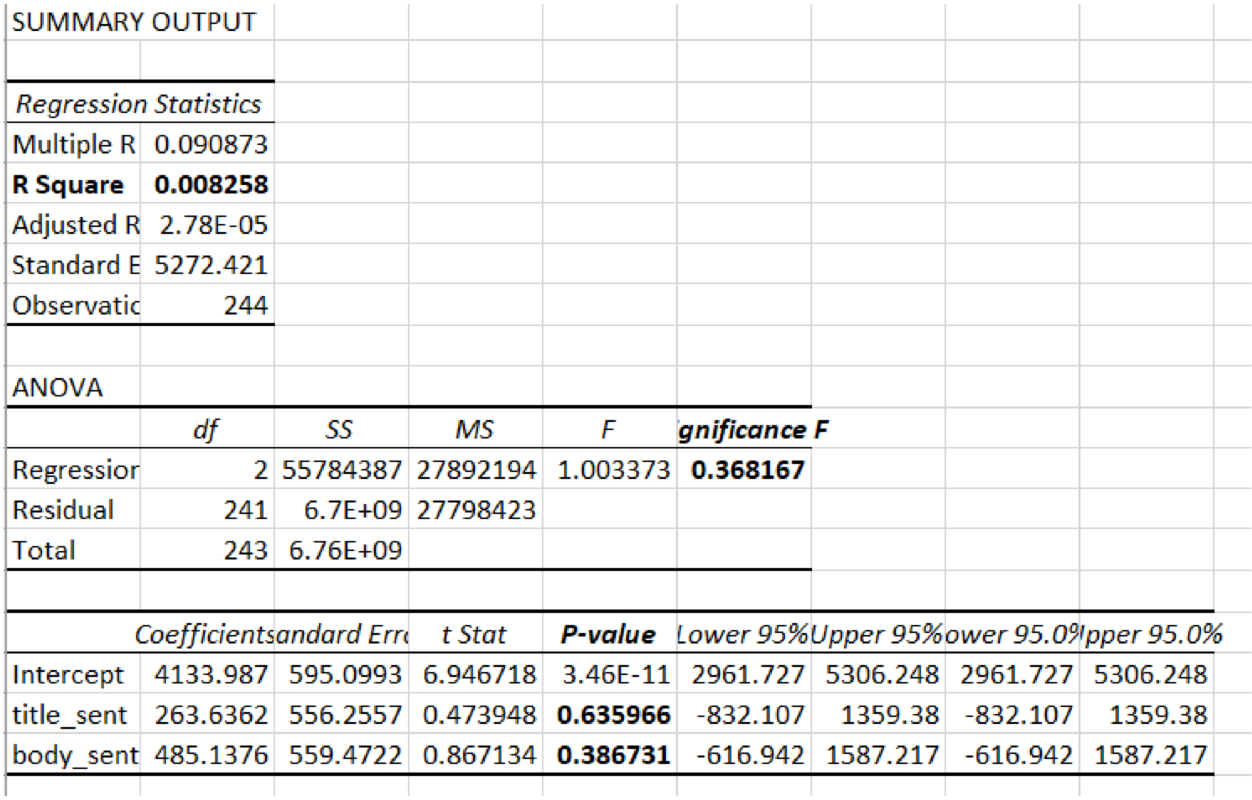
Regression Statistics	
Multiple R	0.23832
R Square	**0.056796**
Adjusted R	0.036981
Standard E	0.884281
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	5	11.20657	2.241313	2.8663	**0.015586**
Residual	238	186.1049	0.781953		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.351908	0.168825	8.007739	5.18E-14	1.019326	1.684491	1.019326	1.684491
title_sent	0.142091	0.093651	1.517244	**0.130532**	-0.0424	0.326582	-0.0424	0.326582
body_sent	0.159762	0.101338	1.57653	**0.116232**	-0.03987	0.359395	-0.03987	0.359395
bubble	0.060318	0.145551	0.41441	**0.678947**	-0.22641	0.34705	-0.22641	0.34705
words	0.00015	0.00012	1.257584	**0.209775**	-8.5E-05	0.000386	-8.5E-05	0.000386
nag	-0.08512	0.080592	-1.05619	**0.291951**	-0.24389	0.073644	-0.24389	0.073644

[1] Daily High & [II] Sentiment Loaded Variables



SUMMARY OUTPUT

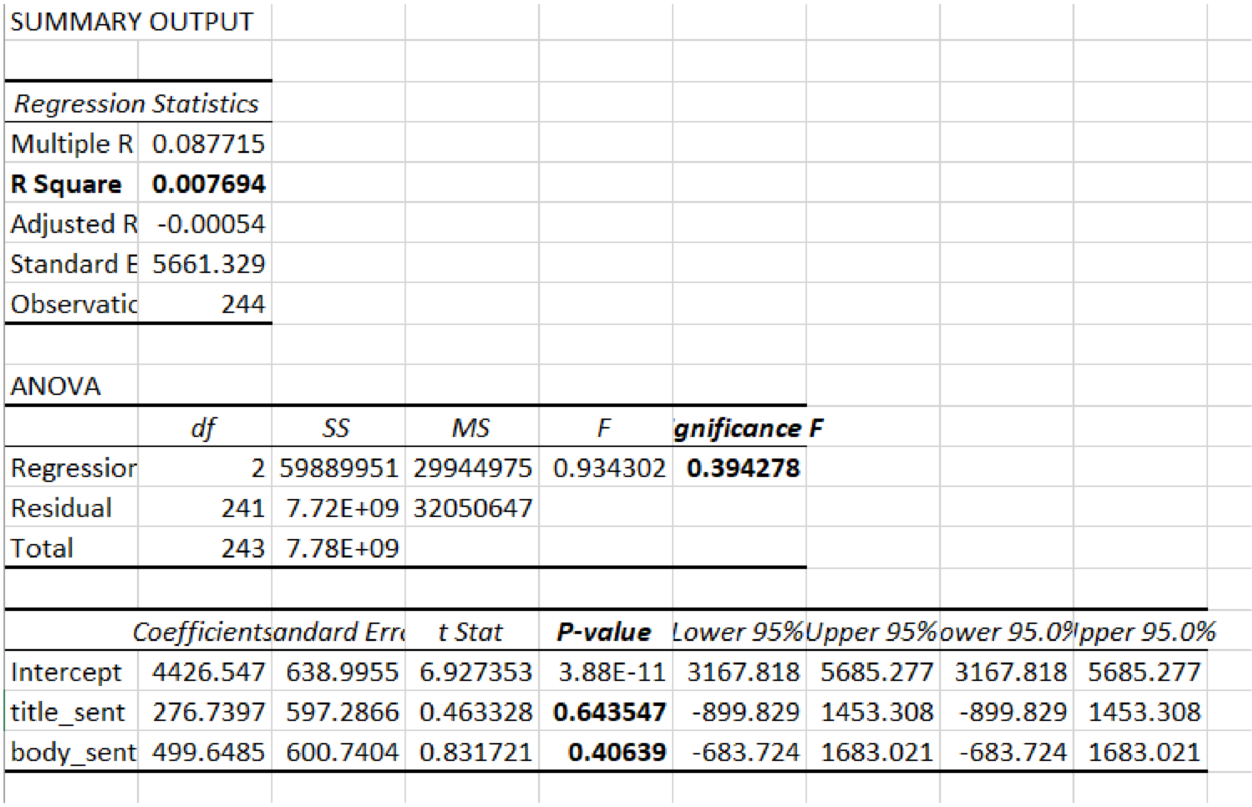
Regression Statistics	
Multiple R	0.090873
R Square	**0.008258**
Adjusted R	2.78E-05
Standard E	5272.421
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	2	55784387	27892194	1.003373	**0.368167**
Residual	241	6.7E+09	27798423		
Total	243	6.76E+09			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4133.987	595.0993	6.946718	3.46E-11	2961.727	5306.248	2961.727	5306.248
title_sent	263.6362	556.2557	0.473948	**0.635966**	-832.107	1359.38	-832.107	1359.38
body_sent	485.1376	559.4722	0.867134	**0.386731**	-616.942	1587.217	-616.942	1587.217

[2] Daily Average & [II] Sentiment Loaded Variables



SUMMARY OUTPUT

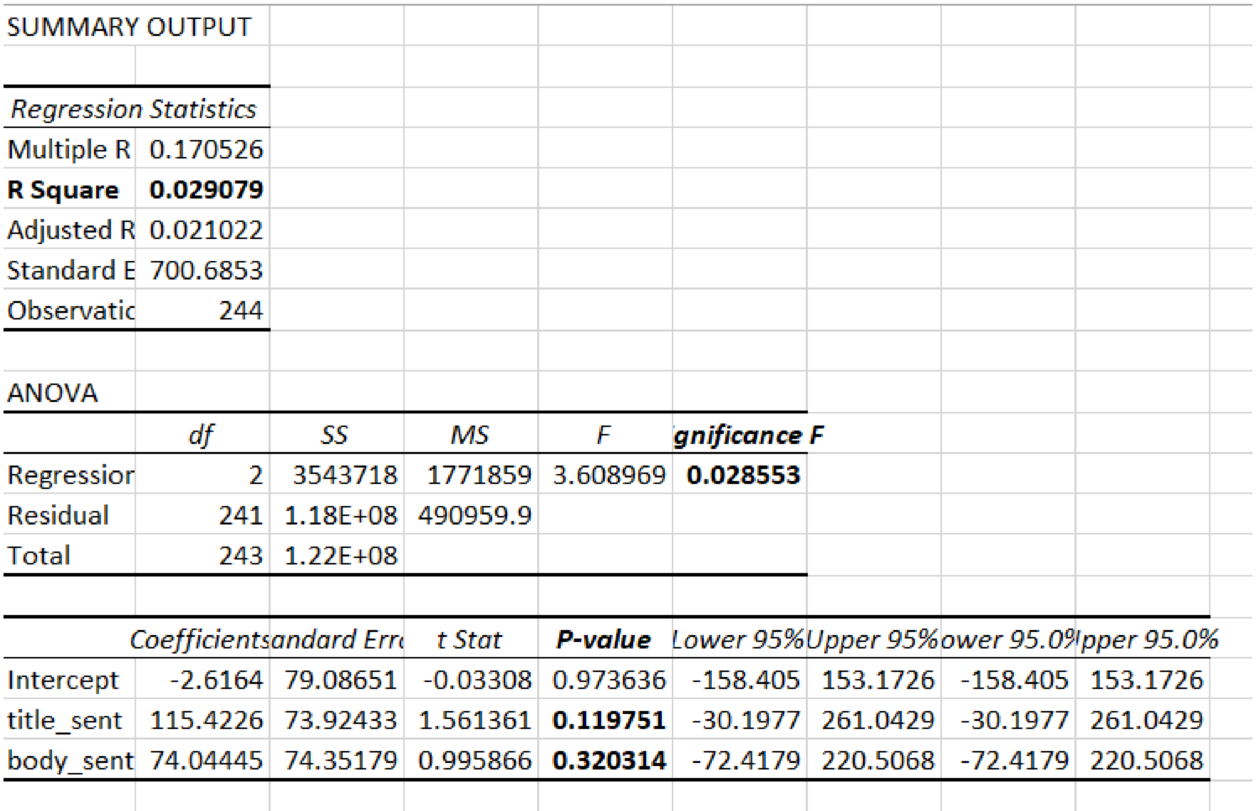
Regression Statistics	
Multiple R	0.087715
R Square	**0.007694**
Adjusted R	-0.00054
Standard E	5661.329
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	2	59889951	29944975	0.934302	**0.394278**
Residual	241	7.72E+09	32050647		
Total	243	7.78E+09			

	Coefficients	*andard Erro*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	4426.547	638.9955	6.927353	3.88E-11	3167.818	5685.277	3167.818	5685.277
title_sent	276.7397	597.2866	0.463328	**0.643547**	-899.829	1453.308	-899.829	1453.308
body_sent	499.6485	600.7404	0.831721	**0.40639**	-683.724	1683.021	-683.724	1683.021

[3] Absolute Change & [II] Sentiment Loaded Variables



SUMMARY OUTPUT

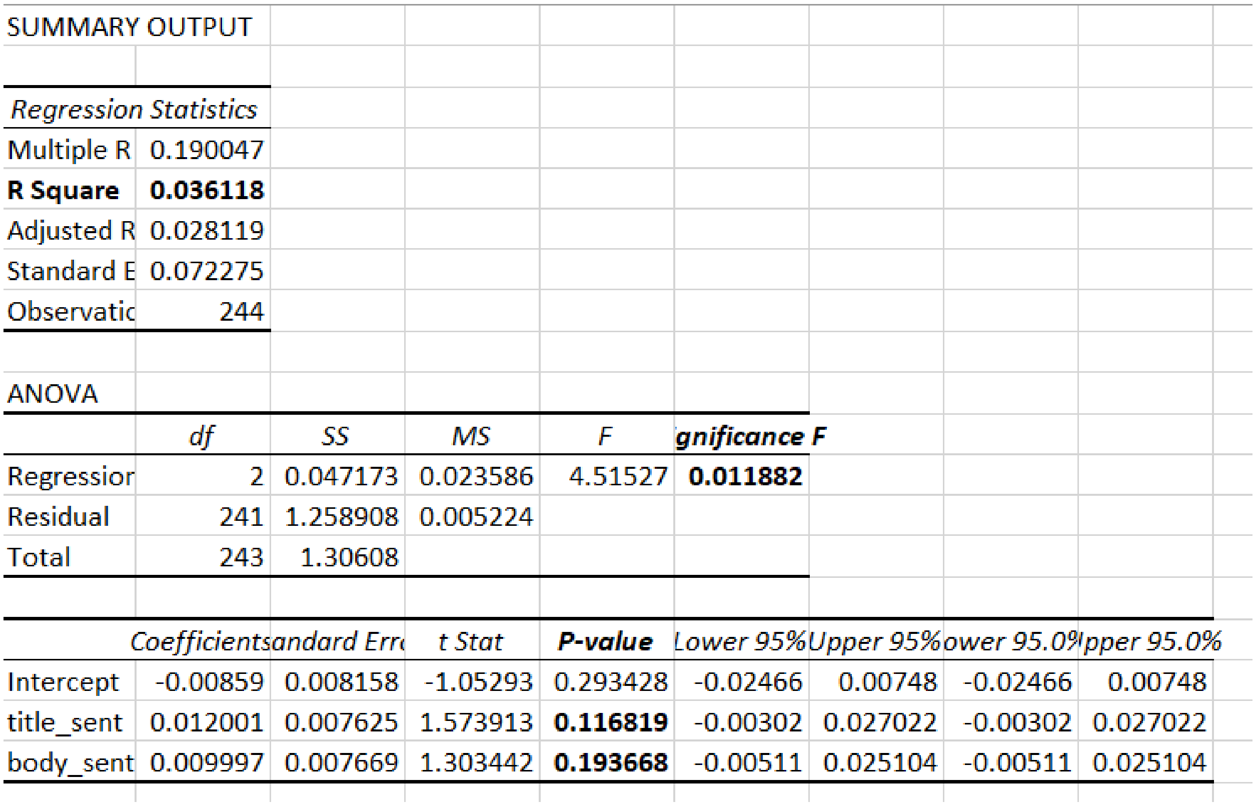
Regression Statistics	
Multiple R	0.170526
R Square	**0.029079**
Adjusted R	0.021022
Standard E	700.6853
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	2	3543718	1771859	3.608969	**0.028553**
Residual	241	1.18E+08	490959.9		
Total	243	1.22E+08			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-2.6164	79.08651	-0.03308	0.973636	-158.405	153.1726	-158.405	153.1726
title_sent	115.4226	73.92433	1.561361	**0.119751**	-30.1977	261.0429	-30.1977	261.0429
body_sent	74.04445	74.35179	0.995866	**0.320314**	-72.4179	220.5068	-72.4179	220.5068

[4] Percentage Change & [II] Sentiment Loaded Variables



SUMMARY OUTPUT

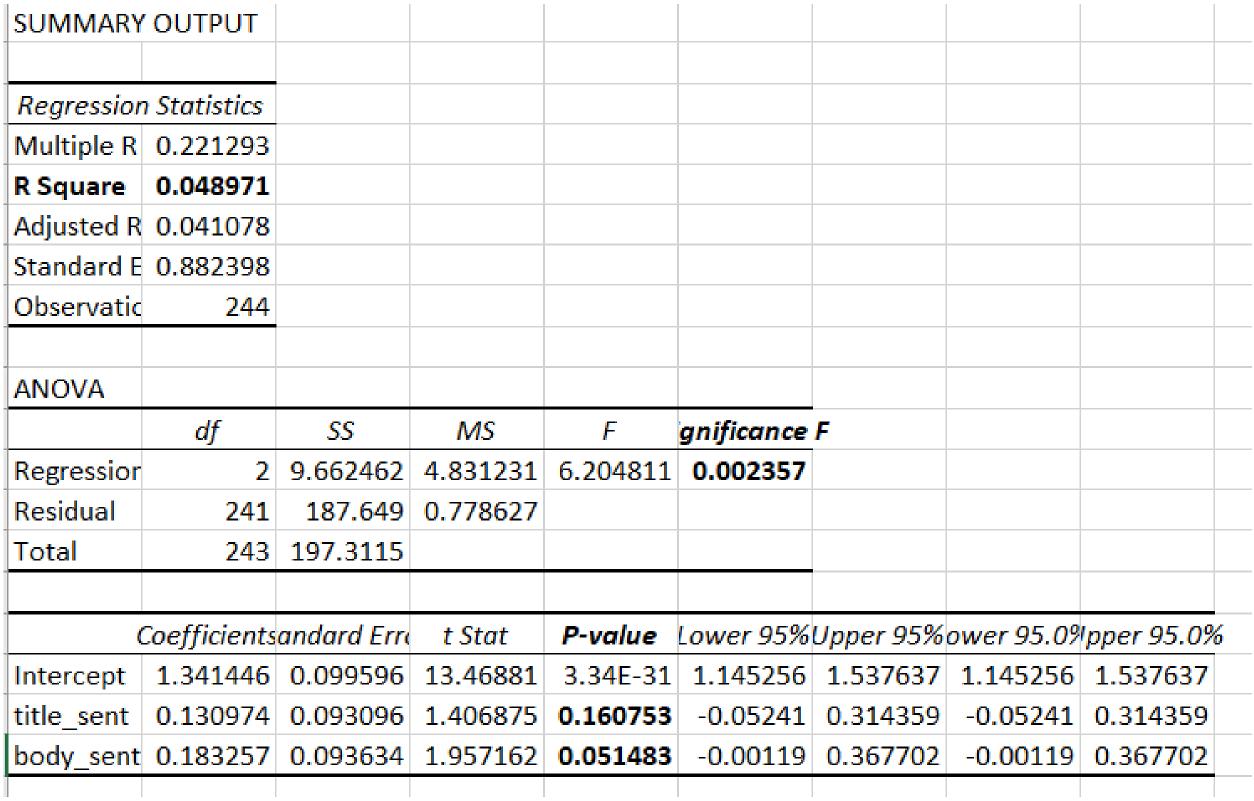
Regression Statistics	
Multiple R	0.190047
R Square	**0.036118**
Adjusted R	0.028119
Standard E	0.072275
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	2	0.047173	0.023586	4.51527	**0.011882**
Residual	241	1.258908	0.005224		
Total	243	1.30608			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-0.00859	0.008158	-1.05293	0.293428	-0.02466	0.00748	-0.02466	0.00748
title_sent	0.012001	0.007625	1.573913	**0.116819**	-0.00302	0.027022	-0.00302	0.027022
body_sent	0.009997	0.007669	1.303442	**0.193668**	-0.00511	0.025104	-0.00511	0.025104

[5] Change Tag & [II] Sentiment Loaded Variables



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.221293
R Square	**0.048971**
Adjusted R	0.041078
Standard E	0.882398
Observatic	244

ANOVA

	df	*SS*	*MS*	*F*	***gnificance F***
Regressior	2	9.662462	4.831231	6.204811	**0.002357**
Residual	241	187.649	0.778627		
Total	243	197.3115			

	Coefficients	*andard Err*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.341446	0.099596	13.46881	3.34E-31	1.145256	1.537637	1.145256	1.537637
title_sent	0.130974	0.093096	1.406875	**0.160753**	-0.05241	0.314359	-0.05241	0.314359
body_sent	0.183257	0.093634	1.957162	**0.051483**	-0.00119	0.367702	-0.00119	0.367702

RANDOM FOREST OUTPUTS

In [1]:

```
import pandas as pd
import numpy as np
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
from datetime import datetime
from sklearn.preprocessing import OneHotEncoder
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline

pd.set_option('display.max_rows', 15)
```

In [2]:

```
def to_date(x):
    comps = x.split('/')
    return datetime(2000 + int(comps[2]), int(comps[0]), int(comps[1]))
```

In [3]:

```
data = pd.read_csv("btc_sentiment_analysis.csv")

data['nag'] = pd.Categorical(data['nag'])
data_categories = pd.get_dummies(data['nag'], prefix='nag')
data = pd.concat([data.iloc[:, 0], data_categories, data.iloc[:, 1:]], axis=1)

data['premium'] = data['premium'].apply(lambda x: 0 if pd.isnull(x) else 1)
data['chars'] = data['chars'].apply(lambda x: max(int(x.split('/')[0]), int(x.
split('/')[1])))

data['date'] = pd.to_datetime(data['date'].apply(to_date))

del data['nag']

del data['chars']
del data['words']

X = data.iloc[:,1:11]  #independent columns

y_perc = data['change_perc']
y_perc = y_perc.apply(lambda x: x * 10000)
y_perc = y_perc.astype('int')

y_tag = data['change_tag']
y_tag = y_tag.astype('int')

y_change = data['change']
y_change = y_change.astype('int')

single_label_data_perc = pd.concat([X, y_perc], axis=1)
single_label_data_tag = pd.concat([X, y_tag], axis=1)
single_label_data_change = pd.concat([X, y_change], axis=1)
```

```
In [4]: single_label_data_change
```

```
Out[4]:
     nag_ANALYSIS  nag_INTERVIEW  nag_NEWS  nag_OTHER  nag_STORY  title_sent  body_ser
  0             1              0         0          0          0           1
  1             1              0         0          0          0           2
  2             0              0         0          0          1           2
  3             0              0         0          0          1           1
  4             1              0         0          0          0           0
  5             1              0         0          0          0           0
  6             0              0         0          0          1           1
...           ...            ...       ...        ...        ...         ...
237             1              0         0          0          0           1
238             0              0         0          0          1           1
239             0              0         0          0          1           1
240             1              0         0          0          0           1
241             1              0         0          0          0           1
242             0              0         0          0          1           1
243             1              0         0          0          0           2

244 rows × 11 columns
```

In [5]:
```
data
```

Out[5]:

	date	nag_ANALYSIS	nag_INTERVIEW	nag_NEWS	nag_OTHER	nag_STORY	title_sent	b
0	2015-01-16	1	0	0	0	0	1	
1	2015-01-20	1	0	0	0	0	2	
2	2015-01-20	0	0	0	0	1	2	
3	2015-02-03	0	0	0	0	1	1	
4	2015-02-03	1	0	0	0	0	0	
5	2015-02-03	1	0	0	0	0	0	
6	2015-02-15	0	0	0	0	1	1	
...	...	...	...	...	...	...	...	
237	2017-12-14	1	0	0	0	0	1	
238	2017-12-14	0	0	0	0	1	1	
239	2017-12-14	0	0	0	0	1	1	
240	2017-12-15	1	0	0	0	0	1	
241	2017-12-15	1	0	0	0	0	1	
242	2017-12-15	0	0	0	0	1	1	
243	2017-12-17	1	0	0	0	0	2	

244 rows × 19 columns

In [6]:
```
X.dtypes
```

Out[6]:
```
nag_ANALYSIS      uint8
nag_INTERVIEW     uint8
nag_NEWS          uint8
nag_OTHER         uint8
nag_STORY         uint8
title_sent        int64
body_sent         int64
bubble            int64
alphaville        int64
premium           int64
dtype: object
```

In [7]:
```
y_perc.dtypes
```

Out[7]:
```
dtype('int64')
```

In [8]:
```
X
```

Out[8]:

	nag_ANALYSIS	nag_INTERVIEW	nag_NEWS	nag_OTHER	nag_STORY	title_sent	body_sei
0	1	0	0	0	0	1	
1	1	0	0	0	0	2	
2	0	0	0	0	1	2	
3	0	0	0	0	1	1	
4	1	0	0	0	0	0	
5	1	0	0	0	0	0	
6	0	0	0	0	1	1	
...	...	...	...	...	...	...	
237	1	0	0	0	0	1	
238	0	0	0	0	1	1	
239	0	0	0	0	1	1	
240	1	0	0	0	0	1	
241	1	0	0	0	0	1	
242	0	0	0	0	1	1	
243	1	0	0	0	0	2	

244 rows × 10 columns

In [9]:
```
y_perc
```

Out[9]:
```
0      -129
1      -124
2      -124
3      -446
4      -446
5      -446
6      -890
       ... 
237      95
238      95
239      95
240     651
241     651
242     651
243    -154
Name: change_perc, Length: 244, dtype: int64
```

Univariate Selection

In [32]:

```
# https://stats.stackexchange.com/questions/24179/how-exactly-does-chi-square-
feature-selection-work
```

In [10]:

```
#apply SelectKBest class to extract top 10 best features
bestfeatures = SelectKBest(score_func=chi2, k=6)
fit = bestfeatures.fit(X, y_tag)
dfscores = pd.DataFrame(fit.scores_)
dfcolumns = pd.DataFrame(X.columns)
#concat two dataframes for better visualization
featureScores = pd.concat([dfcolumns,dfscores],axis=1)
featureScores.columns = ['Specs','Score']  #naming the dataframe columns
print(featureScores.nlargest(10,'Score'))  #print 10 best features
```

```
          Specs     Score
2      nag_NEWS  7.085707
6     body_sent  6.915646
5    title_sent  5.117973
8    alphaville  2.417011
3     nag_OTHER  2.395709
9       premium  1.710777
1 nag_INTERVIEW  1.674330
4     nag_STORY  1.523090
7        bubble  1.245417
0  nag_ANALYSIS  0.644100
```

Feature Importance

In [ ]:

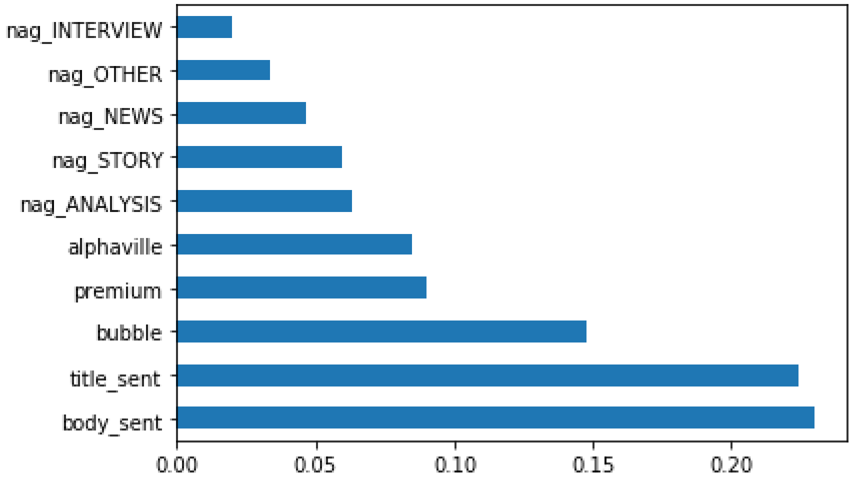
```
# https://towardsdatascience.com/feature-selection-using-random-forest-26d7b74
7597f
```

Against Change Percentage

In [11]:

```
from sklearn.ensemble import ExtraTreesClassifier
import matplotlib.pyplot as plt
%matplotlib inline
model = ExtraTreesClassifier(100)
model.fit(X, y_perc)
print(model.feature_importances_) #use inbuilt class feature_importances of tree based classifiers
#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

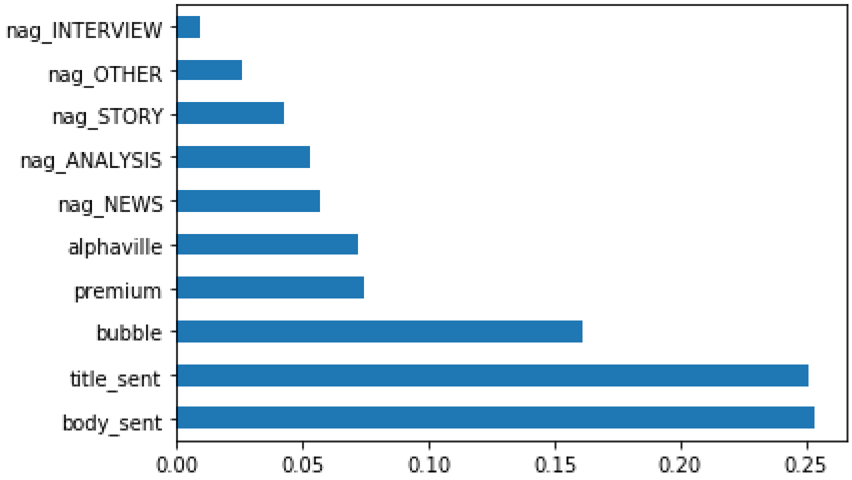
```
[0.06315346 0.01980092 0.04661252 0.0337949  0.05966783 0.22448904
 0.23018087 0.14759374 0.08474311 0.0899636 ]
```



Against Change Tag

```
In [12]: from sklearn.ensemble import ExtraTreesClassifier
         import matplotlib.pyplot as plt
         %matplotlib inline
         model = ExtraTreesClassifier(100)
         model.fit(X, y_tag)
         print(model.feature_importances_) #use inbuilt class feature_importances of tr
         ee based classifiers
         #plot graph of feature importances for better visualization
         feat_importances = pd.Series(model.feature_importances_, index=X.columns)
         feat_importances.nlargest(10).plot(kind='barh')
         plt.show()
```

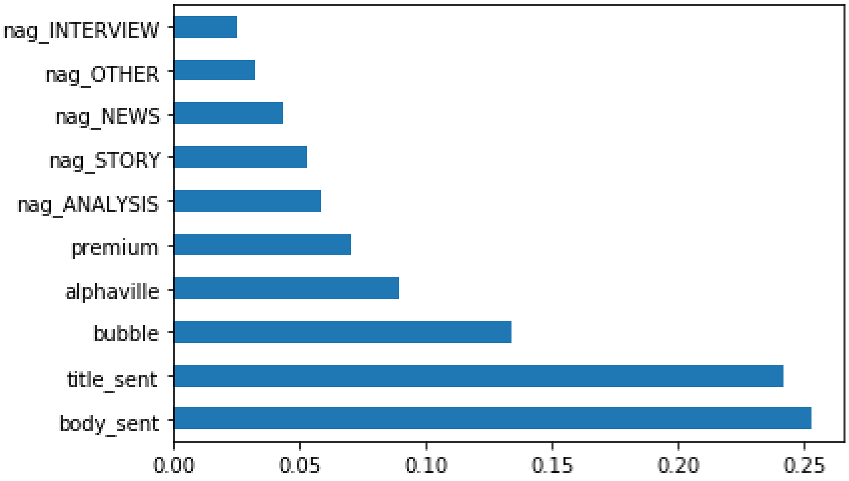
```
[0.05283087 0.00913154 0.05698698 0.02594775 0.04248568 0.25140382
 0.25364048 0.1614551  0.07176143 0.07435635]
```



Against Absolute Change

In [13]:
```
from sklearn.ensemble import ExtraTreesClassifier
model = ExtraTreesClassifier(100)
model.fit(X, y_change)
print(model.feature_importances_) #use inbuilt class feature_importances of tr
ee based classifiers
#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

```
[0.05860946 0.02505341 0.04316355 0.03257819 0.05282877 0.24133596
 0.25306543 0.13391132 0.08893992 0.070514  ]
```

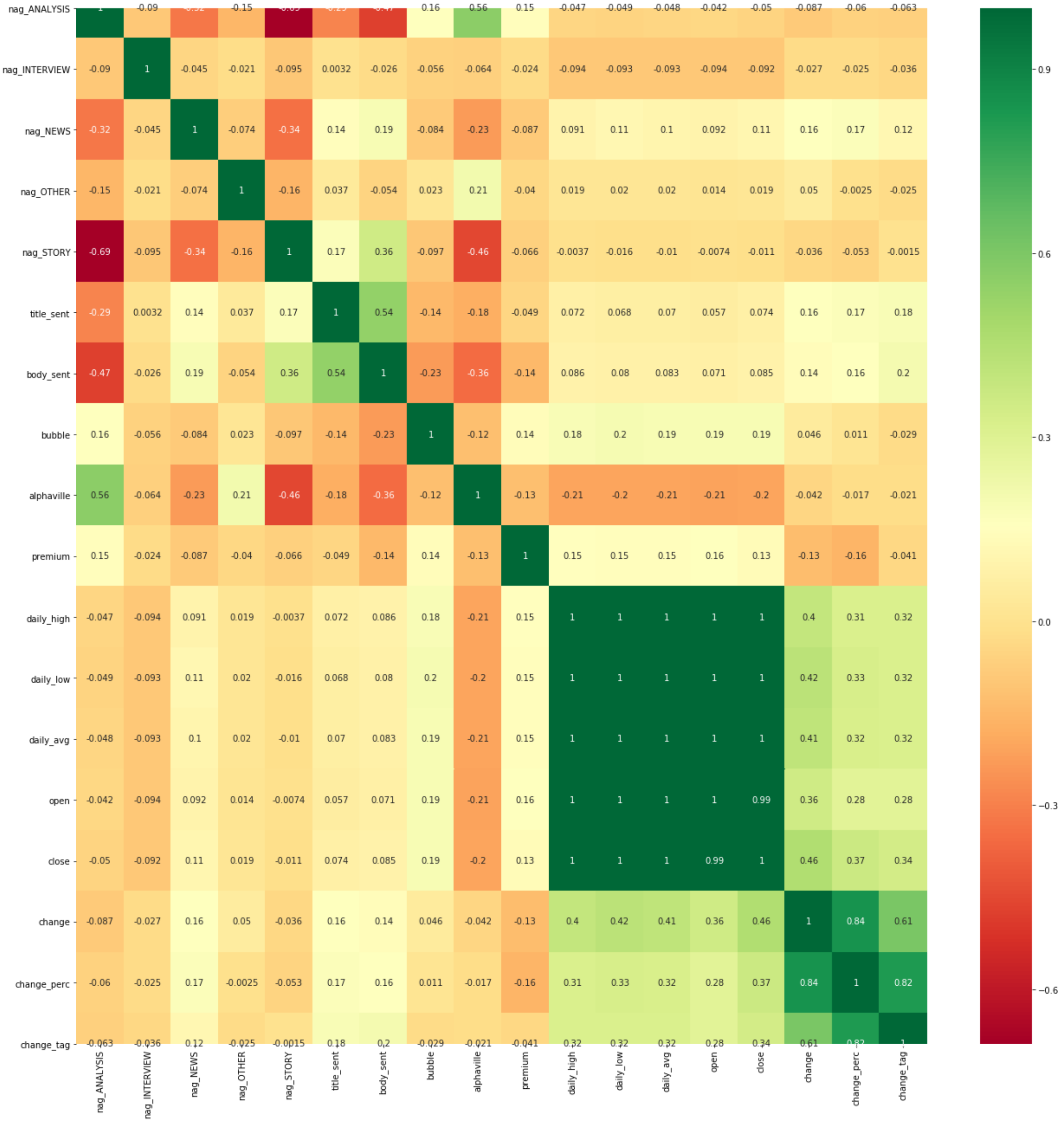


Correlation Matrix

In [33]:
```
# https://www.statisticshowto.datasciencecentral.com/correlation-matrix/
```

In [14]:

```
corrmat = data.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(22, 22))
# plot heat map
g=sns.heatmap(data[top_corr_features].corr(), annot=True, cmap="RdYlGn")
```



In [ ]:

```
```

Principal Component Analysis

In [35]:

```
from sklearn.decomposition import PCA

pca = PCA(n_components=3)
fit = pca.fit(X)
# summarize components
print(fit.explained_variance_ratio_)
print(fit.components_)

principal_components = pca.fit_transform(X)
principal_df = pd.DataFrame(data = principal_components)

# we see in the first principal component that title sentiment and the body se
ntiment features have the highest relevance
# those are the 5.74496588e-01 and 6.45111132e-01 values respectively
```

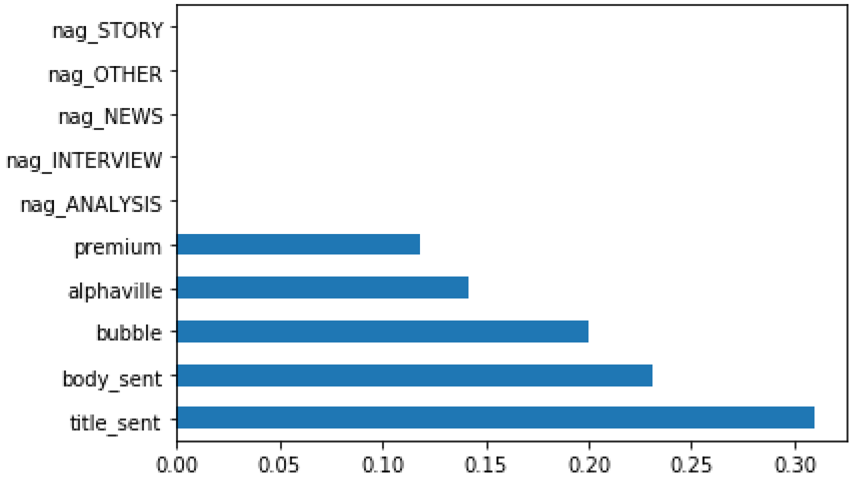
```
[0.46214426 0.18129422 0.10458358]
[[-3.39789020e-01 -1.55596349e-04  7.50466464e-02 -5.59921886e-03
   2.70497189e-01  5.74496588e-01  6.45111132e-01 -9.80092315e-02
  -2.21822850e-01 -2.61806997e-02]
 [-4.40244667e-01  8.86104190e-04 -7.21364395e-02 -2.92387170e-02
   5.40733719e-01 -6.14072657e-01 -1.91491605e-02  4.91320751e-02
  -3.57582998e-01 -9.19288572e-04]
 [-9.87761618e-02  6.91895535e-04 -2.75872267e-01  1.58783299e-02
   3.58078203e-01  5.29010830e-01 -6.34214322e-01  2.91505490e-01
  -1.17178177e-01  6.98478500e-02]]
```

ANALYSIS Nag Only

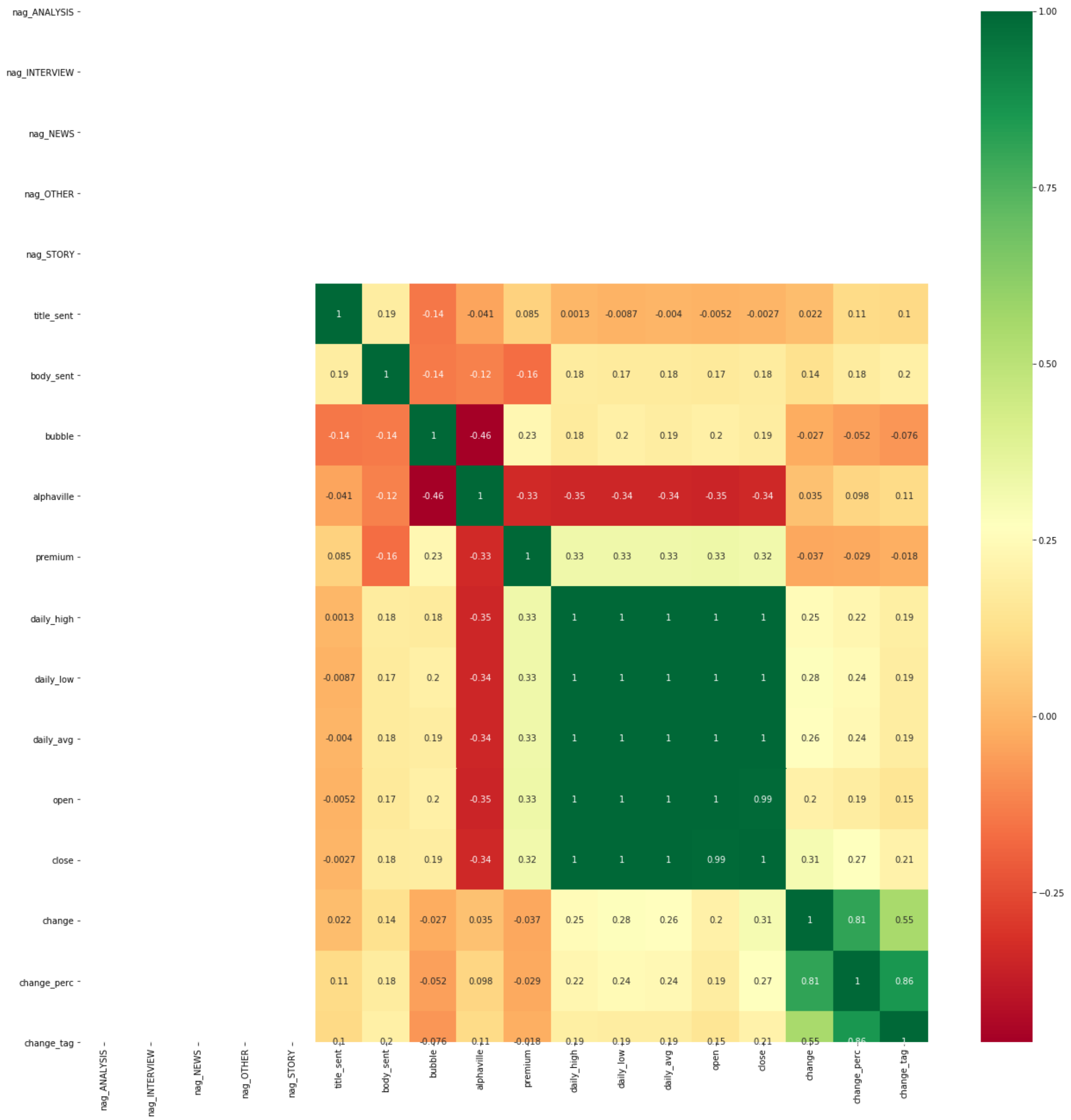
In [23]:

```
from sklearn.ensemble import ExtraTreesClassifier

data_analysis = data[data.nag_ANALYSIS == 1]

X_analysis = data_analysis.iloc[:,1:11]  #independent columns

y_analysis_change = data_analysis['change']
y_analysis_change = y_analysis_change.astype('int')

model = ExtraTreesClassifier(100)
model.fit(X_analysis, y_analysis_change)
print(model.feature_importances_) #use inbuilt class feature_importances of tr
ee based classifiers
#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X_analysis.colu
mns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

```
[0.         0.         0.         0.         0.         0.31012704
 0.23080373 0.19970667 0.14153218 0.11783039]
```



In [24]:
```
corrmat = data_analysis.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(22, 22))
# plot heat map
g = sns.heatmap(data_analysis[top_corr_features].corr(), annot=True, cmap="RdY
lGn")
```

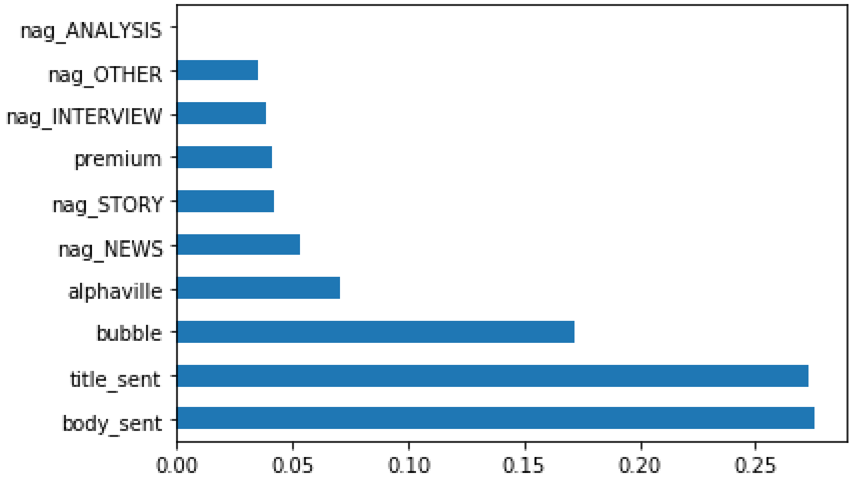


Everything apart from ANALYSIS Nag

In [29]:

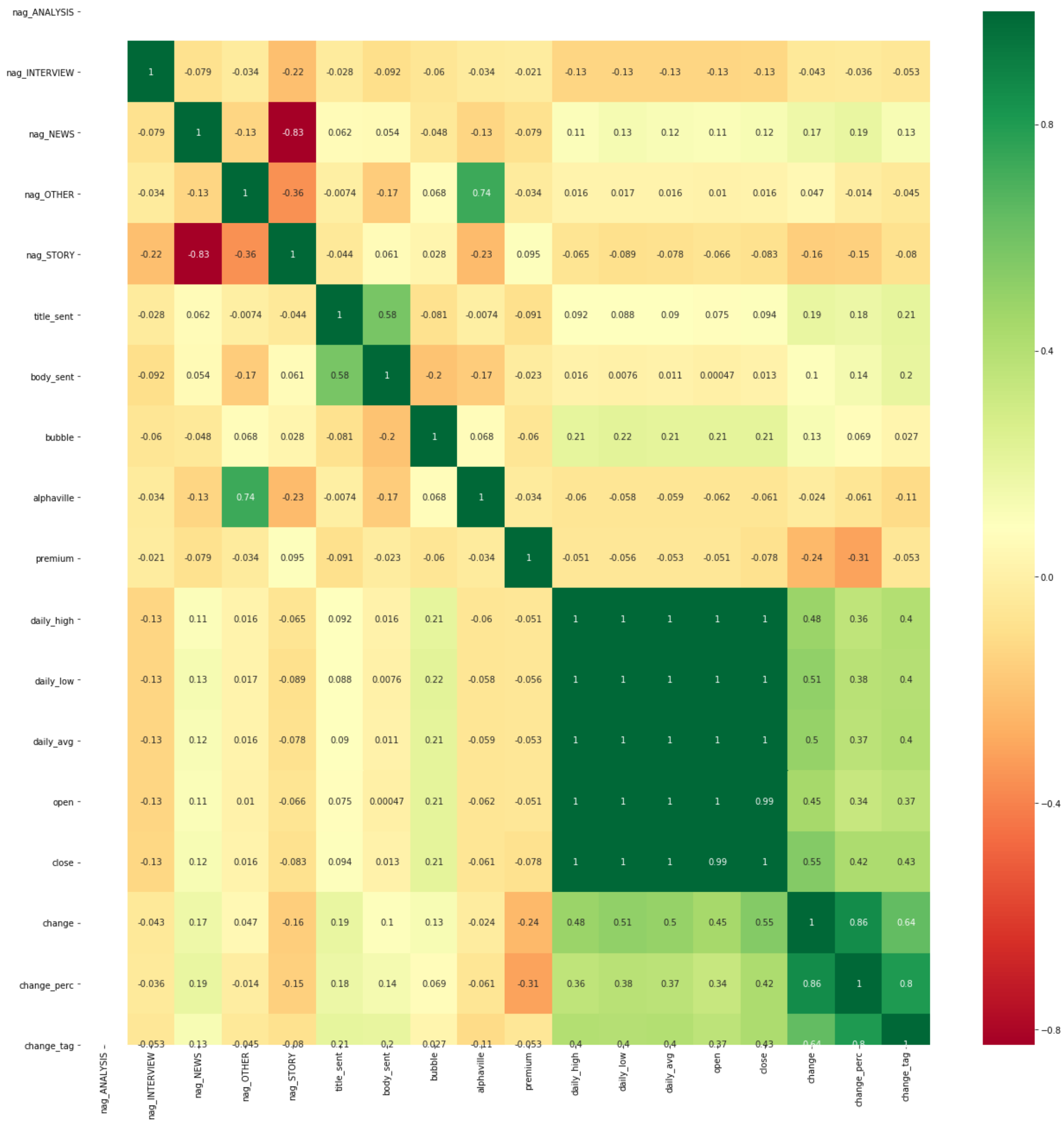
```
from sklearn.ensemble import ExtraTreesClassifier

data_not_analysis = data[data.nag_ANALYSIS == 0]

X_not_analysis = data_not_analysis.iloc[:,1:11]  #independent columns

y_not_analysis_change = data_not_analysis['change']
y_not_analysis_change = y_not_analysis_change.astype('int')

model = ExtraTreesClassifier(100)
model.fit(X_not_analysis, y_not_analysis_change)
print(model.feature_importances_) #use inbuilt class feature_importances of tr
ee based classifiers
#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X_not_analysis.
columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

```
[0.         0.03815888 0.05324108 0.03466422 0.04237872 0.27260144
 0.27582289 0.17142612 0.07018703 0.04151962]
```



In [30]:

```
corrmat = data_not_analysis.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(22, 22))
#plot heat map
g=sns.heatmap(data_not_analysis[top_corr_features].corr(), annot=True, cmap="RdYlGn")
```



In [ ]:

In [1]:

```
import pandas as pd
import numpy as np
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
from datetime import datetime
from sklearn.preprocessing import OneHotEncoder
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm

%matplotlib inline

pd.set_option('display.max_rows', 35)
```

In [2]:

```
def to_date(x):
    comps = x.split('/')
    return datetime(2000 + int(comps[2]), int(comps[0]), int(comps[1]))
```

In [3]:

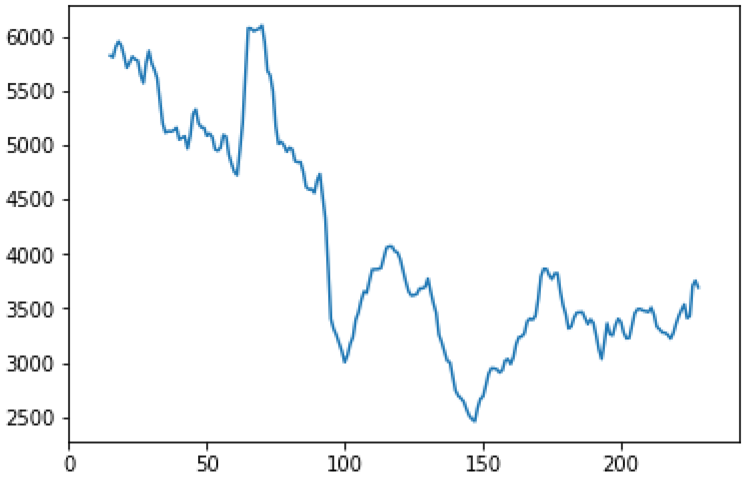
```
data = pd.read_csv("btc_sentiment_analysis.csv")

data['nag'] = pd.Categorical(data['nag'])
data_categories = pd.get_dummies(data['nag'], prefix='nag')
data = pd.concat([data.iloc[:, 0], data_categories, data.iloc[:, 1:]], axis=1)

data['premium'] = data['premium'].apply(lambda x: 0 if pd.isnull(x) else 1)
data['chars'] = data['chars'].apply(lambda x: max(int(x.split('/')[0]), int(x.
split('/')[1])))

data['date'] = pd.to_datetime(data['date'].apply(to_date))

del data['nag']

X = data.iloc[:,1:11]  #independent columns

y_perc = data['change_perc']
y_perc = y_perc.apply(lambda x: x * 100)
y_perc = y_perc.astype('int')

y_tag = data['change_tag']
y_tag = y_tag.astype('int')

y_change = data['change']
y_change = y_change.astype('int')

single_label_data_perc = pd.concat([X, y_perc], axis=1)
single_label_data_tag = pd.concat([X, y_tag], axis=1)
single_label_data_change = pd.concat([X, y_change], axis=1)

# data = data.set_index('date')
data.index

analysis_only = data[data.nag_ANALYSIS == 1]
non_analysis = data[data.nag_ANALYSIS == 0]
```

Trend of length over time

https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/ (https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/)

In [4]:
```
sm.tsa.seasonal_decompose(data['chars'], model='additive', freq=30).trend.plot
()
```

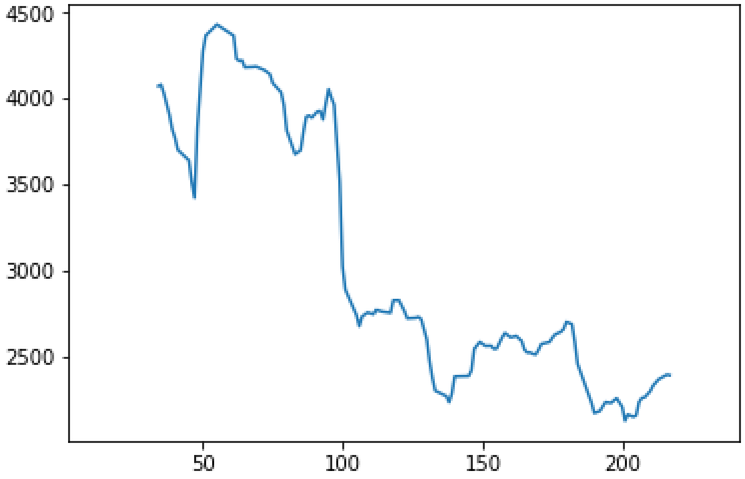
Out[4]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x120d27f28>
```



Trend of non_analysis length over time

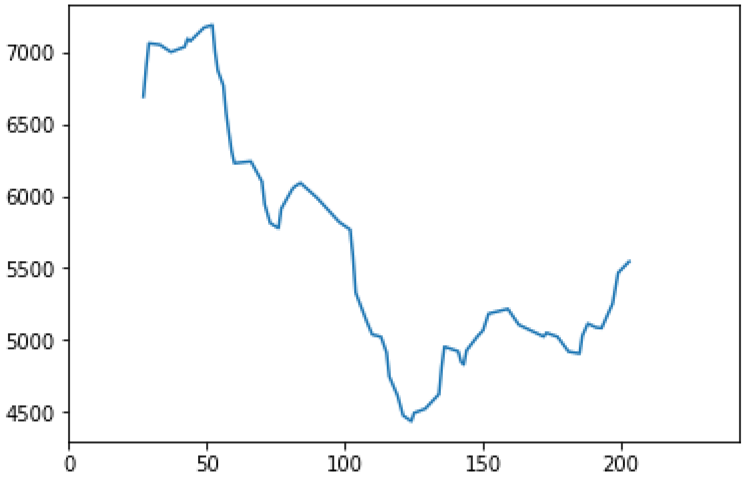
In [5]:
```
sm.tsa.seasonal_decompose(non_analysis['chars'], model='additive', freq=30).tr
end.plot()
```

Out[5]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x120eedcf8>
```



Trend of analysis length over time

```
In [6]: sm.tsa.seasonal_decompose(analysis_only['chars'], model='additive', freq=30).t
        rend.plot()

Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x12300ad30>
```

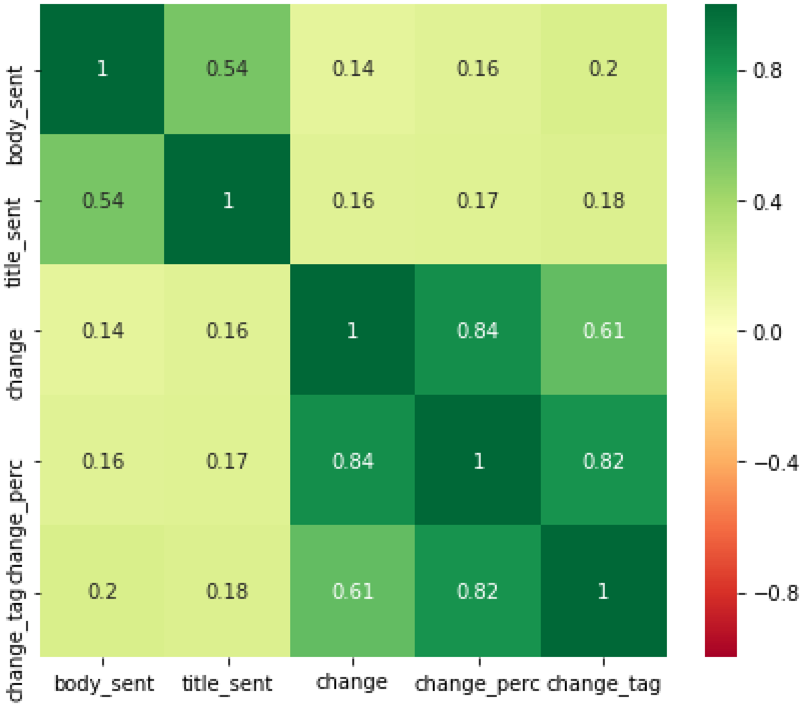


Sentiment of news and price action

[https://www.statisticshowto.datasciencecentral.com/correlation-matrix/ (https://www.statisticshowto.datasciencecentral.com/correlation-matrix/)](https://www.statisticshowto.datasciencecentral.com/correlation-matrix/)

In [7]:

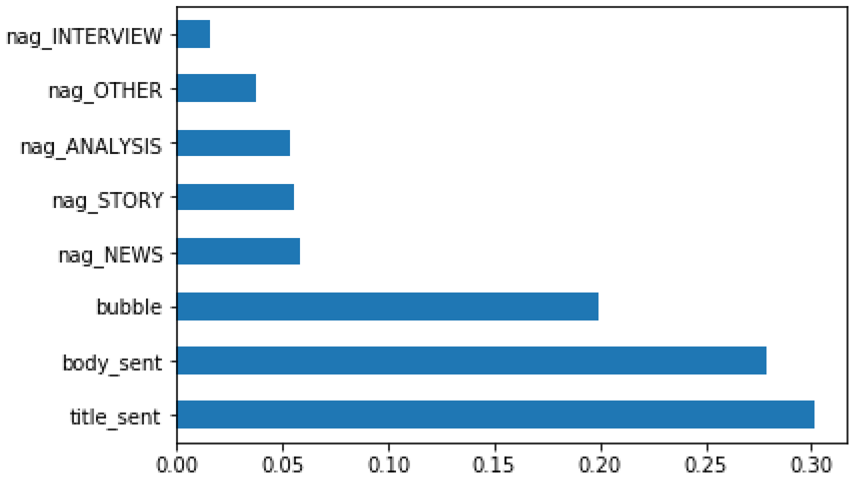
```
length_change = data[['body_sent', 'title_sent', 'change', 'change_perc', 'change_tag']]

corrmat = length_change.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(7, 6))
#plot heat map
g = sns.heatmap(length_change[top_corr_features].corr(), vmin=-1, vmax=1, annot=True, cmap="RdYlGn")
```



https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f (https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f)

In [8]:
```
from sklearn.ensemble import ExtraTreesClassifier

X_no_length = X[X.columns.difference(['words', 'chars'])]

model = ExtraTreesClassifier(100)
model.fit(X_no_length, y_perc)
print(model.feature_importances_)
feat_importances = pd.Series(model.feature_importances_, index=X_no_length.col
umns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

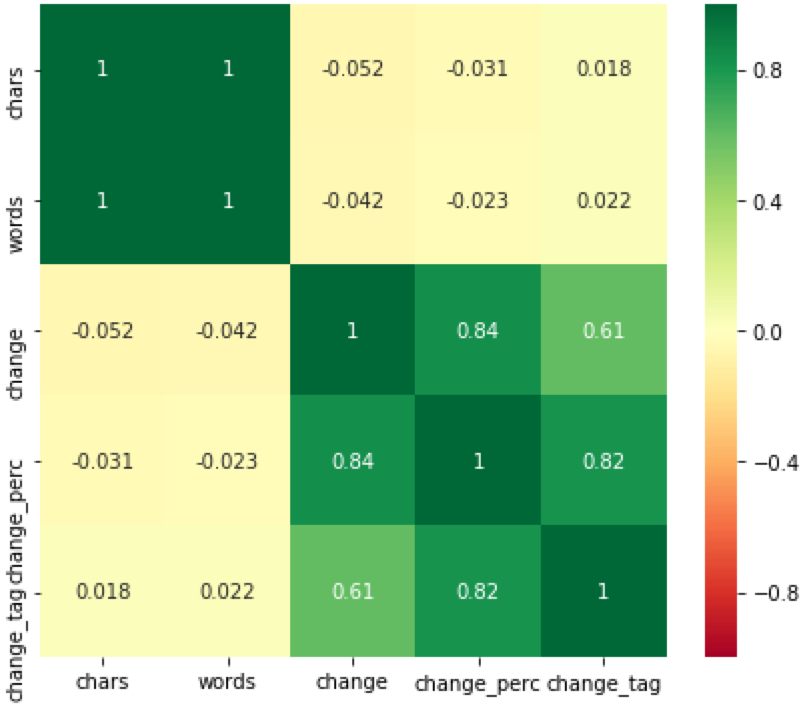
```
[0.27892421 0.19887292 0.05321159 0.01586323 0.05835282 0.03705889
 0.05584297 0.30187336]
```



Length of news and price action

https://www.statisticshowto.datasciencecentral.com/correlation-matrix/
(https://www.statisticshowto.datasciencecentral.com/correlation-matrix/)

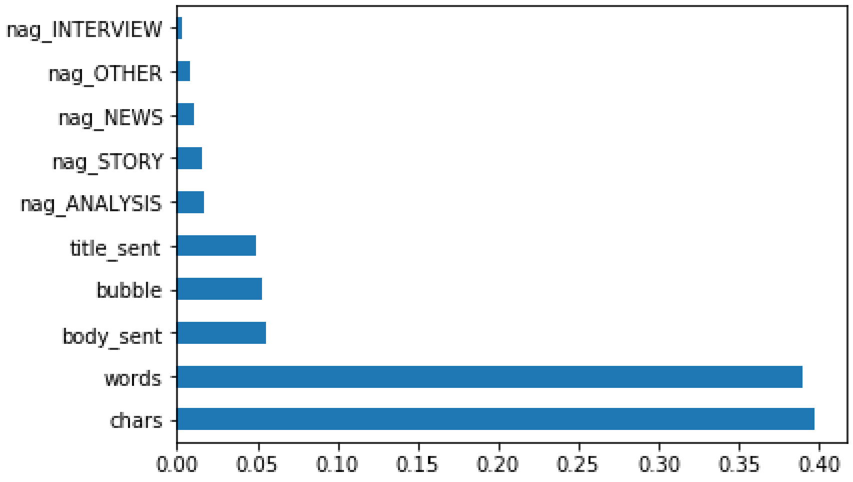
```
In [9]: length_change = data[['chars', 'words', 'change', 'change_perc', 'change_tag'
        ]]

        corrmat = length_change.corr()
        top_corr_features = corrmat.index
        plt.figure(figsize=(7, 6))
        g = sns.heatmap(length_change[top_corr_features].corr(), vmin=-1, vmax=1, anno
        t=True, cmap="RdYlGn")
```



https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f (https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f)

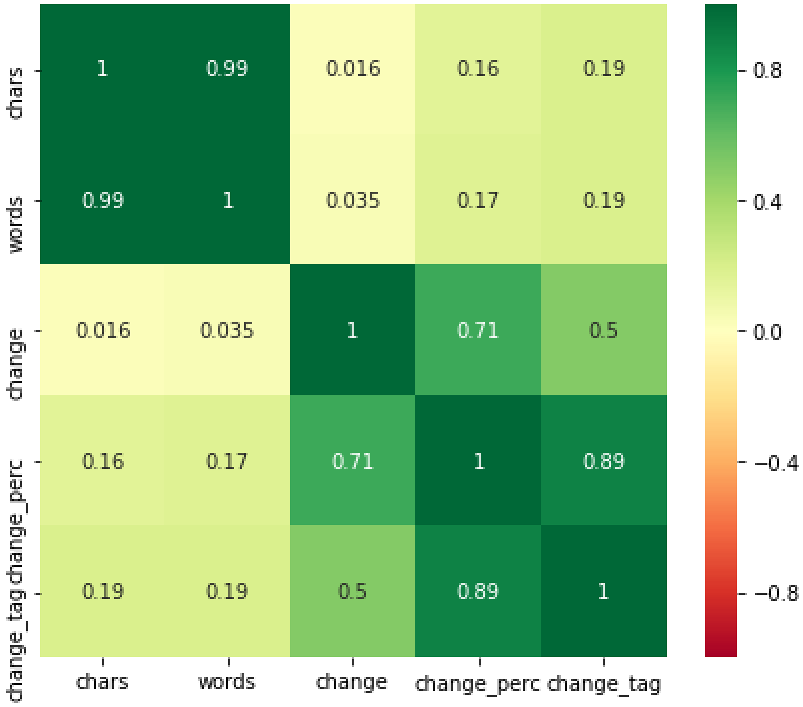
```
In [10]: from sklearn.ensemble import ExtraTreesClassifier
         model = ExtraTreesClassifier(100)
         model.fit(X, y_perc)
         print(model.feature_importances_)
         feat_importances = pd.Series(model.feature_importances_, index=X.columns)
         feat_importances.nlargest(10).plot(kind='barh')
         plt.show()
```

```
[0.01722604 0.00277084 0.01088641 0.00755505 0.01600563 0.04978062
 0.05575228 0.05276376 0.38950843 0.39775094]
```



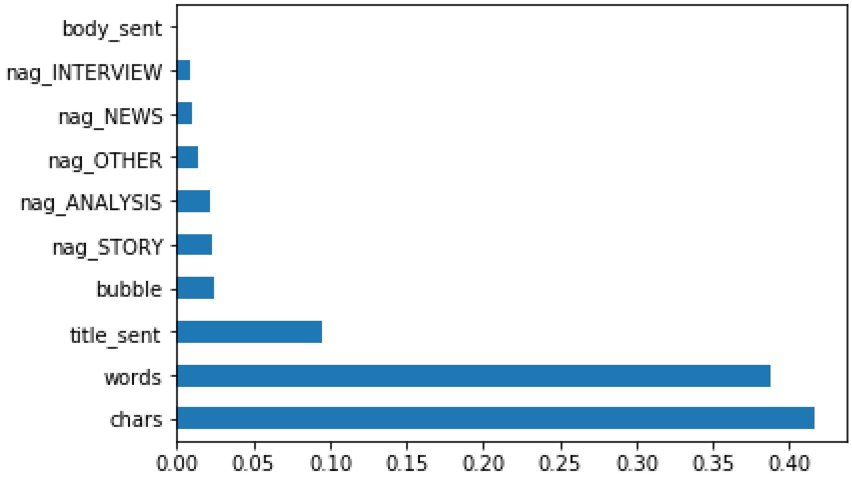
Length of negative news and price action

```
In [11]: negative_data = data[data.body_sent == 0]
         negative_length_change = negative_data[['chars', 'words', 'change', 'change_perc', 'change_tag']]

         corrmat = negative_length_change.corr()
         top_corr_features = corrmat.index
         plt.figure(figsize=(7, 6))
         g = sns.heatmap(negative_length_change[top_corr_features].corr(), vmin=-1, vmax=1, annot=True, cmap="RdYlGn")
```



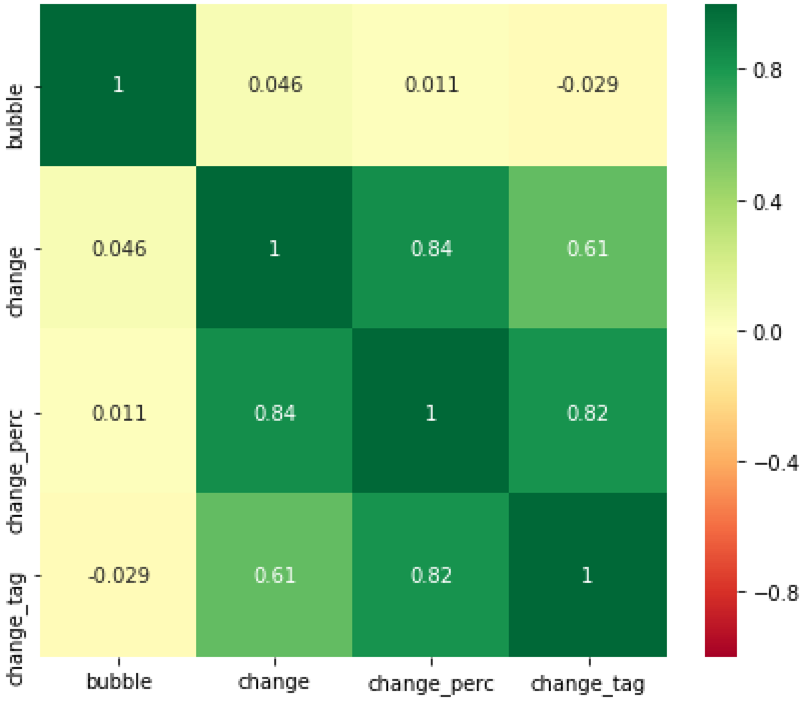
```
In [12]: from sklearn.ensemble import ExtraTreesClassifier

         X_negative = negative_data.iloc[:,1:11]  #independent columns
         y_negative_perc = negative_data['change_perc']
         y_negative_perc = y_negative_perc.apply(lambda x: x * 10000)
         y_negative_perc = y_negative_perc.astype('int')

         model = ExtraTreesClassifier(100)
         model.fit(X_negative, y_negative_perc)
         print(model.feature_importances_)
         feat_importances = pd.Series(model.feature_importances_, index=X_negative.colu
         mns)
         feat_importances.nlargest(10).plot(kind='barh')
         plt.show()
```

```
[0.02212087 0.00794961 0.00985367 0.01430674 0.02266329 0.09440915
 0.         0.02462898 0.38732223 0.41674545]
```



Price action and articles suggesting that BTC might be in a bubble state

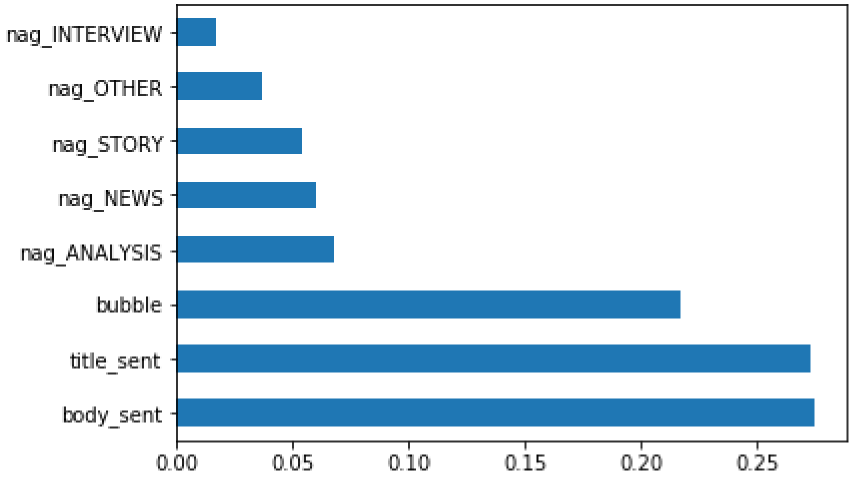
```
In [13]: bubble = data[['bubble', 'change', 'change_perc', 'change_tag']]

         corrmat = bubble.corr()
         top_corr_features = corrmat.index
         plt.figure(figsize=(7, 6))
         g = sns.heatmap(bubble[top_corr_features].corr(), vmin=-1, vmax=1, annot=True,
         cmap="RdYlGn")
```



In [14]:

```
from sklearn.ensemble import ExtraTreesClassifier

X_no_length = X[X.columns.difference(['words', 'chars'])]

model = ExtraTreesClassifier(100)
model.fit(X_no_length, y_perc)
print(model.feature_importances_)
feat_importances = pd.Series(model.feature_importances_, index=X_no_length.col
umns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

```
[0.27511957 0.21746709 0.06768672 0.01682016 0.0596378  0.03654486
 0.05397913 0.27274467]
```


nag_INTERVIEW
nag_OTHER
nag_STORY
nag_NEWS
nag_ANALYSIS
bubble
title_sent
body_sent