

POLITECNICO DI TORINO

Master 's Degree in
Management Engineering



Master Thesis

Initial Coin Offerings
An analysis of success factors

Supervisor

Prof.ssa Elisa Ughetto

Candidate

Alfredo Scermino

A.A. 2020-2021

A special thanks to my family and best friends
who have always stood beside me
through this journey

SUMMARY

1	Chapter 1.....	7
1.1	LITERATURE REVIEW	7
1.2	ICOs introduction	14
1.3	Origin of Blockchain	15
1.3.1	The beginnings	16
1.3.2	Reusable Proof Of Work.....	17
1.3.3	Bitcoin network	18
1.3.4	Ethereum.....	20
1.4	Initial Coin Offerings.....	20
1.5	Comparison of ICOs to conventional financing methods.....	22
1.6	Phases of ICO.....	23
1.7	ICOs regulations	24
1.8	ICOs Geography.....	28
1.9	ICOs trends and challenges	30
2	Chapter 2.....	33
2.1	Database methodology	33
2.2	List of sources.....	33
2.3	Methodology	35
2.3.1	Information from Coinschedule	36
2.3.2	Information from ICObench	38
2.3.3	Information from Coinmarketcap.com	43
2.3.4	Information from web sources.....	45
3	Chapter 3.....	47
3.1	Adjustments to database	47
3.1.1	Data saved in eth:.....	47
3.1.2	Data saved in usd:	47
3.1.3	Data saved in other currencies:	48
3.2	Descriptive Statistics	48
3.2.1	Arithmetic mean.....	48
3.2.2	Standard Error	49
3.2.3	Median	49
3.2.4	Mode	49
3.2.5	Standard deviation	50

3.2.6	Variance.....	50
3.2.7	Kurtosis.....	50
3.2.8	Asymmetry (skewness).....	51
3.2.9	Interval	51
3.2.10	Minimum value	51
3.2.11	Maximum value.....	51
3.3	Descriptive statistics – output.....	51
3.3.1	Analysis of the entire database.....	52
3.3.2	Analysis of subsamples divided by country.....	53
3.3.3	Analysis of subsamples divided by category	55
3.3.4	Analysis of subsample divided by Platform.....	57
3.3.5	Analysis of subsample characterized by Pre-ICO existence	58
3.3.6	Analysis of subsample characterized by existence or absence of source code ..	59
3.3.7	Analysis of subsample characterized by usability of token as currency	60
3.3.8	Analysis of subsample characterized by positive and negative market sentiment 61	
3.3.9	Analysis of subsample characterized by presence of bonus scheme	63
3.3.10	Analysis of subsample characterized by disclosure of the team.....	63
3.4	Output descriptive graphs.....	64
3.5	Conclusion	69
4	Chapter 4.....	70
4.1	Database Methodology	70
4.2	Source.....	70
4.3	Methodology	71
4.3.1	Information from LexisNexis	72
4.3.2	Information from Alexa	73
4.3.3	Analysis of news texts	73
4.3.4	Theoretical framework of the variables.....	74
5	Chapter 5.....	77
5.1	Adjustments to database	77
5.1.1	ANOVA analyses	77
5.1.2	Anova hypothesis	78
5.1.3	Statistical test	78
5.1.4	ANOVA between Categories and Lexical diversity	80
5.1.5	ANOVA between Categories and Sentiment.....	81

5.1.6	ANOVA between Categories and Posemo.....	83
5.1.7	ANOVA between Categories and FocusFuture	84
5.1.8	ANOVA between Categories and FocusPresent.....	86
5.1.9	ANOVA between Countries and Sentiment	87
5.1.10	ANOVA between Countries and Posemo	90
5.1.11	ANOVA between Countries and FocusFuture	92
5.1.12	ANOVA between Countries and FocusPresent	94
5.1.13	ANOVA between Source Code and Sentiment and Posemo.....	97
5.1.14	ANOVA between Team Disclosure and Sentiment and Posemo	98
5.1.15	ANOVA between Pre-ICO Phase and Sentiment and Posemo	99
5.1.16	ANOVA between Bonus Program and Sentiment and Posemo	100
5.1.17	ANOVA between MVP presence and Sentiment and Posemo.....	101
5.2	Correlations between news factors and capital raised.....	102
5.2.1	Pearson Correlation.....	103
5.2.2	Correlation parameters.....	104
5.2.3	Conclusions.....	106
6	Avenues for future research	108
7	Bibliography	110

1 CHAPTER 1

1.1 LITERATURE REVIEW

In the latest years, from 2015 until now, cryptocurrencies (cryptos) and all the financial instruments and events linked to it has become very popular following a hype wave of excitement.

These cryptos, like Bitcoin or Ether, are traded thanks to the existence of the blockchain technology, which is the concept of a distributed ledger (DL) system where all new transactions are added to the chain thanks to the due diligence of people who are part of the network. They are called 'miners' and assure the correctness of the code of the transaction and deliver security to the system, which is not based on a trusted centralized authority.

Talking about entrepreneurial finance related to this technology, a very interesting financing tool for new businesses and start-ups in early stages are ICO's (initial coins offerings), similar to IPOs (initial public offerings) where a young firm offers rewards or securities using specific crypto-tokens in exchange of cryptocurrencies to fund their business.

A high number of studies have been conducted on these revolutionary and disruptive topics giving to the community more and more insights and revealing many interesting aspects, starting with research agendas about Blockchain, Bitcoin and ICOs, offering nontechnical explanations of the basics of blockchain and its applications, cryptocurrencies and ICOs (Kher, Terjese, & Liu, 2019).

But a much relevant and discussed topic in the literature is finding the success factors that characterize the ICOs. The ones first studied and analysed, indeed, are factors, like the presence and visibility of the source code, the existence of documents called "whitepapers", where it should be written, voluntarily by the issuers, information like the aim and final goal of the business and the organization

of a token pre-sale. (Adhami, Giudici, & Martinazzi, Why do businesses go crypto? An empirical analysis of initial coin, 2018)

It has been found that probability of success of an ICO is unaffected by the availability of a “whitepaper” (and pre-sales are only marginally significant for the probability of success of the campaign) but is strongly and positively affected by the presence of a set of codes for the blockchain project.

Sharing the source code is a tool that reaches only the share of technologically prepared investors, though, more aware of the quality of the business. While it does not give much information to people who do not have that technological knowledge. The pre-sale helps understanding better the value of the tokens being sold and overcome as much as possible the under-pricing problem.

Following, other relevant factors were relating exploring the influence of the ICO presale period, Bitcoin spot and futures returns, and sentiment extracted from social media on ICO returns. (Domingo, Pineiro-Chousa, & Lopez-cabarcos, 2020)

The results show that Bitcoin spot and futures’ returns exert a positive influence on ICO returns, the existence of a presale period exerts a negative influence and lastly social media sentiment positively influences ICO returns.

The very interesting aspect of correlation between ICOs and ether/bitcoin cryptocurrencies and market cycles has been deeply analysed finding that the hype surrounding one ICO positively influences subsequent ICOs and innovations in either bitcoin or ether positively influence ICOs. (Masiak & Block, Initial coin offerings (ICOs): market cycles and relationship, 2020). It has been found an effect of the growth rates of ICO volumes on cryptocurrency returns, sustaining Domingo et al. findings, and no significant effect at all for the volatility of cryptocurrency on ICO volumes. Moreover, the results show that ICO timing is very important, and this can lead entrepreneurs to eventually decide to postpone their ICOs in ‘hot’ market periods.

An important insight is for investors who want to diversify their portfolio and reduce risk, they should be aware of investing in both ICOs and cryptos since the results of the study show that cryptocurrencies and ICOs returns are correlated.

Other relevant factors have been studied, following the path to determine the factors that lead to success/failure (like team quality, vision of projects and adverse industry events) of an ICOs basing their analysis on outcomes like first day returns, gross proceeds, and time to market. The results showed that the quality of the management team is a good predictor of ICO success, since this entails the fact that the business has strong technological background. On the other hand, it has been shown that highly visionary projects fail more often, since they are perceived by investors as more difficult to realize. (Momtaz, Entrepreneurial Finance and Moral Hazard: Evidence from Token, 2020)

Another subject of study is the role of digital social capital in ICO success. Several variables have been explored, such as official website, ICO activity on social networks like Twitter, the community built and the activeness of the community (Perez, Sokolova, & Konate, 2020). The results highlight that website social capital is a strong determinant of ICO performance and in addition is a signal of non-scam. The role of twitter instead is revealed to be of much minor contribution with a small linear effect on performance. Lastly the overall activity of the community seems to be more important than twitter social capital.

A limitation of the early literature on token offerings is its lack of conceptualization of token offering dynamics. On this topic the application of signalling theory to examine the dynamics of the funding amount in token offerings was crucial, showing that seller firms want to 'signal' their quality to the investors through whitepapers technicality, high source code quality or following Ethereum based standards in order to reduce the information asymmetry belonging to the system (Fisch, Initial coin offerings (ICOs) to finance new ventures, 2019). The information content of whitepapers is highly significant in relation to the successful ending of the linked

ICOs (Florisyak & Schandlbauer, 2020). Their length and complexity lead to higher amount raised, telling us that investors perceive projects with more complex whitepaper as more successful. Over time the average length of white papers has increased suggesting that ICO issuers understood the relevance of white papers and started to voluntarily disclose information. (Samieifair & Baur, 2020)

On the other hand, potential signal like patents are not related to the success of ICOs like in traditional entrepreneurial finance, this might be because of its scarce usability in the DL technology because code is not generally patentable. (Fisch, Initial coin offerings (ICOs) to finance new ventures, 2019) Analysing more in detail the concept of signalling previously explored by Fisch there is reasonable incentive by entrepreneurs to exaggerate their signals to investors leading to moral hazard behaviours. (Momtaz, Entrepreneurial Finance and Moral Hazard: Evidence from Token, 2020) The entrepreneurs are incentivized to do so because they can only tap the market once, contrary to what might be understood by the term ICO (initial coin offering) that suggest that could be more than one offering, so it's not important for them to satisfy the investors repetitively. This happens also because there are several obstacles to detect these signals ex ante and there could not be ex post punishment due to the anonymity of users.

Given this fact they tend to exaggerate about the information provided and the investors fail in perceiving this, probably the reasons are diverse. One explanation could be because token offerings require high technological knowledge, or maybe investors 'fear to miss out' investment opportunities in this new field and lastly the ICOs emerging every day are about three or four so the investors do not have the sufficient time to gather enough information about the venture. The efficiency of token market improves in the aftermarket where the trading of tokens makes investors realize of exaggerated projects and dump their tokens. Manifestation of moral hazard backfire on token issuers as soon as dispersed investors can pool information.

Going forward, given that decentralization is a key characteristic of the blockchain environment a characterization phenomenon of raising funds for entrepreneurial ventures is the geography of ICOs.

ICOs take place more frequently in countries with developed financial systems, public equity markets and advanced digital technologies (Huang, Meoli, & Vismara, The geography of initial coin offerings, 2020). It has been revealed that the presence of equity crowdfunding platform is positively associated with ICOs; in both cases an entrepreneurial firm seeks funding from a broad crowd of (mostly unprofessional) investors. But the two funding methods are very different in terms of information asymmetry involved, potential risk, investors motivations and behaviours and investors background (Block, Groh, Hornuf, & Vismara, 2020). Huang et al. analysis demonstrates also that private equity or debt markets are not positively correlated to ICOs and the countries with ICO-friendly regulations have more ICOs.

An important factor that lead to success of a token sale project is about the diverse investors' motives and behaviours in ICOs. Investors are driven by ideological, technological and financial motives and, at the moment, the most important found is the technological one, followed by financial and ideological. (Fisch, Masiak, Vismara, & Block, 2020) ICOs are not yet seen as a purely financial investment, this because professional financial investors have not yet started to adopt ICOs as investment vehicle because of their novelty and high risk associated, so the ICO world is populated with a high number of unprofessional investors. Some contributors often invest in more than one campaign and these serial investors contribute earlier; however, it does not seem they are more informed and choose better quality campaigns. Only large serial investors invest more in more successful campaigns. (Boreiko & Risteski, 2020)

On this topic it has been employed a qualitative research, relied on interviews, that focus on the rationales and motives of ICO issuers to pursue the blockchain based financing mechanism rather than using traditional funding channels, like VCs, angels or crowdfunding. (Schuckes & Gutmann, 2020)

The reasons found are several, a high percentage of issuers show strong resentment through traditional financing system blaming that the biggest part of the wealth generated goes in the pockets of VCs or angels while thanks to ICOs reaches the community of buyers making the process more democratic, it can incorporate a diverse range of stakeholders, incentivize them and democratize decision making in a start-up, therefore leading to community building and a more collective way of engaging in entrepreneurship.

Another (economical) reason is that ICO is seen as an easy and fast vehicle that does not require heavy resources, still delivering considerable funding outcomes.

Other benefits are related to the wisdom-of-crowd, in fact having a base of investors of relevant size can help entrepreneurs to understand better the response of the market on their product, receive management suggestions and improve their quality.

Lastly all the investors have a community feeling that make them very loyal and dedicated to the business idea they invested in.

In addition, following papers studied the potential determinants of adoption of the infrastructure supporting bitcoin's use as an alternative financial system.

Bitcoin adoption is empowered by perception of failure of traditional financial institutions. Specifically, Bitcoin is more popular in regions with low trust in banks and the financial system. Moreover the utilization of bitcoin could bring to countries with large percentage of population that cannot exploit bank services, but it is demonstrated that this has not happened yet, the use of bitcoin is still a complement and not a substitute to existing financial institutions (Saiedi, Brostrom, & Ruiz, 2020). But, unfortunately, bitcoin adoption could be driven by its usefulness in allowing illicit business activity, thanks to the anonymity/pseudonymity involved in the system. And in countries with strict law enforcement the adoption rises in order to circumvent this 'issue' (Saiedi, Brostrom, & Ruiz, 2020).

In conclusion, it is important to understand the influence of countries' decisions about ICOs regulation on other countries decisions.

For example, even though the blockchain system nature is decentralized, discovering the China/South Korea ban of ICOs in 2017 affected the subsequent market of ICOs in short-term and long-term. (Bellavitis, Cumming, & Vanacker, Ban, Boom, and Echo!Entrepreneurship and Initial Coin Offerings, 2020) In short-term future, a ban in a country gives incentives to other ICOs issuers to rush in other countries, this would mean that issuers have much less time to dedicate on the business reflecting negatively on the quality of the ICO itself.

The short time gives more incentives to low-quality ICOs to rush since they fear the future cost linked to ICOs will increase kicking them out of the market.

Overall, in short-term there would be an increase in low-quality token sales.

Considering long-term instead the outcomes are the opposite.

The highest-quality ICO issuers, that also are the ones who have strong ideological motives related to blockchain technology, want to keep doing their businesses exploiting this technology and so are incentivised to raise the quality, fearing future strict regulations, and signal the low-quality ones.

The lowest-quality issuers instead do not have the incentive anymore to run their token sales because of higher costs related.

Overall, in long-term there would be an increase in high-quality ICOs and less low-quality ones. (Bellavitis, Cumming, & Vanacker, Ban, Boom, and Echo!Entrepreneurship and Initial Coin Offerings, 2020) The results have been validated also by Momtaz confirming that adverse industry events, like China ban or the DAO hack, destroy token holder value, at least in the short-term.

Innovative studies have been conducted changing the topic of interest and shifting the light on CEOs emotional/mood traits. (Momtaz, CEO Emotions and Firm Valuation in Initial Coin Offerings: An Artificial Emotional Intelligence Approach, 2020)

In a context of scarcity of available information like ICOs investors rely much on the little number of information they do have, including issuers affects. The results demonstrate that CEOs showing negative affects are more prone to behave like the typical firm in its market, leading to low-risk taking decisions (in this study explained by the ICO's under-pricing level, which, in that case, is more pronounced). Also, investors, perceiving negative affects by CEOs, are less willing to pay a premium price and increase the under-pricing effect. Investors receive a negative signal.

It is shown that negative affects influence importantly the performance of an ICO while positive ones do not at all. A possible explanation is that positive affects are the 'norm', on the other hand negative affects are rarer.

1.2 ICOs INTRODUCTION

Initial Coin Offerings are a typology of raising capital exploiting cryptocurrencies¹, in which companies require certain kinds of cryptos in exchange of their company tokens².

The tokens can be set as 'Utility' or 'Security' ones, the first give advantages to the token holder in terms of usage of service offered by the company while the latter is a digital form of a stock, bond, or option.

ICOs usually have a maximum cap of capital to be raised and a minimum one that if not reached makes the project fail and the restitution of the money collected to the investors necessary.

An ICO could be a company's source of capital alternative to traditional existing ones, like IPOs³ or crowdfunding⁴. Their peculiarity that differentiates them from the others is

¹ A cryptocurrency is a digital or [virtual currency](#) that is secured by cryptography, which makes it nearly impossible to counterfeit or double-spend. Source: Investopedia.com

² Digital securities issued by the company.

³ An initial public offering (IPO) refers to the process of offering shares of a private corporation to the public in a new stock issuance. Public share issuance allows a company to raise capital from public investors. Source: Investopedia.com

⁴ Crowdfunding is the practice of funding a project or venture by raising small amounts of money from a large number of people, typically via the Internet. Source: Wikipedia

that ICOs are not exposed to strict and stringent regulations and can be a way to circumvent the necessary intervention of intermediaries like banks, venture capitalists and stock exchanges that may require greater scrutiny or some percentage of ownership.

The low-level of regulations, though, also brings diffusion of a lot of scams and frauds conveyed by the ICOs.

The ICOs, exploiting cryptocurrencies, rely on the blockchain, a transparent system continuously updated by a network of personal computers that record the exchange of cryptocurrencies by solving difficult codes puzzles; This system is not fully owned by anyone in the chain, so it is called decentralized, making extremely difficult for anyone to hack the system.

The reason of existence of the blockchain is that the people working to make every transaction secure earn a percentage of every transaction they validate and a fixed amount of cryptocurrency, they are called 'miners'.

1.3 ORIGIN OF BLOCKCHAIN

Blockchain is the technology behind cryptocurrencies. It permits not to have an intermediary to validate transaction but a chain of 'miners' who does it, leaving out the necessity to trust seller and buyer as in common online transactions. It also overcomes the double spending problem, which means spending the same digital money twice copying the digital information behind it, avoiding the necessary presence of intermediaries.

The blockchain is a specification of the more general concept of Distributed Ledger Technology (DLT), that are systems based on a distributed ledger, in which all nodes, or participants, of a network have the same copy of a database that can be read and modified independently by the individual nodes.

The fundamental characteristics that distinguish the DLT and Blockchain systems is the fact that data is stored and added to the network in form of blocks, which are then connected to each other using cryptography and hash techniques, thus forming a chain of blocks, hence the name Blockchain.

The Blockchain solutions add one further feature that is not necessarily present in Distributed Ledger systems: Transfers of value through data information.

Based on the type of network, a distinction is made between Blockchain systems:

- **Permissioned** - networks in which to access you must register and identify yourself and therefore be authorized by a central body or by the network itself.
- **Permission-less** - networks that anyone can access without permission.

In permissioned systems, the consent mechanism is simpler: when a node proposes an addition of a transaction, its validity is checked, and a majority vote is taken on whether to add it to the register.

In permission-less systems, on the other hand, the consent mechanisms are more complex (based for example on Proof of Work or Proof of Stake) to prevent a malicious person from creating numerous fictitious identities and influencing the registry modification process.

Specifically, Blockchain solutions are those in which the ledger is structured as a chain of blocks containing multiple transactions and the blocks are linked together by cryptography. Allowing to make transfers or more generally transactions of cryptocurrency. These transfers can be simple or more advanced depending on the level of programmability allowed by the platform.

1.3.1 THE BEGINNINGS

The history of blockchain technology couldn't exist without the birth of BitGold and Bitcoin.

In 1998 Nick Szabo, exponent of the Cypherpunks, that are individuals advocating widespread use of strong cryptography and privacy-enhancing technologies as a route to social and political change, built the BitGold project, a mechanism for a decentralized digital currency, on which the Proof Of Work⁵ consensus was based. In Szabo's BitGold structure, a participant would dedicate computer power to solving cryptographic puzzles, then, solved puzzles would be sent to a public registry; each solution would become part of the next challenge, creating a growing chain of new property. This mechanism later became the Bitcoin consensus mechanism where 'miners' solve difficult cryptographic problems.

1.3.2 REUSABLE PROOF OF WORK

In 2004, IT expert and cryptographic activist Harold Thomas Finney II introduced a system called RPoW, Reusable Proof Of Work, variant of the Szabo POW. The system works according to this scheme: it receives a Hashcash-based⁶, non-exchangeable or non-fungible proof of work token, and in return creates a signed token with an encryption that can be transferred from person to person.

The RPoW system finds a solution to the double spending problem by registering ownership of the tokens on a trusted server, designed to allow users from all over the world to verify their correctness and integrity in real time.

The RPoW system can be considered a first prototype and a significant first step in the history of cryptocurrencies (academy.binance.com, s.d.)

⁵ Proof of work (PoW) is a form of cryptographic zero-knowledge proof in which one party (the prover) proves to others (the verifiers) that a certain amount of computational effort has been expended for some purpose. Source: Wikipedia.com

⁶ Hashcash is a proof-of-work algorithm that requires selectable amounts of work to be done, where such proof of efficiency can be shown without any problem. Source: Wikipedia.com

1.3.3 BITCOIN NETWORK

In late 2008, a white paper introducing a decentralized peer-to-peer e-money system, called 'Bitcoin: a peer-to-peer electronic cash system' is posted on a crypto-enthusiast mailing list called 'cryptography' by a person or group using the pseudonym Satoshi Nakamoto.

The problems he highlighted in the economic system were the excessive dependence on referents, called "Trusted Third Parties", in addition to the weaknesses of the system highlighted by the erroneous and excessive use of the "model based on trust".

Nakamoto says that transactions based on these two models have meant that financial institutions have kept management costs high, consequently putting them on users and the problem becomes much more concrete especially in small transactions. (Nakamoto, 2008) The system is based on the proof of work Hashcash algorithm, but instead of using trusted hardware such as RPoW, Bitcoin double-spending protection is guaranteed by a decentralized peer-to-peer protocol for monitoring and verifying transactions.

The main idea behind the proof-of-work system is to make it expensive for a single user or a group of users to rewrite the history of transactions once it has been accepted as definite; the aim is to prevent malicious users from double spending their Bitcoins. (Matthias Lischke, 2016) The solution that Nakamoto proposed is the use of a server that takes the hash of a block of items, timestamps it, and widely publishes the hash. (Nakamoto, 2008) A hash transforms a line of text into a series of letters and numbers, then the proof of work uses hash algorithms to find a specific value. In order to make every line more difficult to manipulate, final digits are added at the end called nonce, which stands for 'number only used once'.

If the hash meets the requirements set forth in the target, then the block is added to the blockchain.

There is a predetermined target difficulty that is updated for every 2016 blocks that have been generated. This ensures that the time it takes to generate one block is on average about 10 min. The block is only accepted by users if all transactions in it are valid and

the Bitcoins have not been spent previously. Users show their acceptance by using the newly found hash in the “previous hash” section of the next block they attempt to generate; thus, adding a new block to the chain. This chain is called the block chain or transaction log and contains the entire history of all transactions that have been carried out in the network. The generation of blocks by users is called mining and is achieved through providing a certain amount of computation power to the network to solve the proof-of-work problem. The expending of computation power is rewarded when generating a block. There is competition to get the reward, and the more computation power a user or group possesses the better the chance to get it. The reward is predetermined and started at 50 BTC. It will decrease by half every 210,000 blocks. In that way new Bitcoins are introduced to the network. This procedure will continue until the predetermined final amount of 21 million Bitcoins is in circulation, around the year 2140. This predetermination tries to overcome the problem of inflation that an ever-ending cryptocurrency could easily generate. Below it is explained the Bitcoin economy in **Errore. L'origine riferimento non è stata trovata.**

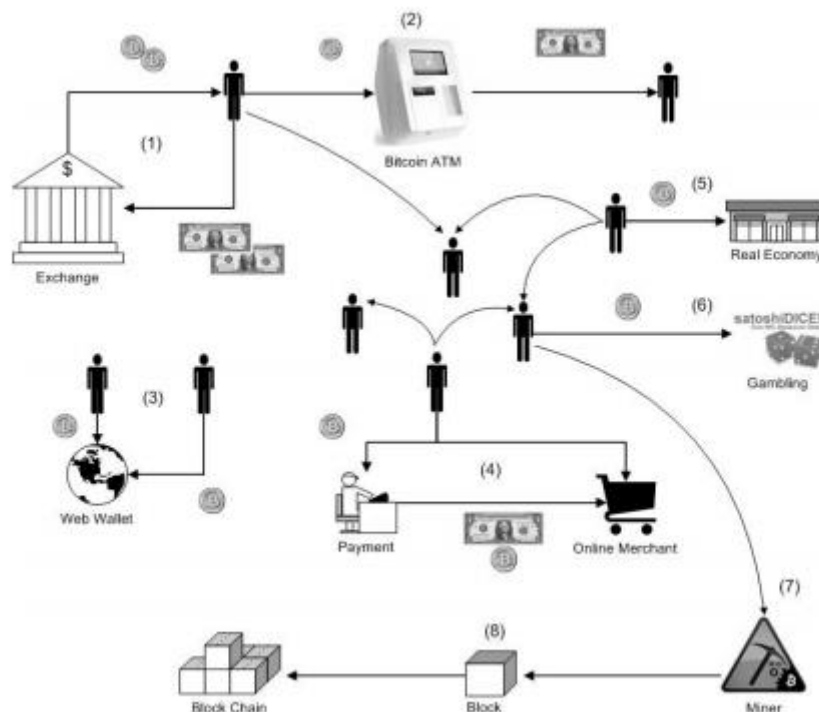


Figure 1 - Bitcoin Economy (Matthias Lischke, 2016)

1.3.4 ETHEREUM

In 2013, Vitalik Buterin, programmer and co-founder of Bitcoin Magazine, declared that Bitcoin needed a scripting language for building decentralized applications. Failing to convince the community, Vitalik begins to develop a new distributed computing platform based on the blockchain, Ethereum, which features a scripting feature, called a smart contract; Ethereum platform is aiming to improve the concepts brought by Satoshi Nakamoto and is using currency inside its network -Ether. (Buterin, 2013)

Ethereum, like Bitcoin and other cryptocurrencies, allows to transfer digital money. However, it is even stronger because gives the possibility to deploy personal codes and interact with applications created by other users; also, diverse sorts of programs can be launched on Ethereum.

The main idea behind Ethereum is that developers can create and launch code which runs across a distributed network instead of existing on a centralized server. This means that, in theory, these applications cannot be shut down or censored so anyone, anywhere, can launch applications that cannot be taken offline. These applications can set conditions on how value is transferred, the programs that make up applications are called smart contracts. In most cases, they can be set to operate without human intervention, giving a certain return to the user who sends money if certain conditions are met. Ethereum is currently the second-most-valuable cryptocurrency. (academy.binance.com, s.d.)

1.4 INITIAL COIN OFFERINGS

As soon as cryptocurrencies were discovered as a new means to clear payments, the opportunity immediately arose to use them to raise money through ICOs.

ICOs are a very recent phenomenon, in fact, the first ICO was conducted in July 2013 by Mastercoin, a digital currency built on Bitcoin's blockchain. Since then, hundreds of ICOs

have followed and both the number of ICOs and the amount of capital raised have exploded since 2017.

In an initial coin offering (ICO), new ventures raise capital by selling tokens to a crowd of investors; Often, this token is a cryptocurrency, a digital form of value exchange based on the distributed ledger technology.

A token corresponds to a unit of value issued by a venture and could be of different typologies. Usually, tokens can either provide a utility or function as securities. Hence, “utility tokens” are generally distinguished from “security tokens” even though no legally binding classification of token types exists (Sameeh, 2018)

Utility tokens, that are the majority of ICO tokens, provide to the investor the possibility to exploit a certain kind of utility related to the business. There are other types of tokens that provide utility, such as reputation or reward tokens, giving in exchange gadgets and objects to the investors (Sameeh, 2018). Security tokens, instead, are a digital form of traditional securities like stocks and bonds. Some ventures have created their own cryptocurrencies by issuing their tokens and making trade them on their own platform and ecosystem.

Typically ventures funding their business through ICOs make public a document called “whitepaper” that usually includes the business model of the project, the team, the IT protocols, the blockchain adopted and other meaningful information.

Campaigns usually accept Bitcoin and ether as funding vehicle but usually accept also others forms like traditional fiat money.

Distributed ledger technologies allow the design of complex claim structures called “smart contracts” that return a certain event if certain circumstances are met; For example, money could be released only on occurrence of certain events. Alternatively, particular rights could be attributed to early contributors, such as the ownership of a token that can be spent to obtain a service or be traded, or polls might be organized

among pledgers. (Adhami, Giudici, & Martinazzi, Why do businesses go crypto? An empirical analysis of initial coin, 2018)

It is important to note that the venture-issued tokens often do not have a tangible product or a functioning service, instead, they entitle the holder to future participation in a project that uses the tokens in its respective utility-providing function (Kaal, 2018). Security tokens' value is given by the ownership of certain financial rights like shares of ownership or 'coupon payments' like traditional bonds. They can be equity tokens, implying ownership or control and emulating traditional stocks, or they may emulate all kinds of traditional securities, entitling the investor to get dividends, or other financial benefits. (Sameeh, 2018) Most tokens can be traded in a secondary market after the conclusion of the ICO. Although this is very rare because of the very high listing costs imposed by the listing platforms and the long procedure of due diligence to ensure that the ICO is not a scam, possibly damaging the exchange platform reputation.

1.5 COMPARISON OF ICOs TO CONVENTIONAL FINANCING METHODS

This section gives a brief comparison of ICOs to conventional financing methods such as reward and equity crowdfunding, venture capital, and initial public offerings (IPOs). Differently from ICOs traditional funding mechanisms are related to specific stages of the business development. Crowdfunding, for example, is used to fund early stages, venture capital covers all stages until a firm goes public, and IPOs are used to raise big amounts of capital through an exit event. ICOs, on the other hand, can theoretically be employed during all funding stages.

While ICOs are suitable to attract all different kinds of investors (from early adopters to institutional investors), conventional financing methods usually attract specific types of investors. Reward and equity crowdfunding attract early adopters and angel investors, respectively. Venture capital and IPOs are traditionally more attractive to sophisticated and experienced investors. Going forward, the motivation of investors differs among these financing methods. Venture capitalists and IPO investors are more likely to be driven by financial motives, while ICO and crowdfunding investors are often driven more

by non-financial motives (altruism, product interests, feedback provision, etc.) than financial ones.

The peculiar characteristic that drives ICOs popularity is the very low, almost zero, transaction costs and low level of regulation, but, at the same time, enables new ventures to possibly raise substantial funding comparable to costly and highly regulated venture capital transactions or IPOs.

Another reason that can push investors in investing is the after-market liquidity. Although very low percentage of tokens get listed on a token exchange platform. Another relevant characteristic of ICOs is that tokens do not convey voting rights to investors, this lack of corporate governance could be a potential issue in the development of the project. Finally, the exit method could be one of the most important advantages of ICOs. In crowdfunding or venture capital exits do not happen before a certain maturity stage level and surely not realizable in the short run as a potential acquirer needs to be identified or an IPO needs to be prepared. On the other hand, ICOs provide the quickest exit by delegating the future development of a platform to a decentralized network of developers and supporters often before a product prototype or service is developed. (Momtaz, Entrepreneurial Finance and Moral Hazard: Evidence from Token, 2020)

1.6 PHASES OF ICO

There are different stages that an ICO may follow.

- **PRE-ICO**

The pre-ICO is optional but many companies follow this path because its important in terms of visibility of their projects and raising money to cover the costs of the actual ICOs. The main subject of the pre-ICO is the disclosure of the white paper where the digital firm explains the business and the technology to the public. The entrepreneurial firm normally announces an advisory board (to signal the quality of the ICO project) and hires experts for conducting the ICO campaign, in exchange for either capital or a

considerable number of tokens. Smaller firms lacking finance and resources tend to purchase external expertise to indicate their quality commitment to potential investors and to differentiate their ICO from other campaigns. Another possible use of this phase is to assess the market acceptance of the token. Usually, the tokens are sold at a discount in this phase. Pre-ICOs can also be seen as instruments to assess a fair value to the price of tokens issued. (Masiak & Block, Initial coin offerings (ICOs): market cycles and relationship, 2020)

- **ICO**

The mechanics of the actual ICO are easy. The project creates an address to which the funds will be sent. The company discloses the currencies that project accepts as payment for its token. Investors send then funds, in the accepted currencies, to the address and receive the equivalent number of tokens. To promote the ICO campaign, the venture usually provides bonus schemes for ICO investors.

- **LISTING**

A venture that has conducted an ICO needs fiat, or other cryptos, to make investments and develop the product or service. Trading exchanges offer the opportunity to change tokens to fiat or other cryptocurrencies. To trade tokens, ICOs have to be listed on a trading exchange, which typically takes time, from several weeks to even six months as an average. In addition to the ICO firms, investors aim to increase the value of the tokens that they receive and sell them if their value rises considerably. (Masiak & Block, Initial coin offerings (ICOs): market cycles and relationship, 2020) The post-ICO phase is a critical one. The characteristics that a token must have to be exchanged on a platform are opaque, and the biggest share of ICOs fail in this important achievement that could potentially bring a major part of liquidity to stakeholders.

1.7 ICOS REGULATIONS

ICO regulation is a crucial determinant in the evolution of the ICO market. Regulation influences the level of ICO activity as well as its geographical location (Bellavitis, Ban,

Boom, and Echo! Entrepreneurship and initial coin offerings, 2020). The ICO sector is characterized by a quite low degree of regulation (Bellavitis, Cumming, & Vanacker, Ban, Boom, and Echo! Entrepreneurship and Initial Coin Offerings, 2020). The absence of regulation potentially enables ventures to raise large amounts of funding with lower effort and costs (Fisch, Initial coin offerings (ICOs) to finance new ventures, 2019). But this also potentially increases investment risk because of opportunistic or malevolent behaviour. Indeed, the ICO industry has a high prevalence of fraudulent events resulting from lack of regulation (Momtaz, Initial coin offerings, 2020). For example, “exit scams”, in which the venture team disappears after raising funds, seem relatively common (Fisch, Initial coin offerings (ICOs) to finance new ventures, 2019). The lack of regulation makes prosecution of such scams also very difficult. High investment risk and fraudulent activity have attracted scrutiny from regulators having different reactions. Regulators have sought to determine how they can prevent the use of the coins for money laundering initiatives or to fund illegal activity, like terrorism. One of the other significant issues that regulators have run into is a lack of universal compliance.

Some countries have taken a hard approach toward cryptocurrencies while others, like Switzerland, have been much more progressive in their monitoring of digital currencies. Having a firm understanding of how regulations vary from country to country can help an organization when establishing its ICO.

Below, it is provided a breakdown of some of the regulations currently in place. These regulations are always subject to change and are likely to do so in the future as governments seek to control and monitor the currencies.

- **COUNTRIES IN WHICH CRYPTOCURRENCIES ARE PERMITTED**

The following countries have permitted the use of cryptocurrencies without any exceptions or regulations:

Australia, Belarus, Belgium, Bosnia and Herzegovina, Brazil, Bulgaria, Chile, Colombia, Croatia, Cyprus, Czech Republic, Finland, Greece, Iceland, Italy, Lebanon, Lithuania,

Malta, Norway, Poland, Romania, Saudi Arabia, Slovakia, Slovenia, Trinidad and Tobago, Turkey, Ukraine.

- **COUNTRIES IN WHICH CRYPTOCURRENCIES ARE BANNED**

The following countries have not permitted the use of digital currencies and have not provided any exceptions:

Bangladesh, Bolivia, China, Ecuador, Morocco, Nepal, Pakistan.

- **COUNTRIES IN WHICH CRYPTOCURRENCIES ARE PERMITTED WITH STIPULATIONS**

Many countries permit the use of digital currencies but have taken efforts to regulate or further define them. For instance, Argentina allows initial coin offerings. When doing so, they do not recognize altcoins as legal tender, even though they perceive them as money. In Cambodia, companies are permitted to introduce initial coin offerings but have discouraged their people from involving themselves with altcoins.

- The **Canadian Securities Administrators** permit ICOs in their country. However, they have recently announced that they plan to regulate ICOS and altcoins on a case-by-case basis, classifying them as securities. Additionally, Canada has also placed a regulatory sandbox in place to help monitor projects that might fall outside of its standard regulatory scheme. Lastly, the country deems altcoins to be intangible assets. All dealers must register themselves as a money-service company. Canadian banks have temporarily placed a hold on the purchase of altcoins until the issue is further clarified.
- **Estonia** not only allows ICOs, but it has even considered starts its own for fundraising purposes. However, there is a contentious debate on how the Eurozone rule on nation states would impact the fundraising campaign, so its status is up in the air.
- **The European Union** permits ICOs but requires strict regulation. All ICOs must abide by Know Your Customer and Anti-Money Laundering regulations. The

European Securities and Markets Authority announced that they view ICOs as extraordinarily volatile and risky to investors.

- **France** permits ICOs but regulates them. So too does Germany, whose Federal Financial Supervisory Authority issued a strong warning about ICO investments. The group said that the consumer takes on all risks associated with ICOs because there is a noticeable lack of regulatory and transparency requirements. They also warned consumers that there is no way to ensure their data is safe.
- **Gibraltar** permits ICOs but plans to pass laws that codify legal protections for altcoins. Hong Kong is similar in that they allow ICOs but subject them to regulation, stating that some altcoins must be viewed as securities. ICO use is permitted in Hungary. It is also allowed in Indonesia, although there it is only done so as a commodity. Iran sanctions ICOs but has announced future regulation.
- **India** allows ICOs although places them under heavy scrutiny. Their government has strongly discouraged the use of altcoins. Additionally, the Reserve Bank of India does not allow altcoins in its banking system. Israel permits altcoins but subjects them to a capital gains tax of 25 percent. Additionally, miners must also pay corporate income tax on earnings. The Isle of Man permits ICOs but is looking to solidify its regulations moving forward.
- **Jamaica** is progressive when it comes to ICOs, noting that they provide an opportunity for growth. Jordan, on the other hand, does not allow its financial institutions to use altcoins. Kyrgyzstan does not allow the use of altcoins as currency. Luxembourg, on the other hand, supports the use of Bit License of business. Malaysia currently supports ICOs but could subject them to further regulation.
- **Mexico** supports ICOs and views them as virtual assets. Namibia permits ICOs but forbids the use of altcoins for payment. Nicaragua allows ICOs and has not released an official stance. Nigeria has banned altcoins but not virtual currencies. The Philippines recognizes some cryptocurrencies as a form of remittance payment but has admitted that identity verification is needed in the future.

- **Russia** allows ICOs but requires altcoin miners to register and pay taxes on their earnings. It appears the country may soon rule altcoins illegal. Singapore recently provided a guide to its citizens on Digital Token Offerings, viewing them as capital market products. South Africa, similarly, sees altcoins as intangible assets.
- **South Korea** has not banned altcoins but seeks to strike down malicious ICOs. Individuals cannot trade altcoin futures or derivatives. Sweden is currently in an intense legal battle over ICOs as they are now subject to regulation by the Swedish Financial Supervisory Authority. Switzerland could possibly soon regulate ICOs, even though previous attempts have failed previously.
- **Taiwan** permits ICOs although advised banks not to accept altcoins as a form of currency. Those in Thailand are not allowed to invest in altcoins or exchange coins for currency, although the government has yet to ban ICOs entirely. In Vietnam, individuals can use the coins as currency with no laws on trading. And in Zimbabwe, the government has yet to take a stance on altcoins.
- The **United Kingdom** recently issued a warning to investors about ICOs, calling them “experimental.” In the United States, ICOs are subject to burdensome regulations. However, the laws vary from state to state. If the ICO trades securities, the organization must first register with the Securities and Exchange Commission.

(AU10TIX, s.d.)

1.8 ICOs GEOGRAPHY

The most successful scene of ICOs is in:

- **USA:** In the early stages of the industry, the US had a large share of the market, comprising 30% of ICOs. However, this early dominance has disappeared, with the US responsible for only 1 ICO in Q2 2020. (Bellavitis, Cumming, & Vanacker, Ban, Boom, and Echo! Entrepreneurship and Initial Coin Offerings, 2020)

- **Estonia:** Estonians built an efficient, secure and transparent Internet ecosystem. The country tops the list of number of ICOs per inhabitant.
- **Cyprus:** The country has never lacked talents: several educational establishments graduate specialists in economy and finance. Today, Cyprus has become a hub for international ICOs.
- **Slovenia:** The small country of Slovenia positioned itself as the leading destination of Blockchain technology in the European Union (EU). The government is also studying the potential applications of the technology in public administration while many tech start-ups have filled the news with amazing ICO successes.
- **Switzerland:** The Swiss state has accounted for four of the 10 largest ICOs with investors attracted by the country's business-friendly regulations and digital expertise. But during 2020 the country, very important and popular at the beginning of the spreading of ICOs, did not keep up with its historical trend. (Bellavitis, Cumming, & Vanacker, Ban, Boom, and Echo! Entrepreneurship and Initial Coin Offerings, 2020)
- **Russia:** Russia has valued a lot the possibilities of blockchain & cryptocurrencies. The first guarantee system for Initial Coin Offering (ICO) investments has been launched in Russia in 2018. The system, has been developed by Globex Bank, which is a subsidiary of state-owned Vnesheconombank (VEB) bank, together with the Russian Association of Cryptocurrency and Blockchain (RACIB), and the CrowdHub platform. Sberbank, one of the largest banks in Russia, has announced a blockchain laboratory to develop and test blockchain-based solutions. Russia as Switzerland was popular in early stage of the game but lost popularity.
- **Panama:** Panama is an emerging player on the blockchain stage. The Blockchain Embassy Panama has opened its doors and intends to serve the general public, connoisseurs, enthusiasts, developers and entrepreneurs in the blockchain ecosystem.

- **Singapore:** Initial Coin Offerings has seen tremendous growth and has increasingly seen mainstream adoption by start-ups as a method to raise fund in Singapore. In 2018 year-alone Singapore witnessed more than 200 ICOs with over \$3.3 billion USD raised.
- **UK:** 2017 was the year that ICOs diffusion exploded in UK. Start-ups in Great Britain raised over 3 billion USD in 2018. (Unlock, s.d.)

If we consider data from Icobench.com the diffusion of ICOs by the number of them sees the USA at first place with 717 events, Singapore at second place with 587 events, UK at third place with 514 events, Russia at fourth place with 328 events and Estonia at fifth place with 300 events. Looking at funds raised by ICOs also British Virgin Islands and Switzerland enter the top 5 list of Countries.

This tells us that the phenomenon of ICOs is much more spread in some countries than others. There are several reasons behind this fact and have been explained by Huang, Meoli and Vismara (Huang, The geography of initial coin offerings, 2019). Firstly, ICOs are more diffused in countries with a well-developed financial system and where the presence of digital ventures is stronger, and this happens where digitalization is more present. Following, high ICT investment rate is another important factor that makes the spreading heavier. Finally, the stronger intention of regulating ICOs, rather than banning or doing nothing, by countries might positively affect their diffusion.

1.9 ICOs TRENDS AND CHALLENGES

Companies launching ICOs are becoming more professional. Starting from the second half of 2018 over 70% of ICOs set both a soft cap and a hard cap; in 2017 this figure was only around 30-40%. A possible explanation could be that companies participating in ICOs are becoming more and more transparent with their financial goals. Average ICO duration has been increasing steadily. This mark has gone much higher since November

2017, when the average ICO duration was around 1 month. This could indicate that now Initial Coin Offerings need more time to achieve their goals due to the general decline of the crypto market and less active investors. On the other hand, nuanced projects need more time to adequately share their ideas with the community, and investment now seems to become more measured and less hasty. Moreover, a longer average lifespan requires steady funding, something that might be inaccessible to smaller, risky projects. The average success rate has significantly increased. In 2017 only a quarter of all projects were successful, their number has been increasing since September 2018, and in 2019 almost two thirds of ICOs reached their targets. It can be said that this is occurring because of the industry maturation factor that has produced more experienced teams and more valuable expertise. (IcoBench, s.d.)

But in 2020 there was the decline of ICOs, the number of ICOs dropped off to only 87 ICOs in Q1 2020. The drop in ICOs was accompanied by an even more dramatic decline in the amount of money raised, down to only 31m USD in Q1 2020. This recent decline could also partly be due to an increased occurrence of IEOs⁷ and STOs⁸. (Bellavitis, Ban, Boom, and Echo! Entrepreneurship and initial coin offerings, 2020) Considering the Bellavitis study another possible reason causing the decline is the bigger increasing presence of regulations.

ICOs are characterized also by quite severe disadvantages in their broad recognition as stable and trustworthy investment mechanisms. One of the major issues is the low level of regulation that provides soil for scams, warning potential investors. A second relevant problem is the dominance of information asymmetry, in the market, that linked to the absence of mediating institutions could lead to a so-called “market of lemons”, inducing

⁷ An Initial Exchange Offering, as its name suggests, is conducted on the platform of a cryptocurrency exchange. Contrary to Initial Coin Offerings, an IEO is administered by a crypto exchange on behalf of the startup that seeks to raise funds with its newly issued tokens. Source: Cryptopato.com

⁸ A security token offering (STO) / tokenized IPO is a type of public offering in which tokenized digital securities, known as security tokens, are sold in cryptocurrency exchanges, or security token exchanges. Tokens can be used to trade real financial assets such as equities and fixed income, and use a blockchain virtual ledger system to store and validate token transactions. Source: Wikipedia.com

the issuers of the project to provide exaggerated information and have moral hazard behaviours. (Momtaz, Entrepreneurial Finance and Moral Hazard: Evidence from Token, 2020)

2 CHAPTER 2

2.1 DATABASE METHODOLOGY

A relevant part of the work has been the determination of considerable variables for the goal of understanding the best success factors of an ICO. It has also been important to recognize the best sources from which capture these information.

All the variables found were added to an already existing database made by Sellitto, previous author of thesis about funding success factors that characterize ICOs, and after this all new ICOs, not included in the database, were added in it with all variables related.

In doing so the variables took in consideration for the future analysis were the ones not yet studied by the literature but considered as important to analyse in the future research.

After the decision of the relevant variables, it has been conducted a subsequent analysis on which were the best sources to extract these data, also for this purpose the approach has been to check on the literature the sources used and considered more reliable.

2.2 LIST OF SOURCES

This section of the chapter will highlight and describe the different sources examined, with the aim of obtaining the data required for the population of the database.

- ICOBench.com

This website is the internet source with most numerous amounts of information on ICOs.

ICOBench.com comprises a total of 5781 projects developed from 2015 to 2020.

On this website it is possible to find: a description regarding the project, the beginning and ending dates of the crowdfunding phase (and, frequently but not always, also data about the pre-sale phase), information about the country issuing the project, the

restricted areas (if any), the availability of an MVP or prototype, bonus, or bounty program and Pre-ICO phase with related dates about duration of it.

It can also be seen the composition of the team of the project, the token issued by the company, information about the contract type and several other financial information such as price of the token, how many tokens have been issued and sold, currencies accepted, soft and hard cap. Furthermore, the website made accessible other information, such as the amount of money that have been raised by the company, a project rating (from 0, lowest, to 5, highest) given by ICOBench.com team, a link to the company's website and also a section where it is possible to download the technical whitepaper.

-CoinMarketCap

This source has been used to extract market prices of listed tokens of the various projects, also prices of Bitcoin and Ethereum, in order to calculate variables useful for the work like "Under-pricing" or "token holders returns" for example. The website supplies an historical data list of prices of a specific token or cryptocurrency from its listing until the present moment in time. Other features are a link to the company website and a redirection link to Github.com to have a look at the source code of the project. Lastly, other interesting information provided are a list of articles related to the project in consideration, to understand the social visibility, a rating that shows the riskiness of the project and finally a list of market pairs.

-Github

GitHub is a web and cloud-based service that helps developers store and manage their code and track and control changes. It offers a cloud-based Git repository hosting service. In particular, Git is a distributed versioning system, which means that the entire base code and history are available on each developer's computer, which allow them to easily create branching and mergers. With branching, a developer duplicates some of the source code (called a repository). The developer can then securely make changes to that part of the code without affecting the rest of the project. Then, once the developer

makes his code part work properly, he can merge that code into the main source code and make it official (www.Kinsta.com, s.d.).

This website has been used to understand if the project taken into consideration had a source code available for the public or not.

-Twitter

Twitter is a free microblogging platform that allows users to share messages up to 140 characters long (called tweets), it is commonly used to share emotions and passions, but also to professionals to promote their activities, public figures to have direct contact with ordinary people. For the purpose of this work the website has been used to collect numbers about the visibility of projects on social media (proxied by Twitter) finding the amounts of followers of the project account and the number of tweets shared.

2.3 METHODOLOGY

After having analysed all the sources taken into consideration, the database has been populated manually line by line, because of the absence of the possibility to automatically download the data of interest.

For all the new ICOs added, from December 2019 to December 2020, not all information was available on the internet since the primary source, Coinschedule.com, from which the data were collected for constructing the first database has been closed in the meantime.

When previous author was populating the database with information from multiple sources, it happened to find uncoherent information for the same parameters according to the two different platform consulted. This happened especially for the field referring to the ending date of the token sales period and for the one referring to the amount raised by the company. To clarify as much as possible in the database all these information, 2 parameters about each of these two voices have been inserted, specifying in the field's name the source from which it has been taken. In projects where the information from Coinschedule.com and those from ICObench.com coincided or

were similar, the field corresponding to the second source has been left blank, otherwise both the parameters corresponding to the two sources have been completed by the author. All the values extracted from Coinschedule.com have been inserted in the database, whether it was identical, similar or different to the same parameter taken from ICObench.com: this is due to the fact that it has been downloaded automatically from the web platform. It must also be underlined that some of the ICOs taken from Coinschedule.com were replicated because they have been registered on multiple records, according to the different crowdfunding phases (i.e., two or more different records for pre-sale, private sale and final sale phases), so the records not directly referred to the final period of the ICO were removed. After previous author detected and eliminated ICOs according to these criteria, from the original 1671 projects downloaded the database was then populated by 1475 of them. Out of those 1475 left, 1354 projects disposed of the technical whitepaper, which have been downloaded. The rest of those ICOs, for which the whitepaper was not available on the Internet, have however been included into the dataset in consideration, of which 1475 have been taken from Coinschedule.com before its closure from middle 2016 to the end of 2019, the latter 65 have been take from ICObench.com and added manually, from the end of 2019 up to now. Each ICO has 39 fields that describe it.

In this section of the chapter it will be described each of the variables available in the database's structure. The fields will be divided according to the sources from which they have been taken, which may be Coinschedule (from previous author) ICObench.com CoinMarketCap.com Github or Twitter.

2.3.1 INFORMATION FROM COINSCHEDULE

The information, regarding the 1475 ICOs that were initially taken from the website, were about:

- **Name:** the name decided by the company of the issued project, it may happen that does not match with the name of the company issuing it.

- **Category:** it explains the category in which the company issuing the ICO is going to operate when the completion of it will be achieved.
- **Ending date (Coinschedule):** it is related to the ending of the token main sales stage. Inside the database have been included projects whose ending date is comprised between the beginning of 2016 and the end of 2019.
- **Amount raised (Coinschedule):** it represents the total amount of money raised by the company during the funding periods (including all the different phases). The currency indicated for this field is United States Dollar (USD) for all the ICOs.
- **Percentage of Hard Cap:** this parameter refers to the total amount of money that has been raised by the company with respect to the Hard Cap. An ICO Hard Cap refers to the maximum capital amount that it is aiming to accumulate. Since it is an estimate it may happen that the actual amount raised exceeds the hard cap, giving as result a percentage higher than 100%. In the whole dataset 49 ICOs reported a percentage higher than 100%.

“Another parameter, not directly taken from Coinschedule.com, that is included in the final database refers to the univocal code **ICO_ID** that has been used to indicate the projects. This field of the database can be considered as a primary key for the database development. The parameter, that has been set by the author, is a non-repeatable number that has been assigned to each project according to the order in which Coinschedule.com listed them, when downloading the information. To each of the 1671 ICOs that were originally downloaded was given a univocal code from 1 to 1671: eventually, as explained beforehand, some ICOs has been removed with the intention

of preventing information from being duplicated, but the univocal codes for the ICOs still have been left as they were listed, going from 1 to 1671.

When downloading whitepapers, the files were saved in a dedicated folder, in which all the whitepapers were collected together and named according to their corresponding univocal code in the database.” (Sellitto)

2.3.2 INFORMATION FROM ICOBENCH

Data from ICObench have been manually added.

During database’s population of these information by previous author, when an ICO from Coinschedule.com was not found on ICOBench.com, the project was removed from the dataset.

The parameters taken from ICObench are:

- **Description:** this field consists of a brief description about the business model of the company and the goal the project wants to achieve.
- **Starting date:** It tells the reader when the funding process has begun. This information was not available on Coinschedule and also for around the 8% of the ICOs included in the dataset.
- **Ending date (ICOBench):** Since this field of the database was mainly populated by Coinschedule given the automatic download of its data the only situations in which the data has been taken from Icobench was when the information from Coinschedule.com was uncoherent or different with the ones available on ICOBench.com. This phenomenon happened for around the 13% of the ICOs in the database. As said before for all the 2020 ICOs the ending date was taken solely from this website.

- **Token:** this field tells the name used to indicate the token issued to investors. In this case there was never problem of different projects having the same name because the token name is strictly univocal. Only 3% of total number of ICOs was not specified any token name.
 - **Platform:** this field represents the platform on which the token was developed. There were 75 different platform used. 1351 ICOs have been developed on Ethereum (88%). The second most used platform is waves for 27 ICOs. While Bitcoin only used for 8 projects (3.6%).
 - **Type of token:** A token can be a utility or a security or payment token, the difference is explained in the first chapter. The 98% of totality are utility tokens. This wide spreading is probably related to the fact of much lower presence of regulations on those type of tokens.
- Type of contract:** it refers to the smart contract model adopted by the company. The contract type provides a standard set of rules that the token itself should adhere to. In the database there are 60 different types but ERC20 is the dominant one with the 92% of totality. ERC-20 is the universal language that all tokens on the Ethereum network use. It allows one token to be traded with another.
- **Country:** it represents the country where the ICO has been issued. The database includes 100 different countries, the most popular ones are Singapore (193 ICOs), USA (190 ICOs), UK (135), Switzerland (94 ICOs), Russia (92 ICOs) and Estonia (83 ICOs). These countries together account to the 53% of totality.
 - **Restricted areas:** this field contains a list of areas in which is not possible for investors to allocate their money into the selected projects. 55 ICOs had declared that no specific restricted areas was set, for about 940 projects no information were available and for around 500 at least one restricted area was specified.

- **KYC/Whitelist:** since sometimes is required a further control in order to be authorized to buy tokens this field contains if there is or not this requirement. Whitelist registration is like the registration on e-commerce websites for mobile sales, the company's entrepreneurs are capable to track who is receiving their tokens and investing in their ICO, while the KYC stands for Know Your Customer which entails the publication of personal information and documents to the company. In this way, the company's entrepreneurs are capable to track who is receiving their tokens and investing in their ICO. 266 ICOs have the KYC protocol, 49 the Whitelist and 365 both of them.
- **Price in ICO:** this field refers to the price of the token during the capital funding period. This information is only related to the price during the main phase of the ICO and not pre-sale phase. 1443 ICOs had their related price disclosed which is the 94%. For about 1400 ICOs the price was recorded in USD, for around 90 was not available and for the rest was recorded in ETH. The price of pre-sale phases was not included.
- **Tokens for sale:** this information gives the number of tokens issued by the company in order to develop its project. It is referred to all different phases of the ICO in consideration. Almost 22% of totality have not this information disclosed.
- **Sold tokens:** tokens that have been actually sold by the company throughout the ICOs' phases. This information could not be any higher than the value referred to tokens for sale and was available only for 114 ICOs (7.4%).
- **Minimum investment:** it is the minimum amount of money that a potential investor could invest on a specific project. The amounts could be found in USD, ETH, or other currencies. Only 35% of totality has this data available on the website.

- **Currency accepted:** this indicates the list of currencies that the company is willing to accept in order to finance its project. The most popular ones are USD, ETH and BTC. For 1343 ICOs the disclosure of this data was present.
- **Distributed in ICO:** it is the amount of raised money that the company will reinvest in the project. It is a percentage value and cannot exceed the 100%. 72% of projects has this information disclosed. Sometimes this field has been filled not with a single percentage but with a range of percentages.
- **Soft Cap:** Is the level of money raised that the company assumes it would be easily achieved, if not the project fails and cannot go any further. If this happens the company will return all the investments. The setting of the Soft Cap is not mandatory but could be a good practice for ICO issuers since gives a feeling to the investor of the possibility to achieve the minimum goal. The information related are mostly reported in USD or ETH, but several other currencies were present. The 52% of ICOs had a soft cap.
- **Hard Cap:** This instead is the maximum amount of capital that can be raised. This is as before not mandatory to be set and is reported mainly in USD and ETH but also here several other currencies have been used. The 78% of dataset had decided to set it.
- **Amount Raised (ICOBench):** similarly, to the field "Ending Date (ICOBench)", this field has been populated just in case the information from ICOBench.com was different from the one of Coinschedule.com. The information from the two websites differed for this field in 35% of the cases, but in many of them the order of magnitude of their differences was very little, being of some thousands of USD or less. However, there were some cases in which the differences for this field were relevant, having completely different total amounts raised. While for Coinschedule.com data about the amount raised by the project were always available, for several ICOs on ICOBench.com was not specified any information about this field.

- **ICOBench Rating:** refers to an overall rating to the project's reliability given directly by the ICOBench.com website. The rating can go from 0 to 5 and entails also other experts' ratings, all the ICOs in the database had one.
- **MVP/Prototype:** This field tells if the project has developed a Minimum Viable Product or a Prototype. 299 ICOs had developed it (the 19 % of totality).
- **Pre-Ico dummy:** this field explained the presence or absence of a pre-ICO phase. If an ICO had a pre-ICO phase this field contained a 1. A pre-ICO is not mandatory but could be chosen to be done since gives visibility and pre-funding to the project. 599 ICOs had one, which is the 39% of totality.
- **Bonus Dummy:** this field explained the presence or absence of a bonus scheme, it could be discounts or other kind of promotions. If an ICO had a Bonus scheme this field contained a 1. Also this is not mandatory for any ICO. 521 ICOs had one, which is the 34% of totality.
- **Bounty Dummy:** this field explained the presence or absence of a bounty program. If an ICO had a Bounty program this field contained a 1. In a bounty ICO, a token-sale is hold existing purely for the purpose of testing a specific design. By selling tokens that can be redeemed for a guaranteed return, it is created an incentive for people to obtain as many tokens as possible. (www.medium.com, s.d.) 428 ICOs had one, which is the 28% of totality.
- **Team disclosure:** this field contains another dummy variable telling if the project's team was disclosed. 1166 Projects had their team public, the 76% of totality.

- **Number of days between the pre-ICO end date and the ICO:** this variable takes into account the number of days passed between the pre-ICO end date and the start date of the ICO. 635 ICOs had not this information, the possible reasons could be the absence of a pre-ICO phase or the absence of this particular data.
- **Dummy variable that equals 1 if the pre-ICO offered a bonus:** This variable tells if the pre-ICO offered a bonus or not. 315 projects have been signalled as having it.
- **Number of days the pre-ICO lasted:** This field gives the number of days of duration of the pre-ICO phase. 531 ICOs had this data available.
- **Dummy variable that equals 1 if the pre-ICO had a hard cap:** This variable tells if the pre-ICO phase had an hard cap, only 1 ICO in the whole dataset has this characteristic recorded.
- **Jurisdiction:** Jurisdiction is the last variable taken from ICObench. Equals to 1 if the ICO in consideration has chosen Swiss, Singapore or Estonia as jurisdiction or country. The reason why these countries have been chosen is their very permissive regulations on ICO markets, and so it was important to understand if any relation with this and the success of an ICO was present.

2.3.3 INFORMATION FROM COINMARKETCAP.COM

Data from Coinmarketcap.com have been manually added since several calculations were required to obtain the wanted results of the variables.

The parameters taken from Coinmarketcap.com are:

- **Currency:** Dummy if the token can be used as currency.

- **Under-pricing:** This variable tells how much the token price at the end of first day of trading is above the initial price set. 109 ICOs had a positive under-pricing value.
- **Market sentiment proxied by the price change in BTC:** This value gives an information about how the market is trending taking into consideration the change in Bitcoin price in the period of time of the ICO considered. The number of ICOs issued in a positive market period (hot market) are 465. Examining Bitcoin and Ethereum price, it is clear that both the number of ICOs and ICO volume closely track the cryptocurrencies' prices. Most ICOs offer their tokens in exchange for Bitcoin or other cryptocurrencies. In addition, most ICOs state their funding goals in Bitcoin instead of USD, so that the amount collected is more substantial when the price of Bitcoin is higher (Fisch, Initial coin offerings (ICOs) to finance new ventures, 2019).
- **Trading:** Dummy variable that equals 1 if the related token is traded or futures on the token are traded. This variable is considered by looking at CoinMarketCap.com and the presence of the token on the website. This variable showed that only 352 ICOs have been listed after the ICO and actually been trading, a quite low percentage of 23%.
- **After ICO performance:** Dummy variable that equals 1 if the related token is traded at a premium respect the initial ICO price and 0 otherwise. This variable is considered by looking at the current price of the token and its initial one. Only 28 ICOs are trading now with a premium, 1.8% of totality.
- **Natural logarithm of the Ethereum price at the ICO start date**

- **One-month return from investing in bitcoin before the token sale start:** This variable makes understand how much could have been the potential outcome of investing in Bitcoin and selling after one month rather than investing in the token considered.
- **One-month return from investing in ether before the token sale start:** This variable makes understand how much could have been the potential outcome of investing in Ether and selling after one month rather than investing in the token considered.

2.3.4 INFORMATION FROM WEB SOURCES

The information taken from the internet were not present on any ICO website like ICObench.com or Coinmarketcap.com, there was though the possibility to find redirecting links that bring the user to other websites.

Two fields included in the database relate to '**whitepaper link**' and '**website URL**' of the specific ICO, these information have been extracted from general internet sources. These information give the possibility to the reader to inspect more complete details of the project.

- The "Whitepaper" field contains a link that redirects to an online version of the technical whitepaper of the project. Technical whitepapers have been downloaded, when available, by the previous author and me and saved all together into the same folder, with each whitepaper document name saved as the Unicode corresponding to that ICO in the database. The link available on the database redirects from the same source from which the whitepaper has been downloaded. Usually, the sources used by previous author in order to download the papers were ICOsbull.com, Neironix.com, the website of the company issuing

the project or other free online sources. I used as principal website ICObench.com to download whitepapers and also broad internet research. It may be possible that for a record in the database it is available the link to the technical whitepaper, while the whitepaper has not been downloaded: this happened when the consulted source allows the visitor just to view the paper but not to download it. In the database is possible to access to 1362 links to their respective technical whitepapers, while the whitepapers that has been downloaded are 1374. For the remaining ICOs was not possible to find any information about technical whitepapers on the Internet.

- The “Website URL” field contains a link to the website of the company issuing the project or to a dedicated website created by the ICOs’ entrepreneurs specifically for their projects. Some of the ICOs on ICObench.com have available a link to the company’s web page, therefore in some cases the information has been taken from ICOBench.com, but in most cases the website has been looked for through online searches.

The last fields of the database to discuss relate to the ‘**relevance and visibility on Twitter**’, taken as example of social media and finally the ‘**availability of the source code**’ of the project.

The first data was taken from Twitter looking at the number of tweets and followers on each ICO’s Twitter account, if any. While the last information has been taken from GitHub, looking at the specific ICO’s repositories available on the website and searching for the presence and disclosure of the main source code of the project.

Of all projects included in the dataset the 33% had a public source code available, and the 71% had a twitter account.

3 CHAPTER 3

3.1 ADJUSTMENTS TO DATABASE

A preliminary activity to carry on after the completion of the database was its clean-up, especially from a big amount of numbers saved as 'text' so not exploitable for analysis. In order to do so, special excel formulas have been used.

The problem arose when calculating descriptive statistics of the variables taken as proxies of ICOs success, like Soft Cap, Hard Cap and Amount raised. As said before there were a huge amount of numbers saved as string or with another currency type apart from USD.

3.1.1 DATA SAVED IN ETH:

For data saved in ETH there was the necessity to remove from the text the string 'ETH' and switch the format from text to number. To do so the function 'remove and substitute' has been used in all the data columns of interest, removing 'ETH', and substituting it with "" (nothing). The second step has been the switch of format, about this there is an excel formula called VALUE() that transform any format in number. The last step was to exchange the number from ETH to USD, in order to do this a list of all exchange rates (ETH/USD) history has been downloaded and imported in excel. After, all dates of the values that were previously in ETH have been compared with the list of exchange rates thanks to the formula VLOOKUP() and then multiplied by the related exchange rate.

3.1.2 DATA SAVED IN USD:

For data saved in USD as string the process have been much easier, simply removing the string 'USD' and switching the format to number thanks to the VALUE() formula.

3.1.3 DATA SAVED IN OTHER CURRENCIES:

Given the particularly low number of values saved in other currencies, about 10/20, it has been decided to exclude them from the dataset.

3.2 DESCRIPTIVE STATISTICS

Right after the clean-up of the dataset it has been decided to carry on some descriptive statistics.

Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Measures of central tendency include the mean, median and mode, while measures of variability include standard deviation, variance, minimum and maximum variables, and kurtosis and skewness. (www.Investopedia.com, s.d.)

The main variables taken in consideration are the Soft Cap, the Hard Cap and the Amount Raised since they better estimate the level of success of an ICO.

Several statistical tools have been used to analyse and understand the impact of some variables in the dataset on these particular ones. The list of the statistical tools is the following:

3.2.1 ARITHMETIC MEAN

The arithmetic mean is that value that all observations would have if there was no variability (random or systematic). More precisely, it is that value that replaced to each of the n data makes it remain constant the sum given a set of n elements $\{x_1, x_2, \dots, x_n\}$. It is called the simple arithmetic mean of n numbers the number obtained by dividing their sum for n .

$$\bar{x} = \frac{x_1 + x_2 + \dots x_n}{n}$$

We can formally express the average simple arithmetic through the following formula:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

3.2.2 STANDARD ERROR

In statistics the standard error of a measure is defined as the estimate of the standard deviation of the estimator. It is therefore an estimate of the variability of the estimator, that is, a measure of its imprecision. If the estimator is the sample mean of n independent samples with the same statistical distribution, the standard error is:

$$se = \frac{S}{\sqrt{n}}$$

Where S is the standard deviation of the sample.

3.2.3 MEDIAN

Value / modality (or the set of values / modalities) assumed by the statistical units that are in the middle of the distribution. The median is a position index and is part of the set of order statistics. (www.Wikipedia.com, s.d.)

3.2.4 MODE

In statistics, the mode of a frequency distribution X is the value (or class of values) characterized by the maximum frequency. In other words, it is the value that appears most frequently. A distribution is unimodal if it admits only one modal value, it is bimodal if it admits two (i.e.: if there are two values that both appear with the maximum frequency in the given distribution), trimodal if it has three, etc.

3.2.5 STANDARD DEVIATION

The standard deviation of a variable is a summary index of the differences in the values of each observation with respect to the mean of the variable. In fact, each observation has a deviation (also called deviation or deviation) from the mean. This deviation is equal to 0 if the observation has the same value as the average. The deviation will instead be negative if the observation has a value smaller than that of the average. On the contrary, this deviation will be positive if the observation has a value greater than that of the average. However, the sum of all these deviations cannot be used as a measure of variability. If all the differences were added together, the negative deviations would exactly compensate for the positive deviations and their sum would be zero. By definition, in fact, the mean always corresponds to the centre of gravity of a distribution.

3.2.6 VARIANCE

In statistics the variance of a statistical variable or a random variable X is a function which provides a measure of the variability of the values assumed by the variable itself; specifically, the measure of how much they differ quadratically from the arithmetic mean or from the expected value, respectively.

3.2.7 KURTOSIS

Kurtosis in the language of statistics, is a departure from distributive normality, with respect to which there is a greater flattening (platykurtic distribution) or greater elongation (leptokurtic distribution). Its best-known measure is the Pearson index, the ratio between the centred moment of order 4 and the square of the variance. The value of the index corresponding to the normal⁹ (Gaussian) distribution is 0 (if you use the index shown below which, as you can see, is centred in zero since 3 is subtracted). A value less than 0 indicates a platykurtic distribution, while a value greater than 0

⁹ Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve. (www.investopedia.com, s.d.)

indicates a leptokurtic distribution (it is possible that some indices are not centred in zero and therefore the value obtained in the case of normality is 3).

3.2.8 ASYMMETRY (SKEWNESS)

A frequency distribution is asymmetric when the arithmetic mean does not coincide with the central maximum of the frequency curve (mode). The frequency curve of an asymmetric distribution is characterized by one tail that is longer than the other, to the right or to the left of the mode.

3.2.9 INTERVAL

In statistics, an interval represents the difference between the maximum and minimum value of a data group. Shows how the values are distributed in a series. If the range is a large number, the series values are far from each other; if it is small, they are close.

3.2.10 MINIMUM VALUE

It is the minimum value of the whole dataset taken into consideration.

3.2.11 MAXIMUM VALUE

It is the maximum value of the whole dataset taken into consideration.

3.3 DESCRIPTIVE STATISTICS – OUTPUT

As said before these statistical tools have been used to analyse the characteristics that describe the dataset taken into consideration. As first step of analysis the whole dataset has been studied on the three variables taken as proxy of success (Soft Cap, Hard Cap, Amount Raised), Only ICOs with these information available have been considered in order to have a much more precise set of data. Then the dataset has been divided in

subsamples in order to better understand the impact of specific variables on the “success of ICOs”. It is important to say that the study has not been conducted on the token-holder returns as proxy of success because of the so low percentage of positive ones, only 28 on 1540 which is 1,8% of totality. The reasons are that firstly the project needs to have completed a successful ICO in the first place. If their tokens failed to sell out, demand for the token on the secondary markets will be low and exchange platforms are unlikely to be interested. Also, exchanges charge huge amounts, ranging from around \$100,000 to \$3 million for the largest and most liquid exchanges. Among the many tasks that must be performed before a token can be listed is the necessary due diligence to ensure the project is not a scam. Should a token later be discovered to be one, it risks affecting the platform’s reputation. However, the ICOs that listed their tokens and are trading at premium are characterized by:

- 100% of data trading on Ethereum Platform
- 86% characterized by source code availability
- 100% of ICOs had their team disclosed
- 29% had a bonus scheme
- 32% had a pre-ICO phase
- 32% had an MVP or prototype available
- 25% are a finance category ICO

All the others variables did not explain any trend or impact on the positive token-holder returns.

3.3.1 ANALYSIS OF THE ENTIRE DATABASE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	10352846	Mean	47150491	Mean	18306544
Standard error	1950595	Standard error	3544636	Standard error	3173120
Median	2500000	Median	20000000	Median	6797500
Mode	5000000	Mode	30000000	Mode	5000000
Standard Deviation	51975223	Standard Deviation	1,21E+08	Standard Deviation	1,22E+08
Variance	2,7E+15	Variance	1,48E+16	Variance	1,49E+16

Kurtosis	228,4962	Kurtosis	62,32194	Kurtosis	959,5729
Skewness	13,64302	Skewness	7,108003	Skewness	29,29526
Interval	1,02E+09	Interval	1,5E+09	Interval	4,2E+09
Minimum value	0	Minimum value	0	Minimum value	0
Maximum value	1,02E+09	Maximum value	1,5E+09	Maximum value	4,2E+09
Sum	7,35E+09	Sum	5,54E+10	Sum	2,71E+10
Counting	710	Counting	1174	Counting	1478

Looking at the previous table we can see, as we expected, that the mean of the Amount Raised is between the mean of the Soft and Hard Caps. For all the variables the median is always less than the half of the mean, this explains the fact that in the dataset there are few events with very high values of the variables that lead to a big increase in the mean while the median is not so affected. The indicators of variability like standard error, standard deviation and variance take very high values telling us that the variables are characterized by relevant fluctuations in values. Consequently, also the kurtosis have very high values leading to leptokurtotic distributions, this means that distribution is massively concentrated on the mean leading to have smaller tails. Finally, the last statistical indicator to consider is the asymmetry, all three values are positive indicating an asymmetric distribution with a mean always bigger than the mode and so a longer tail on the right side.

3.3.2 ANALYSIS OF SUBSAMPLES DIVIDED BY COUNTRY

SINGAPORE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	8608808	Mean	71048460	Mean	9941593
Standard error	2617370	Standard error	27387193	Standard error	1223974
Median	3487885	Median	20000000	Median	6518600
Mode	5000000	Mode	20000000	Mode	5000000

Standard Deviation	21583387	Standard Deviation	2,26E+08	Standard Deviation	10093144
Variance	4,66E+14	Variance	5,1E+16	Variance	1,02E+14
Kurtosis	32,19025	Kurtosis	28,21325	Kurtosis	1,630941
Skewness	5,485849	Skewness	5,168248	Skewness	1,484352
Interval	1,5E+08	Interval	1,5E+09	Interval	41963140
Minimum value	5000	Minimum value	5000	Minimum value	36860
Maximum value	1,5E+08	Maximum value	1,5E+09	Maximum value	42000000
Sum	5,85E+08	Sum	4,83E+09	Sum	6,76E+08
Counting	68	Counting	68	Counting	68

USA

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	5797956	Mean	26550774	Mean	13664392
Standard error	1411143	Standard error	4459518	Standard error	2464544
Median	3000000	Median	20000000	Median	7894548
Mode	1000000	Mode	25000000	Mode	15000000
Standard Deviation	8228306	Standard Deviation	26003236	Standard Deviation	14370639
Variance	6,77E+13	Variance	6,76E+14	Variance	2,07E+14
Kurtosis	9,791175	Kurtosis	11,6383	Kurtosis	1,183809
Skewness	2,887493	Skewness	2,883838	Skewness	1,400226
Interval	40629000	Interval	1,42E+08	Interval	51885000
Minimum value	100000	Minimum value	1000000	Minimum value	115000
Maximum value	40729000	Maximum value	1,43E+08	Maximum value	52000000
Sum	1,97E+08	Sum	9,03E+08	Sum	4,65E+08
Counting	34	Counting	34	Counting	34

UK

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	5318867	Mean	33776974	Mean	9026131
Standard error	2133998	Standard error	5557937	Standard error	1910708
Median	2000000	Median	25500000	Median	5771857
Mode	1000000	Mode	40000000	Mode	1000000
Standard Deviation	13829889	Standard Deviation	36019546	Standard Deviation	12382801
Variance	1,91E+14	Variance	1,3E+15	Variance	1,53E+14
Kurtosis	33,26747	Kurtosis	17,61814	Kurtosis	16,50897
Skewness	5,570925	Skewness	3,670826	Skewness	3,621904
Interval	88116000	Interval	2,16E+08	Interval	71497437
Minimum value	55000	Minimum value	4267950	Minimum value	502563

Maximum value	88171000	Maximum value	2,2E+08	Maximum value	72000000
Sum	2,23E+08	Sum	1,42E+09	Sum	3,79E+08
Counting	42	Counting	42	Counting	42

For the following analysis, the countries chosen are Singapore, USA and UK because of their bigger presence in the sample. Singapore is present in 194 events, USA in 190 and UK in 135.

As we can see from the tables the mean of amount raised by U.S. ICOs is a bit higher than the others, but with a slightly higher variability given by standard deviation and variance. This information is important to understand that the higher mean for USA ICOs is not consistently higher than the others because of the bigger fluctuations in value. USA values are the ones with lower kurtosis, meaning that the distribution is platykurtic and so having bigger tails leading to higher probability of presence of outliers. All these information bring us to doubt a bit a consistent higher mean of amount raised from USA ICOs.

3.3.3 ANALYSIS OF SUBSAMPLES DIVIDED BY CATEGORY

FINANCE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	7503851	Mean	46511744	Mean	11095879
Standard error	1902074	Standard error	10006725	Standard error	2344123
Median	3000000	Median	24302902	Median	5152200
Mode	5000000	Mode	20000000	Mode	5000000
Standard Deviation	15684900	Standard Deviation	82517571	Standard Deviation	19330137
Variance	2,46E+14	Variance	6,81E+15	Variance	3,74E+14
Kurtosis	19,46273	Kurtosis	31,58867	Kurtosis	32,50939
Skewness	4,138988	Skewness	5,166109	Skewness	5,104081
Interval	99999000	Interval	6E+08	Interval	1,43E+08
Minimum value	1000	Minimum value	0	Minimum value	139000
Maximum value	1E+08	Maximum value	6E+08	Maximum value	1,43E+08
Sum	5,1E+08	Sum	3,16E+09	Sum	7,55E+08
Counting	68	Counting	68	Counting	68

TRADING AND INVESTING

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	31986339	Mean	56735539	Mean	13622472
Standard error	24914900	Standard error	25086008	Standard error	2232846
Median	2500000	Median	22000000	Median	8425501
Mode	1000000	Mode	20000000	Mode	3000000
Standard Deviation	1,6E+08	Standard Deviation	1,61E+08	Standard Deviation	14297193
Variance	2,55E+16	Variance	2,58E+16	Variance	2,04E+14
Kurtosis	40,17541	Kurtosis	31,49375	Kurtosis	2,184887
Skewness	6,312633	Skewness	5,425353	Skewness	1,481202
Interval	1,02E+09	Interval	9,98E+08	Interval	59872000
Minimum value	500000	Minimum value	1800000	Minimum value	128000
Maximum value	1,02E+09	Maximum value	1E+09	Maximum value	60000000
Sum	1,31E+09	Sum	2,33E+09	Sum	5,59E+08
Counting	41	Counting	41	Counting	41

INFRASTRUCTURE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	7212915	Mean	41280838	Mean	16509101
Standard error	3237404	Standard error	11337695	Standard error	6522992
Median	3000000	Median	19646000	Median	6200000
Mode	4000000	Mode	#N/D	Mode	#N/D
Standard Deviation	13735141	Standard Deviation	48101765	Standard Deviation	27674712
Variance	1,89E+14	Variance	2,31E+15	Variance	7,66E+14
Kurtosis	9,340503	Kurtosis	0,542206	Kurtosis	11,63658
Skewness	3,027245	Skewness	1,340507	Skewness	3,21569
Interval	54956947	Interval	1,5E+08	Interval	1,18E+08
Minimum value	43053	Minimum value	86106	Minimum value	47626
Maximum value	55000000	Maximum value	1,5E+08	Maximum value	1,18E+08
Sum	1,3E+08	Sum	7,43E+08	Sum	2,97E+08
Counting	18	Counting	18	Counting	18

The following analysis has been conducted on the category of ICOs most present in the sample, these are Finance with 251 events, Trading and investing with 166 and Infrastructure with 107.

We can deduct from the tables that the higher value of the mean of Amount Raised is for the category Infrastructure, this value is also linked, though, to a more than doubled variance and standard deviation, telling us the much higher volatility of the distribution. The median is quite far from the mean explaining the presence of some events with high values that bring up the mean. The kurtosis is bigger than 3 leading to a leptokurtic curve and so having smaller tails and lower probability of presence of outliers. Finally, the asymmetry is relevantly high enforcing all the hypotheses made up to now leading to a high uncertainty of declaring Infrastructure as the best category for successful ICOs.

Trading and investing have led the distribution to have a lower mean of amount raised respect to infrastructure but with a much closer median, and so having also much lower values of variability. The results are also strengthened by a very low asymmetry and a very close Kurtosis value to 3 leading to have a distribution very similar to the gaussian one.

3.3.4 ANALYSIS OF SUBSAMPLE DIVIDED BY PLATFORM

Ethereum

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	8536223	Mean	40272680	Mean	10125762
Standard error	2599537	Standard error	5329652	Standard error	701786,9
Median	2468750	Median	20000000	Median	5000000
Mode	1000000	Mode	20000000	Mode	3000000
Standard Deviation	52379220	Standard Deviation	1,07E+08	Standard Deviation	14140613
Variance	2,74E+15	Variance	1,15E+16	Variance	2E+14
Kurtosis	351,0519	Kurtosis	103,5804	Kurtosis	27,88195
Skewness	18,17668	Skewness	9,124631	Skewness	4,105201
Interval	1,02E+09	Interval	1,5E+09	Interval	1,43E+08
Minimum value	0	Minimum value	0	Minimum value	36860

Maximum value	1,02E+09	Maximum value	1,5E+09	Maximum value	1,43E+08
Sum	3,47E+09	Sum	1,64E+10	Sum	4,11E+09
Counting	406	Counting	406	Counting	406

The study conducted on the platform is related only on Ethereum platform because it is referred to almost the totality of the database, given 1354 Ethereum platform – based ICOs of 1540. Bitcoin platform has not been taken into considerations because present in only 8 events.

The mean diverges a lot from the median reaching its doubled value, variance and standard deviation are both relatively high characterizing the distribution with quite high volatility. The kurtosis value is bigger than 3 leading to leptokurtic distribution characterized by high concentration on the mean, consequently much tinner tails and so lower number of extremes outliers, this value, though, is not so high. Finally, the asymmetry is positive, meaning that the mean will always be higher than the mode and a longer tail on the right side. All the information induces to think that the values are characterized by high uncertainty.

3.3.5 ANALYSIS OF SUBSAMPLE CHARACTERIZED BY PRE-ICO EXISTENCE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	8811720	Mean	40950535	Mean	10031687
Standard error	2826035	Standard error	5660518	Standard error	746583,1
Median	2401900	Median	20000000	Median	5000000
Mode	1000000	Mode	20000000	Mode	1000000
Standard Deviation	54579801	Standard Deviation	1,09E+08	Standard Deviation	14418914
Variance	2,98E+15	Variance	1,2E+16	Variance	2,08E+14
Kurtosis	324,0229	Kurtosis	104,0982	Kurtosis	28,17681
Skewness	17,48205	Skewness	9,203343	Skewness	4,197065
Interval	1,02E+09	Interval	1,5E+09	Interval	1,43E+08
Minimum value	0	Minimum value	0	Minimum value	36860
Maximum value	1,02E+09	Maximum value	1,5E+09	Maximum value	1,43E+08
Sum	3,29E+09	Sum	1,53E+10	Sum	3,74E+09
Counting	373	Counting	373	Counting	373

The following analysis is referred to all ICOs with a pre-ICO phase in order to see the impact of this event on the overall “success”. The results of the analysis show that the mean of the amount raised is higher than the ICOs that do not have the pre-ICO phase by around 10%. But still all the values are featured by high dispersion. The median is about the half of the mean. Standard deviation and variance are very large values. Kurtosis of 28,17 is still quite relevant, making the curve diverge from a gaussian curve. Asymmetry of 4,19 suggests us the presence of a bigger right-side tail.

3.3.6 ANALYSIS OF SUBSAMPLE CHARACTERIZED BY EXISTENCE OR ABSENCE OF SOURCE CODE

EXISTANCE OF SOURCE CODE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	5850084	Mean	34967343	Mean	11144776
Standard error	800805,2	Standard error	4819281	Standard error	1030881
Median	2500000	Median	20000000	Median	6400000
Mode	5000000	Mode	30000000	Mode	5000000
Standard Deviation	10714044	Standard Deviation	64477591	Standard Deviation	13792250
Variance	1,15E+14	Variance	4,16E+15	Variance	1,9E+14
Kurtosis	36,41067	Kurtosis	37,97484	Kurtosis	20,55626
Skewness	5,171025	Skewness	5,492492	Skewness	3,488238
Interval	99998860	Interval	6E+08	Interval	1,18E+08
Minimum value	1140	Minimum value	0	Minimum value	114145
Maximum value	1E+08	Maximum value	6E+08	Maximum value	1,18E+08
Sum	1,05E+09	Sum	6,26E+09	Sum	1,99E+09
Counting	179	Counting	179	Counting	179

ABSENCE OF SOURCE CODE

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	5517109	Mean	35187810	Mean	8780749
Standard error	3098206	Standard error	7675927	Standard error	2585330
Median	1529100	Median	34250000	Median	4850000
Mode	1000000	Mode	40000000	Mode	#N/D
Standard Deviation	16394166	Standard Deviation	40617187	Standard Deviation	13680279

Variance	2,69E+14	Variance	1,65E+15	Variance	1,87E+14
Kurtosis	26,53422	Kurtosis	16,78928	Kurtosis	17,82855
Skewness	5,097856	Skewness	3,734668	Skewness	3,914682
Interval	88116000	Interval	2,16E+08	Interval	71497437
Minimum value	55000	Minimum value	4550000	Minimum value	502563
Maximum value	88171000	Maximum value	2,2E+08	Maximum value	72000000
Sum	1,54E+08	Sum	9,85E+08	Sum	2,46E+08
Counting	28	Counting	28	Counting	28

Consequently, all ICOs with source code existence and absence have been studied. It can be understood that the mean of amount raised of ICOs with the source code available is much bigger than ICOs without this characteristic, higher of about 27%. Also, these subsamples are both characterized by high volatility but a bit less pronounced kurtosis value respect to previous subsample. The median, in both cases, is almost the half of the mean meaning that the values are quite strongly dispersed in the distribution. Deviation standard and variance are quite large values since they are even bigger than the mean. But the asymmetry is more relaxed featuring the curve with a slightly bigger tail on the right side. Still the results of this analysis must be taken carefully given the high volatility.

3.3.7 ANALYSIS OF SUBSAMPLE CHARACTERIZED BY USABILITY OF TOKEN AS CURRENCY

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	8810844	Mean	21317897	Mean	12768805
Standard error	4222160	Standard error	4422186	Standard error	3916811
Median	1750000	Median	16763418	Median	2978375
Mode	1000000	Mode	30000000	Mode	#N/D
Standard Deviation	19803685	Standard Deviation	20741892	Standard Deviation	18371470
Variance	3,92E+14	Variance	4,3E+14	Variance	3,38E+14
Kurtosis	14,71028	Kurtosis	4,73727	Kurtosis	4,016668
Skewness	3,719648	Skewness	1,83637	Skewness	1,944921
Interval	89594358	Interval	90000000	Interval	71642350
Minimum value	405642	Minimum value	0	Minimum value	357650
Maximum value	90000000	Maximum value	90000000	Maximum value	72000000
Sum	1,94E+08	Sum	4,69E+08	Sum	2,81E+08

Counting	22	Counting	22	Counting	22
----------	----	----------	----	----------	----

The characteristic of usability of token as currency leads the distribution to have a relevantly higher mean of amount raised, but, at the same time, to have a much higher discrepancy between the mean and the median, suggesting us the presence of some outliers that rise the mean. Standard deviation and variance are still very high and so the fluctuations of the distribution. The kurtosis value is bigger than 3 but very close to that value, the distribution is slightly leptokurtic but near to a gaussian curve. Also, the asymmetry is a low number giving as a result the presence of almost normal distribution tails.

3.3.8 ANALYSIS OF SUBSAMPLE CHARACTERIZED BY POSITIVE AND NEGATIVE MARKET SENTIMENT

POSITIVE MARKET SENTIMENT

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	2847053	Mean	15989467	Mean	11175197
Standard error	778226,3	Standard error	2755973	Standard error	5346380
Median	2918581	Median	14216720	Median	4200000
Mode	5000000	Mode	10000000	Mode	1000000
Standard Deviation	2805935	Standard Deviation	9936804	Standard Deviation	19276647
Variance	7,87E+12	Variance	9,87E+13	Variance	3,72E+14
Kurtosis	2,416395	Kurtosis	0,830848	Kurtosis	9,812725
Skewness	1,391846	Skewness	1,256348	Skewness	3,011986
Interval	9945000	Interval	31932475	Interval	71497437
Minimum value	55000	Minimum value	4550000	Minimum value	502563
Maximum value	10000000	Maximum value	36482475	Maximum value	72000000
Sum	37011688	Sum	2,08E+08	Sum	1,45E+08
Counting	13	Counting	13	Counting	13

NEGATIVE MARKET SENTIMENT

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	6486302	Mean	41491743	Mean	7832304
Standard error	3070420	Standard error	7563453	Standard error	1393539
Median	1722040	Median	35000000	Median	5913620

Mode	1000000	Mode	40000000	Mode	1300000
Standard Deviation	16534719	Standard Deviation	40730440	Standard Deviation	7504435
Variance	2,73E+14	Variance	1,66E+15	Variance	5,63E+13
Kurtosis	23,09449	Kurtosis	13,48507	Kurtosis	7,350386
Skewness	4,660347	Skewness	3,250463	Skewness	2,3424
Interval	87744205	Interval	2,16E+08	Interval	35934643
Minimum value	426795	Minimum value	4267950	Minimum value	1065357
Maximum value	88171000	Maximum value	2,2E+08	Maximum value	37000000
Sum	1,88E+08	Sum	1,2E+09	Sum	2,27E+08
Counting	29	Counting	29	Counting	29

The two variables taken in consideration seem to have both a positive and negative impact on the successfulness of an ICO event. Since it has been seen that the volume and success of ICOs rely much on how Bitcoin price is going in the market in that period of time, it is possible to see that positive market sentiment led to have ICOs with greater amount of capital raised while at the same time negative market sentiment led to have ICOs with lower amount of capital raised.

Unfortunately, still the variance of these distributions is very high, in particular all the variance indicators of first table are much higher than the ones of the second one. But given this fact the variance of the 'negative market sentiment' table is not too pronounced. The median is quite close to the mean, meaning that the values are not much dispersed. The deviation standard and variance are much lower than the other analysis table in general. And lastly it is characterized by a relevantly low skewness, leading the curve to be as much similar as a Gaussian one. Only the kurtosis, quite bigger than 3, has a negative impact on the distribution giving to it a shape more concentrated on the mean, and so diverging it from normality. These implications can lead to the qualitative conclusion that negative market sentiment led to have worse success rate in ICOs.

3.3.9 ANALYSIS OF SUBSAMPLE CHARACTERIZED BY PRESENCE OF BONUS SCHEME

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	6289880	Mean	33500539	Mean	8884009
Standard error	3073949	Standard error	7648034	Standard error	2658306
Median	1722040	Median	25000000	Median	5000000
Mode	1000000	Mode	35000000	Mode	1300000
Standard Deviation	16553721	Standard Deviation	41185926	Standard Deviation	14315415
Variance	2,74E+14	Variance	1,7E+15	Variance	2,05E+14
Kurtosis	23,22982	Kurtosis	15,79951	Kurtosis	14,09131
Skewness	4,679633	Skewness	3,671853	Skewness	3,520449
Interval	88116000	Interval	2,16E+08	Interval	71497437
Minimum value	55000	Minimum value	4267950	Minimum value	502563
Maximum value	88171000	Maximum value	2,2E+08	Maximum value	72000000
Sum	1,82E+08	Sum	9,72E+08	Sum	2,58E+08
Counting	29	Counting	29	Counting	29

Studying the impact of the presence of this characteristic on the ICO capital raised we can see that the impact on the mean of the amount raised is much lower than the other variables, bringing to a qualitative conclusion that the bonus scheme does not positively affect the success of ICOs. The results are still featured by high dispersion, the mean diverges a lot from the median, having so very volatile values in the distribution. Standard deviation and variance are very high, giving a negative contribution on the reliability of these results. Also, the kurtosis and skewness have significant high values making the curve diverge a lot from the Gaussian one.

3.3.10 ANALYSIS OF SUBSAMPLE CHARACTERIZED BY DISCLOSURE OF THE TEAM

<i>Soft Cap</i>		<i>Hard Cap</i>		<i>Amount Raised</i>	
Mean	8473277	Mean	39998079	Mean	10102308
Standard error	2543497	Standard error	5220973	Standard error	689139,9
Median	2500000	Median	20000000	Median	5000000
Mode	1000000	Mode	20000000	Mode	5000000
Standard Deviation	51814981	Standard Deviation	1,06E+08	Standard Deviation	14038848
Variance	2,68E+15	Variance	1,13E+16	Variance	1,97E+14

Kurtosis	358,6645	Kurtosis	105,425	Kurtosis	28,12049
Skewness	18,37047	Skewness	9,19587	Skewness	4,111705
Interval	1,02E+09	Interval	1,5E+09	Interval	1,43E+08
Minimum value	0	Minimum value	0	Minimum value	36860
Maximum value	1,02E+09	Maximum value	1,5E+09	Maximum value	1,43E+08
Sum	3,52E+09	Sum	1,66E+10	Sum	4,19E+09
Counting	415	Counting	415	Counting	415

The final statistical variable considered is the team disclosure. The distribution is characterized with a more than double mean respect to the median, suggesting us the presence of outliers high in value. Quite high variance indicators that always make the information to be taken very carefully because of the large volatility. A big kurtosis value that makes the curve to be concentrated on the mean of the distribution. A light asymmetry to the right making the right tail a bit bigger and so the number of right-extreme outliers.

All the other variables present in the database have not been presented since the presence or absence of them had almost no impact on the variables taken as proxies of ICOs success.

3.4 OUTPUT DESCRIPTIVE GRAPHS

Other interesting insights given from the more detailed and deep study of the dataset in analysis are shown in the following graphs.

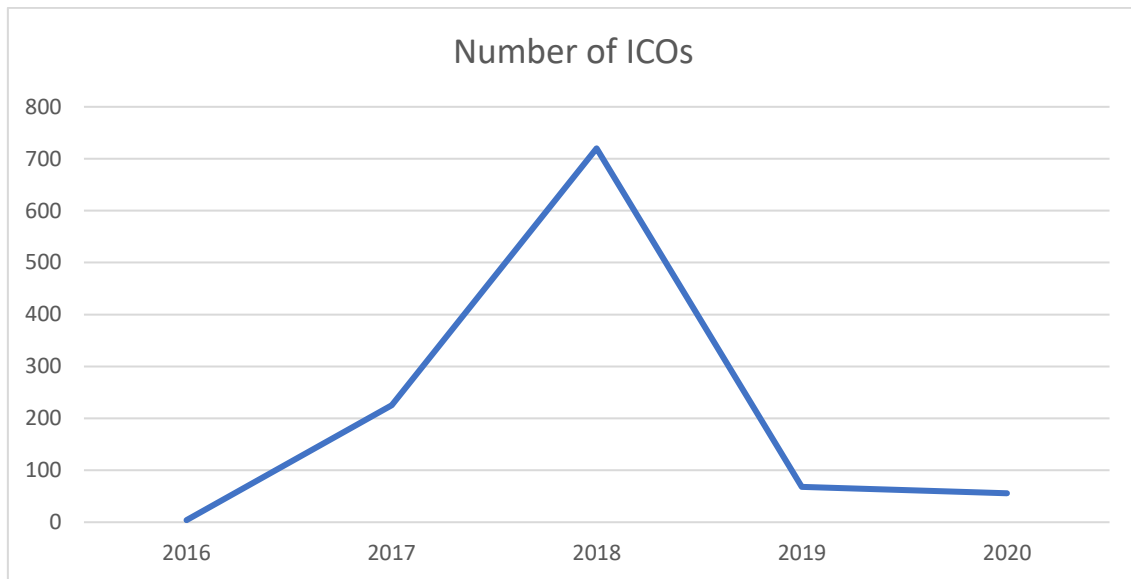


Figure 1- Number of ICOs in time

This first graph gives a view of the diffusion of ICOs in the time horizon of the dataset studied, showing a significative increase in popularity from 2016 to 2018 and a following decrease until 2020. Both Coinschedule and ICO Bench report a decline in ICO popularity altogether, partly due to fear of fraud and exit scamming. Most ICOs inevitably collapse from lack of funds, poor budgeting, etc.

Investors often complain of their intrinsic risks, and the little to no chance of ROI¹⁰(Return on investment) especially when considering more than half of them fail within the first 4 months (www.coindesk.com, s.d.). As the database studied demonstrates, only 1,8% of projects tokens are trading at premium so generating value for token-holders who believed in the business.

Even still, their allure is undeniable, making millionaires out of the investors that strike gold. Regardless of this fact, less than 8 percent of them will be tradable on the main exchanges (www.investopedia.com, s.d.) since exchanges only list single digit coins. So, while the return can be massive, there is also a massive chance return will be 0 percent, or worse.

¹⁰ ROI: Return on Investment (ROI) is a performance measure used to evaluate the efficiency or profitability of an investment or compare the efficiency of a number of different investments. ROI tries to directly measure the amount of return on a particular investment, relative to the investment's cost. (www.investopedia.com, s.d.)

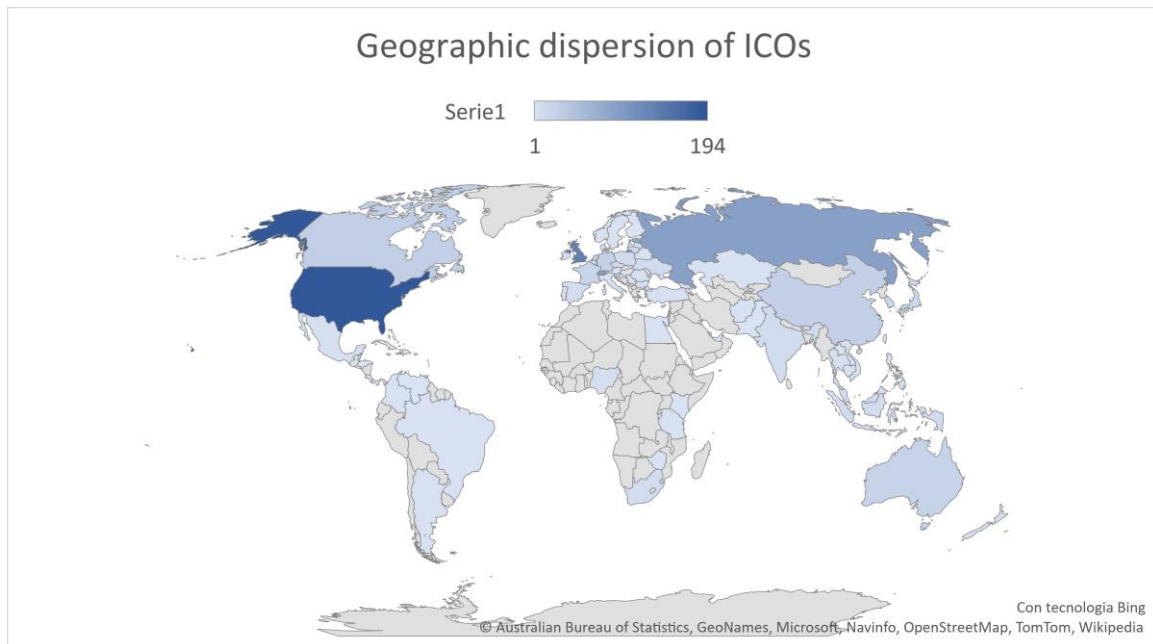


Figure 2- Geographic dispersion

A following interesting aspect given by the analysis of the list of ICOs in the dataset is the geographic popularity of ICOs worldwide as shown in the graph above. It is possible to understand the major use of this source of funding in countries like North America, UK, Russia and, even if not visible because of its small size, Singapore. These events are due to several reasons that have been treated in Chapter 1, specifically in section 1.8 “ICOs Geography”.

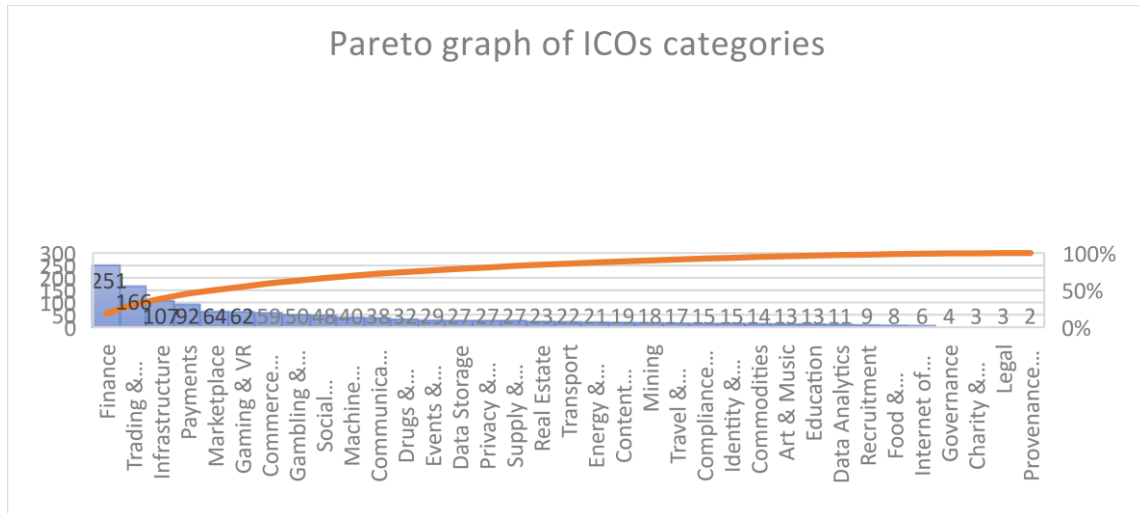


Figure 3- ICOs categories diffusion

Deepening the analysis, we can see from the graph above the percentage of totality of each ICO category present in the database. We can see that majority of the projects are related to Finance, Trading and Investments, Infrastructure, Payments, Gaming etc. This is understandable because of the real aim of the birth of this source of funding and the whole digital system related, that is to re-invent the traditional finance system linked to the fundamental roles of banks.

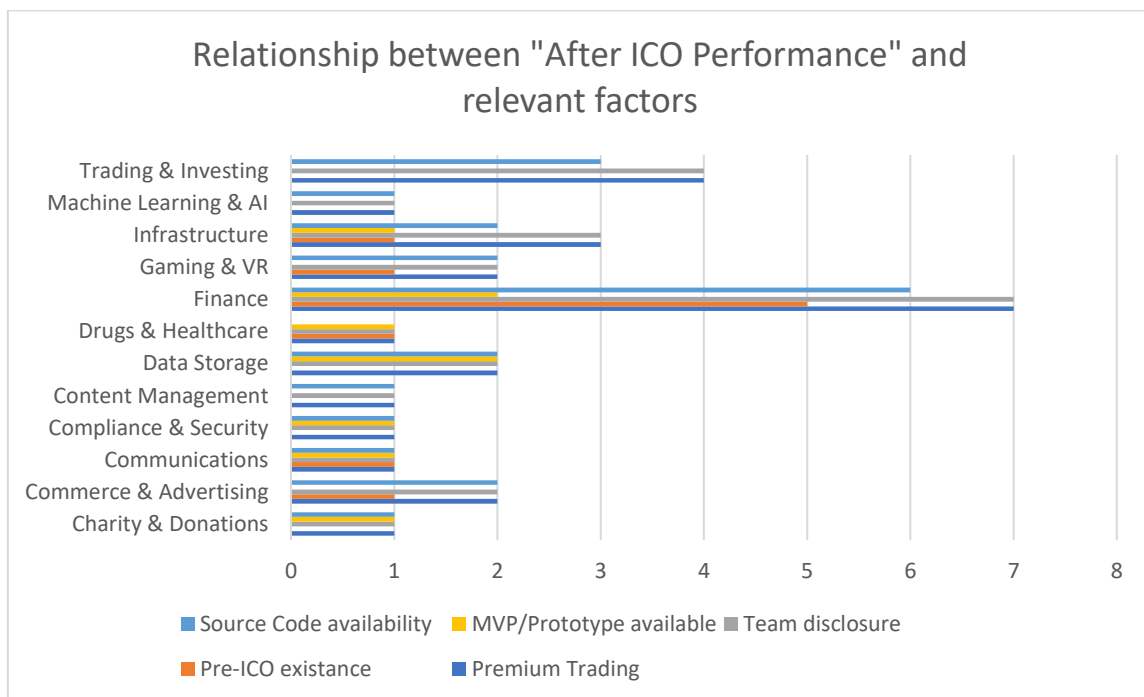


Figure 4- Relationship between "After ICO Performance" and other relevant factors

Talking about ROI of ICOs investors, the graph above shows the link between the main characteristics of a project and the fact that their tokens are trading at premium respect to the ICO price, all divided by category; the diverse categories have specific number of ICOs with these characteristics present. It is possible to understand that the presence of important factors like the Whitepaper, Team disclosure, presence of a pre-ICO period and the availability of the source code of the project does not really affect clearly, either positively or negatively the token performance after the completion of the ICO. The graph shows how the number of ICOs with the respective token trading at premium exists heterogeneously without any specific and clear relation with the presence of the other factors in consideration. The statistical significance of the analysis is, though, a bit compromised by the very low number of ICOs trading at premium, only 28 out of 1540.

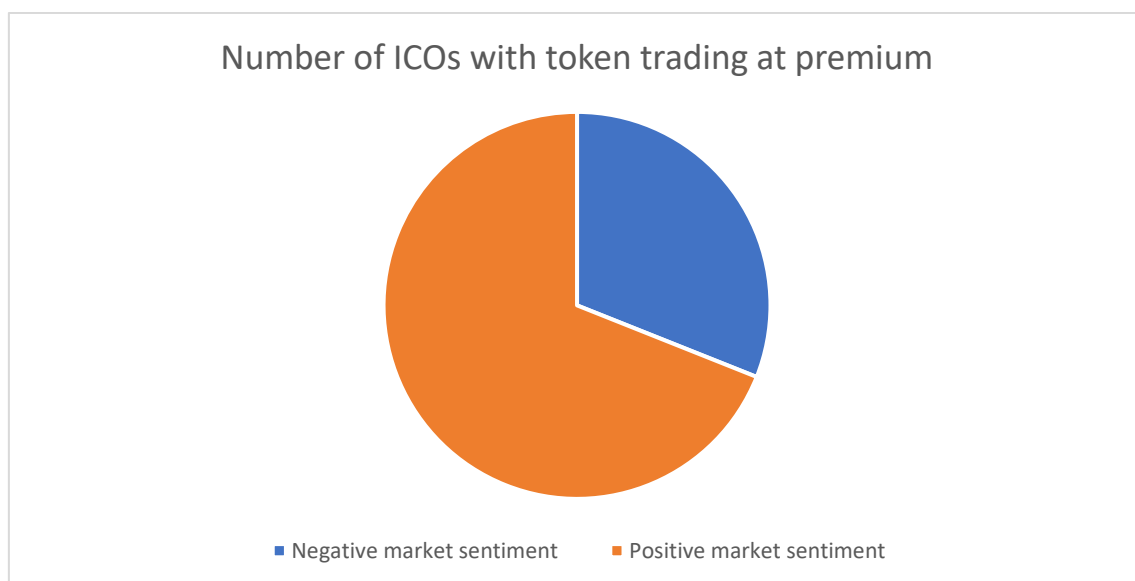


Figure 5- Relationship between "After ICO Performance" and Market Sentiment proxied by change in Bitcoin price

Furthermore, deepening the study on "After ICO performance" a clearer relation exists between the positive performance of the token and the positive perception of the market about Initial Coin Offerings, proxied by the positive delta on Bitcoin price meaning that the market is believing that the digital cryptocurrencies and the linked

projects are in a growing and prosperous period and buys Bitcoin with which possibly could invest in token sales.

3.5 CONCLUSION

As conclusion it can be said that all the variables that have been studied to impact the success of the ICOs are characterized by pronounced variability making each analysis very exposed to very volatile values. Almost all the samples are characterized by some kind of asymmetry and quite relevant divergence from the normal distribution. Given these facts it is anyway possible to say that some variables (like trading and investing, USA) positively affect the amount raised value and others like 'negative market sentiment' negatively affect the amount of capital raised, but this consideration must be taken only qualitatively because of the absence of statistically strong relevant outcomes. Furthermore, it can be said that the positive "After ICO Performance" is not clearly related to the presence of any other factor, this could be also for the very low number of ICOs tokens trading at premium, only 1,8% of totality, that could negatively affect the statistical significance of the results.

4 CHAPTER 4

4.1 DATABASE METHODOLOGY

After the completion of the database about the “success” factors and after a thorough statistical analysis about the impact of the relevant factors on the actual “success” of an ICO, recognized as the capital that the project has been able to raise, the focus of the thesis work has shifted to looking for a social factor that could lead to having a substantial effect on the ICO success.

In the ICO context the media play a fundamental role in delivering information to potential investors who mostly rely on these news, due to the quite unregulated nature of this business about disclosure of relevant and consistent information. So we can assume that the way news are presented and their content may shape investors' beliefs and consequently drive the raise of funds for these projects. This is the reason why the following analysis has been carried out.

4.2 SOURCE

This section of the chapter will highlight and describe the sources examined, with the aim of obtaining the data required for the population of the database.

-LexisNexis

The analysis has been conducted searching for all the news available on a website called LexisNexis, this website has primarily a law research function, and is a massive data repository and aggregator. They house millions of records across various niches of data, both public and private.

“LexisNexis is a corporation providing computer-assisted legal research as well as business research and risk management services. During the 1970s, LexisNexis

pioneered the electronic accessibility of legal and journalistic documents... provides customers with access to billions of searchable documents and records from more than 60000 legal, news and business sources". (www.LexisNexis.com)

LexisNexis pulls their data from databases all over the country and assembles the information in one central location, allowing underwriters the ability to learn more about the person's overall financial standing outside of just a credit report. Federal, state, and county databases provide much of the information housed inside these reports.

-Alexa

Alexa Internet, Inc. is an American web traffic analysis company based in San Francisco. It is a wholly owned subsidiary of Amazon.

Alexa was founded as an independent company in 1996 and acquired by Amazon in 1999 for \$250 million in stock. Alexa provides web traffic data, global rankings, and other information on over 30 million websites. Alexa estimates website traffic based on a sample of millions of Internet users using browser extensions, as well as from sites that have chosen to install an Alexa script. As of 2020, its website is visited by over 420 million people every month. (www.LexisNexis.com, s.d.)

Thanks to this website it has been possible to rank a relevant percentage of the news sources found and understand their main geographic audience.

4.3 METHODOLOGY

After having analysed the source in consideration, it has been chosen to populate the database almost manually, this because the website gives to the user the possibility to automatically download the data of interest but not all the fields that were interesting for the investigation. It has been possible to find 3378 news referring to 666 univocal ICOs.

The path decided to follow was to download in an Excel file all the news related to a specific ICO, this feature downloaded only the title of the news, their date of publication and the source. The sources found are 392, so many news have been posted by same sources.

Since it was crucial to have the text of the news document the only possible way was to download on a Word file all the texts related to the same group of news and manually copy and paste them on the Excel file. Then it was added manually the respective ICO id in order to do following comparisons.

After doing so for all the ICOs present in the original database, it has been chosen to provide information about the date of publication in relation to the date of beginning and ending of the token sales, in particular it has been studied if the news date was prior or not of both date of project beginning and ending. This has been done transporting all the ICO related dates of original database to the new one using the VLOOKUP() Excel formula and through the IF() formula available on Excel that returned a dummy variable of 1 if the news date was prior the ICO begging or ending date, and 0 otherwise. As last step it has been added other relevant information about the news sources from Alexa. But it has been possible to extract these information only for 106 out of the 392 sources. This could be due to the fact that a relevant part of the sources were not websites, since Alexa only have information about websites.

In this section of the chapter it will be described each of the variables available in the database's structure. The fields will be divided according to the sources from which they have been taken, which may be LexisNexis or Alexa.

4.3.1 INFORMATION FROM LEXISNEXIS

The information, regarding the 3378 news taken from the website, are about:

- **Title + Text:** Including the news title and the news text separated by a space.
- **Publication Date:** Containing the publication date.

- **Source:** Explaining the source from which the information has been extracted.

4.3.2 INFORMATION FROM ALEXA

The information, regarding the 106 sources taken from the website, are about:

- **Alexa Global Ranking:** Alexa Rank is a third-party marketing tool that estimates website traffic. Alexa Rank is often used by online businesses for competitive analysis. Alexa rankings are a composite between how many people are estimated to have visited the site and how many pages have been viewed. In this case it has been extracted the value for the whole globe.
- **Alexa Local Ranking:** As above this variable measures the website traffic but specifically in the country in which the website is mainly visited.
- **Main Audience Geography:** In this case the variable measure which is the country with the highest percentage of that website visitors.

4.3.3 ANALYSIS OF NEWS TEXTS

After having collected all the available news texts about each ICO present in the original database they have been studied to extract several information like:

- **World count:** Counting how many words were present in the document.
- **Lexical Diversity:** Measuring how many different words that are used in a text.
- **Sentiment:** It refers to determining the opinions or sentiments expressed on different features or aspects of entities.
- **Posemo:** It determines the percentage of positive emotion in the text.
- **Negemo:** It determines the percentage of negative emotion in the text.
- **FocusPast:** This variable measures the focus on the past time given from the words in the text.

- **FocusPresent:** This variable measures the focus on the present time given from the words in the text.
- **FocusFuture:** This variable measures the focus on the future time given from the words in the text.
- **ComplexityTFIDFavg:** This variable measures the lexical complexity in the text.

And many others that are not taken into consideration for this work. These variables have been chosen because significantly related to potential investors behaviors that could explain a more “successfulness” of a token sale.

World Count, Lexical Diversity and ComplexityTFIDFavg could potentially lead to a perception of higher success of that ICO as explained by Florisyak & Schandlbauer, 2020 and Samieifair & Baur, 2020.

Sentiment, Posemo and Negemo are all related to the investors feelings, this is important because it has been shown in the literature that investors follow the wave of the whole ICO market, so if some players go in a direction than the whole market goes in that direction (Saiedi, Brostrom, & Ruiz, 2020).

Lastly, the FocusPast, FocusPresent and FocusFuture variables have been included because past historical performances of the business or the founders of that business are fundamental to assess the quality of them, eliminating a big part of the business risk. This in relation with variables that focus, in the other way, to the present and the future.

4.3.4 THEORETICAL FRAMEWORK OF THE VARIABLES

The following ANOVA analyses have been conducted on the ICOs news' semantic variables. They have been calculated and extracted following the most well-known semantic variables recognized by the Linguistic Inquiry and Word Count (LIWC).

LIWC is a transparent text analysis program that categorize words in psychologically meaningful categories. Empirical results using LIWC demonstrate its ability to detect

meaning in a wide variety of experimental settings, including to show attentional focus, emotionality, social relationships, thinking styles, and individual differences. (Pennebaker & Tausczik)

POSEMO AND NEGEMO

Positive and Negative Emotions explain the degree to which people express emotion and how they express emotion. People react in radically different ways to traumatic or important events; how people react may say a lot about how they cope with the event and the extent to which the event plays a role in the future. At the heart of reacting and coping with events is people's emotional response. Research suggests that LIWC accurately identifies emotion in language use. For example, positive emotion words (e.g., love, nice, sweet) are used in writing about a positive event, and more negative emotion words (e.g., hurt, ugly, nasty) are used in writing about a negative event (Kahn, Tobin, Massey, & J.A., 2007). LIWC ratings of positive and negative emotion words correspond with human ratings of the writing excerpts (Alpers, Winzelberg, C., & H.D., 2005).

FOCUSPAST, FOCUSFUTURE, FOCUSPRESENT

Personal pronouns provide information about the subject of attention, analyses of the tense of common verbs can tell about the temporal focus of attention. In this case any verb related to past will increase the value of the variable FOCUSPAST, any verb referring to future and present actions will increase the respective semantic variables. (Pennebaker & Tausczik)

LEXICAL DIVERSITY

The lexical diversity indicates the richness of the user's language, so the diversity of a user's vocabulary (Russell, 2013). Vaezi and Kafshgar (Vaezi & Kafshgar, 2012) stated that extraverts tend to write more complexly structured publications, while the speech

of the people more introverse is poor in terms of variety of used words. There have been made interesting findings about the lexical diversity. For example, in the work of Russell (Russell, 2013) the authors state that the diversity can be used for understanding the competence of a person during the discussion of some problems. This can be easily related to the news' writers knowledge about the business their talking about, suggesting that the more disclosed information are available to the market the higher the possibility to find a higher number of news with greater Lexical Diversity, that can ultimately lead the investors to perceive a better quality Initial Coin Offering.

WORD COUNT

The word count is simply a measure of length of a text and it is measured by counting the number of words present in that text. As said before previous studies on the length of whitepapers have led investors to perceive a significantly higher quality of those projects.

SENTIMENT

Sentiment analysis is a text analysis method that detects polarity (e.g. a positive or negative opinion) within a text. Sentiment analysis aims to measure the attitude, sentiments, evaluations and emotions of a speaker/writer based on the computational treatment of subjectivity in a text. The key aspect of sentiment analysis is to analyze a body of text for understanding the opinion expressed by it. Typically, this sentiment is quantified with a positive or negative value, called polarity. The overall sentiment is often inferred as positive, neutral or negative from the sign of the polarity score. (www.Kdnuggets.com, s.d.)

5 CHAPTER 5

5.1 ADJUSTMENTS TO DATABASE

A preliminary activity to carry on after the computation of the news related variables was to import from the previous database all data of interest for the following statistical analyses.

In particular the information imported were about the Country, ICO Category, Capital raised, Source code availability, Team disclosure, Pre-ICO phase availability, Bonus program presence, MVP or prototype available, and it has been done thanks to the Excel formula VLOOKUP().

5.1.1 ANOVA ANALYSES

Right after the import of the dataset it has been decided to carry on some ANOVA analyses on the variables considered more interesting.

ANOVA stands for Analysis Of Variance and was formulated by Ronald Fisher in the year 1918. In practice, it represents a statistical technique that is used to check whether the averages of two or more groups are significantly different from each other. The aim of the technique is to verify ("or disprove") the null hypothesis using a statistical test.

The ANOVA analyses have been conducted on the most relevant news' factors for this study and the relevant categorical variables of the first database.

5.1.2 ANOVA HYPOTHESIS

When performing an ANOVA analysis of variance, two assumptions are made. The first is called the null hypothesis, and is indicated with the letter H_0 , the other H_1 . Usually the null hypothesis tries to state that all the averages of the different groups are equal to each other. In mathematical terms, this translates into:

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ ("all the means of the k groups are equal").

With μ = mean of groups and k = number of groups. On the contrary, however, with the hypothesis H_1 we try to disprove the hypothesis H_0 , that is to verify that there is at least one average different from the others. Formally it is said that:

H_1 : at least one different μ_i ("at least one of the means of the k groups is not equal to the others").

Where μ_i is the mean of the population of group i ($i = 1, 2, \dots, k$).

5.1.3 STATISTICAL TEST

To test the null hypothesis, the one-factor analysis of variance is based on a statistical test, called test F , in honor of Fisher.

The F test is simply a ratio of two variances and follows the value of the F distribution.

Since the F distribution assumes that the null hypothesis is true, we can place the F value from our study in the F distribution to determine how much our results are consistent with the null hypothesis and in order to calculate the probabilities.

First we need to calculate the F -test:

$$\text{Test F} = \frac{\frac{SSB}{k-1}}{\frac{SSW}{N-k}}$$

Equation 1 - F test

With:

SSB (Sum of Squared Between): is the sum of the variances between the means of the various groups;

k - 1 = are the degrees of freedom in the case of SSB;

SSW (Sum of Squared Within): is the sum of the variances within the means of the various groups;

N - k = are the degrees of freedom in the case of SSW (with N the number of valid observations).

We can say that the numerator of the formula follows a chi-square distribution, and also the denominator of the formula follows another chi-square distribution (this depends on the degrees of freedom).

The value of the F test is useful to be compared with the critical statistical value F which depends on the level of significance alpha chosen. The latter usually assumes values equal to 5%, and this means that assuming the null hypothesis, there is less than 5% probability of obtaining the F test, which will make us reject the null hypothesis.

If the probability is low enough (less than 0.05), we can conclude that our data is inconsistent with the null hypothesis, and this means that it is not verified, but disproved (this results in a high F test value).

On the contrary, if the probability is high, we can accept the null hypothesis H0 (this translates into a low value of the F test).

The critical statistical value F is the limit value that tells us whether to accept the null hypothesis by comparing it with the F test. In practice, a value of the F test greater than the critical statistical value F will make us exclude the null hypothesis, on the contrary a value smaller than the critical statistical value F will make us accept the null hypothesis.

5.1.4 ANOVA BETWEEN CATEGORIES AND LEXICAL DIVERSITY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Art & Music	29	13,647	0,470586	0,005766
Charity & Donations	10	4,596	0,4596	0,005627
Commerce & Advertising	121	57,24	0,473058	0,009652
Commodities	23	11,151	0,484826	0,004601
Communications	113	56,696	0,501735	0,005829
Compliance & Security	34	17,187	0,5055	0,004022
Content Management	53	24,839	0,46866	0,004373
Data Analytics	21	10,795	0,514048	0,004187
Data Storage	31	16,845	0,543387	0,014569
Drugs & Healthcare	89	43,418	0,487843	0,007533
Education	7	3,445	0,492143	0,012148
Energy & Utilities	63	30,705	0,487381	0,003514
Events & Entertainment	66	33,647	0,509803	0,005069
Finance	598	295,552	0,494234	0,006069
Food & Beverages	13	5,959	0,458385	0,006504
Gambling & Betting	81	40,652	0,501877	0,005964
Gaming & VR	176	85,164	0,483886	0,007109
Governance	10	4,644	0,4644	0,00196
Identity & Reputation	28	14,291	0,510393	0,00464
Infrastructure	213	105,165	0,493732	0,007088
Internet of Things	3	1,422	0,474	0,001057
Legal	12	6,018	0,5015	0,006989
Machine Learning & AI	101	50,959	0,504545	0,005878
Marketplace	123	61,184	0,497431	0,006322
Mining	30	14,368	0,478933	0,008194
Payments	221	106,516	0,481973	0,00544
Privacy & Security	55	28,354	0,515527	0,004952
Real Estate	40	18,868	0,4717	0,003476
Recruitment	29	14,972	0,516276	0,003804
Social Network	118	59,263	0,502229	0,00465

Supply & Logistics	75	37,935	0,5058	0,004236
Trading & Investing	455	225,071	0,494662	0,00617
Transport	33	15,7	0,475758	0,006065
Travel & Tourism	28	13,881	0,49575	0,008507

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	0,431751	33	0,013083	2,132687	0,000184	1,440197
Inside Groups	18,82123	3068	0,006135			
Total	19,25298	3101				

Firstly, checking the statistical significance of these results and seeing that the F test is higher than F critical we can say that the groups means are different.

Looking at the means values in the first table it is possible to see that there is not relevant difference between the groups since the means are quite close to each other. The only higher value is related to the category Data Storage with 0,54, while the lowest values belong to the categories Charity and Donations and Food and Beverages.

The results show that almost every category has about the same lexical diversity in the related news.

5.1.5 ANOVA BETWEEN CATEGORIES AND SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Art & Music	29	27,2025	0,938017	0,063212
Charity & Donations	10	9,9698	0,99698	1,17E-05
Commerce & Advertising	121	97,7471	0,807827	0,21339
Commodities	23	16,3561	0,711135	0,261049
Communications	113	110,0988	0,974326	0,005701
Compliance & Security	34	33,59	0,987941	0,000523
Content Management	53	48,2733	0,910817	0,104965
Data Analytics	21	20,4231	0,972529	0,003134

Data Storage	31	21,9276	0,707342	0,327756
Drugs & Healthcare	89	84,9034	0,953971	0,030387
Education	7	6,2658	0,895114	0,060409
Energy & Utilities	63	61,3013	0,973037	0,003835
Events & Entertainment	66	47,9044	0,725824	0,336205
Finance	598	535,9937	0,896311	0,113909
Food & Beverages	13	10,7936	0,830277	0,296208
Gambling & Betting	81	76,0764	0,939215	0,062174
Gaming & VR	176	159,3084	0,905161	0,096899
Governance	10	9,9619	0,99619	1,68E-05
Identity & Reputation	28	25,5001	0,910718	0,066131
Infrastructure	213	200,386	0,940779	0,053658
Internet of Things	3	2,985	0,995	1,26E-05
Legal	12	11,9476	0,995633	9,39E-06
Machine Learning & AI	101	83,8961	0,830654	0,17846
Marketplace	123	101,5631	0,825716	0,226484
Mining	30	23,1533	0,771777	0,245183
Payments	221	200,7979	0,908588	0,113959
Privacy & Security	55	41,2276	0,749593	0,311714
Real Estate	40	35,3887	0,884718	0,10289
Recruitment	29	24,8944	0,858428	0,116569
Social Network	118	105,483	0,893924	0,141052
Supply & Logistics	75	72,9621	0,972828	0,0141
Trading & Investing	455	399,6629	0,87838	0,137534
Transport	33	32,1256	0,973503	0,006824
Travel & Tourism	28	24,2732	0,8669	0,163684

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	11,08237	33	0,335829	2,805129	2E-07	1,440197
Inside Groups	367,3001	3068	0,11972			
Total	378,3825	3101				

Also in this case it is possible to see that there is statistical significance in the results. This time the means are much more dispersed, there is relevant differences between the values and it can ben said that the highest sentiment variables belong to the categories: Charity and Donations, Internet of things, Compliance and Security and Legal. It can be assumed that the high value about the latter is due to the aim of insist more on the regularization of token sales and this could be positively seen.

The lower values are attributed to Commodities, Data storage and Privacy and Securities that do not attract to much interest from the news writers.

5.1.6 ANOVA BETWEEN CATEGORIES AND POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Art & Music	29	114,41	3,945172	9,690254
Charity & Donations	10	49,45	4,945	2,095939
Commerce & Advertising	121	378,31	3,126529	3,256801
Commodities	23	52,12	2,266087	2,718598
Communications	113	417,31	3,693009	2,221759
Compliance & Security	34	129,77	3,816765	2,833459
Content Management	53	154,11	2,907736	1,080218
Data Analytics	21	65,5	3,119048	2,085899
Data Storage	31	101,61	3,277742	2,682311
Drugs & Healthcare	89	325,93	3,662135	1,815653
Education	7	16,45	2,35	0,9392
Energy & Utilities	63	233,26	3,70254	1,753784
Events & Entertainment	66	196,08	2,970909	1,949593
Finance	598	1839,95	3,076839	1,653811
Food & Beverages	13	34	2,615385	1,069194
Gambling & Betting	81	318,79	3,935679	2,219787
Gaming & VR	176	530,19	3,012443	1,336246
Governance	10	35,86	3,586	0,648227
Identity & Reputation	28	83,11	2,968214	1,756186
Infrastructure	213	677,41	3,180329	1,243565
Internet of Things	3	10,58	3,526667	1,102533
Legal	12	52,12	4,343333	0,733497
Machine Learning & AI	101	280,52	2,777426	2,239041
Marketplace	123	392,12	3,187967	2,122544
Mining	30	86,8	2,893333	0,553182
Payments	221	671,06	3,036471	1,454859
Privacy & Security	55	196,32	3,569455	2,357924
Real Estate	40	107,34	2,6835	1,977039
Recruitment	29	92,9	3,203448	2,029123
Social Network	118	448,89	3,804153	2,942708
Supply & Logistics	75	223,7	2,982667	0,733901
Trading & Investing	455	1368	3,006593	1,634911
Transport	33	103,3	3,130303	1,107134

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	354,9377	33	10,75569	5,757995	4,11E-23	1,440197
Inside Groups	5730,892	3068	1,867957			
Total	6085,83	3101				

Again, also for this analysis there is statistical significance in the results that demonstrate that the categories that were related to a higher percentage of Sentiment, from previous ANOVA, are also related to a high percentage of Posemo (positive emotions). In fact we find Charity and Donations and Legal as the highest values and then others like Gambling and Betting and Art and Music, the last one though has a much higher variance respect to the others so the result has to be taken carefully.

The lowest values are related to Commodities and Education, probably because their positive correlation with ICOs, seen by the crowd, is scarce.

5.1.7 ANOVA BETWEEN CATEGORIES AND FOCUSFUTURE

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Art & Music	29	39,98	1,378621	0,685312
Charity & Donations	10	12,96	1,296	0,116049
Commerce & Advertising	121	142,41	1,176942	0,728268
Commodities	23	17,03	0,740435	0,225586
Communications	113	170,4	1,507965	1,276865
Compliance & Security	34	49,48	1,455294	0,945638
Content Management	53	60,5	1,141509	0,33839
Data Analytics	21	38,27	1,822381	0,701249
Data Storage	31	34,24	1,104516	0,666506
Drugs & Healthcare	89	125,61	1,411348	0,513396
Education	7	10,46	1,494286	2,302762
Energy & Utilities	63	85,22	1,352698	0,660004
Events & Entertainment	66	69,84	1,058182	0,515637
Finance	598	818,45	1,368645	0,617613

Food & Beverages	13	13,49	1,037692	0,243636
Gambling & Betting	81	137,18	1,69358	0,890548
Gaming & VR	176	239,86	1,362841	0,553694
Governance	10	19,52	1,952	1,309262
Identity & Reputation	28	28,32	1,011429	0,494939
Infrastructure	213	288,54	1,354648	0,671684
Internet of Things	3	3,57	1,19	1,1221
Legal	12	16,28	1,356667	0,735588
Machine Learning & AI	101	135,3	1,339604	0,96735
Marketplace	123	170,73	1,388049	0,731258
Mining	30	27,8	0,926667	0,284464
Payments	221	303,15	1,371719	0,58342
Privacy & Security	55	64,21	1,167455	0,569705
Real Estate	40	50,85	1,27125	0,877298
Recruitment	29	41,04	1,415172	0,595319
Social Network	118	145,46	1,232712	0,426066
Supply & Logistics	75	91,56	1,2208	0,560118
Trading & Investing	455	638,77	1,40389	0,878774
Transport	33	46,74	1,416364	0,315236
Travel & Tourism	28	33,4	1,192857	1,019925

ANALISYS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	61,17426	33	1,853765	2,680696	7,59E-07	1,440197
Inside Groups	2121,595	3068	0,691524			
Total	2182,77	3101				

First thing to look at, as usual, is the statistical significance and it can be understood that in this case there is.

This ANOVA analysis studies the relation with categories and the presence of words with a focus on the future on their respective news texts.

The results show that the highest value is referred to Data Analytics with a value of 1,82. Also Education has a high result with a value of 1,49, but at the same time it has a quite high variance, much higher than the others, also it is possible to say that these results may be so because of the very low population of ICOs belonging to Education category.

While the lowest value is related to category Commodities, reflecting a very low focus on the future time horizon.

5.1.8 ANOVA BETWEEN CATEGORIES AND FOCUSPRESENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Art & Music	29	212,6	7,331034	8,605138
Charity & Donations	10	87	8,7	3,215889
Commerce & Advertising	121	812,97	6,71876	4,069891
Commodities	23	151,96	6,606957	2,061049
Communications	113	770,18	6,815752	6,336485
Compliance & Security	34	222,65	6,548529	3,586746
Content Management	53	384,08	7,246792	3,913707
Data Analytics	21	142,43	6,782381	2,446209
Data Storage	31	179,67	5,795806	4,433538
Drugs & Healthcare	89	601,95	6,763483	3,814941
Education	7	47,9	6,842857	5,300524
Energy & Utilities	63	416,18	6,606032	3,66984
Events & Entertainment	66	414,96	6,287273	3,154346
Finance	598	4064,51	6,796839	3,54854
Food & Beverages	13	117,45	9,034615	4,663694
Gambling & Betting	81	547,63	6,760864	4,707563
Gaming & VR	176	1128,3	6,410795	3,776893
Governance	10	81,47	8,147	0,877357
Identity & Reputation	28	194,66	6,952143	3,149106
Infrastructure	213	1443,77	6,778263	4,447896
Internet of Things	3	19,55	6,516667	3,626433
Legal	12	97,84	8,153333	5,615442
Machine Learning & AI	101	663,99	6,574158	3,289855
Marketplace	123	825,7	6,713008	5,268362
Mining	30	193,51	6,450333	5,948624
Payments	221	1573,31	7,11905	3,927993
Privacy & Security	55	338,5	6,154545	3,136718
Real Estate	40	260,09	6,50225	4,531936
Recruitment	29	211,14	7,28069	4,116764
Social Network	118	850,26	7,205593	3,10908
Supply & Logistics	75	454,17	6,0556	3,749157
Trading & Investing	455	3070,04	6,747341	3,48192
Transport	33	220,28	6,675152	3,624013
Travel & Tourism	28	214,48	7,66	5,30603

ANALISYS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	389,5104	33	11,80334	2,99647	2,46E-08	1,440197
Inside Groups	12085,11	3068	3,939083			
Total	12474,62	3101				

In this case the ANOVA measures the relation between categories and the text words characteristic of focusing about the present time.

Checking the F test it shows statistical significance in the results. The highest means of 9,03 and 8,7 are attributable to categories Food & Beverages and Charity and Donations but there are also a quite high variance of 4,66 and 3,21 that interferes with the results and make their value to take carefully. Other high results belong to Legal and Governance with values of about 8,15 but most significative is Governance since it has a much lower variance of only 0,87 respect to Legal's variance of 5,61.

The lowest result is for Supply & Logistics with value of 6,05 but quite high variance of 3,74.

5.1.9 ANOVA BETWEEN COUNTRIES AND SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Afghanistan	2	1,9861	0,99305	3,28E-05
Anguilla	5	4,8762	0,97524	0,00037
Argentina	13	12,0904	0,930031	0,048197
Australia	74	65,0976	0,879697	0,150507
Austria	5	4,982	0,9964	7,47E-06
Belgium	10	9,9022	0,99022	0,000324
Belize	52	43,0272	0,827446	0,250886
Bermuda	1	0,9984	0,9984	#DIV/0!
Brazil	11	10,8924	0,990218	7,42E-05
British Virgin Islands	50	48,9598	0,979196	0,00231

Bulgaria	12	9,8034	0,81695	0,262244
Cambodia	10	9,967	0,9967	1,76E-06
Canada	59	54,7458	0,927895	0,048676
Cayman Islands	174	149,7219	0,860471	0,193976
China	56	51,5709	0,920909	0,080936
Colombia	1	0,9988	0,9988	#DIV/0!
Costa Rica	1	0,9958	0,9958	#DIV/0!
Croatia	1	0,9684	0,9684	#DIV/0!
Cyprus	18	14,6743	0,815239	0,263031
Czech Republic	8	7,9145	0,989313	3,79E-05
Denmark	3	2,9957	0,998567	2,62E-06
Estonia	183	164,9484	0,901357	0,112566
Estonia	5	4,9773	0,99546	1,1E-05
France	80	78,1824	0,97728	0,004259
Georgia	3	2,9932	0,997733	8,23E-07
Germany	49	43,0453	0,878476	0,165326
Gibraltar	107	97,5044	0,911256	0,087212
Hong Kong	85	78,8131	0,927213	0,08416
India	6	5,9749	0,995817	5,91E-06
Indonesia	12	8,8348	0,736233	0,124582
Indonesia	2	-1,0298	-0,5149	0,227138
Ireland	7	6,973	0,996143	1,03E-05
Isle of Man	10	9,9102	0,99102	8,44E-05
Israel	36	35,3537	0,982047	0,002061
Italy	5	4,9926	0,99852	3,4E-06
Japan	27	25,5586	0,946615	0,017927
Laos	10	7,8231	0,78231	0,383822
Liechtenstein	12	6,1601	0,513342	0,482702
Lithuania	18	15,7253	0,873628	0,189188
Luxembourg	14	10,9525	0,782321	0,193826
Malaysia	23	9,0965	0,3955	0,510561
Malta	53	50,7964	0,958423	0,02624
Marshall Islands	6	5,9417	0,990283	0,000156
Mexico	11	10,6082	0,964382	0,009587
Netherlands	48	46,3982	0,966629	0,032256
New Zealand	11	8,691	0,790091	0,3588
Nigeria	4	3,7836	0,9459	0,003678
Norway	5	4,9799	0,99598	3,59E-06
Philippines	4	3,9093	0,977325	0,000985
Poland	21	18,4598	0,879038	0,155601
Portugal	3	0,7753	0,258433	0,455491
Romania	6	5,8496	0,974933	0,001063
Russia	162	144,6009	0,892598	0,103128
Russia	1	0,9856	0,9856	#DIV/0!
Saint Kitts and Nevis	16	15,7431	0,983944	0,000364

Saint Vincent and the Grenadines	3	1,3118	0,437267	0,303252
Samoa	6	5,9815	0,996917	3,99E-06
San Marino	5	4,9875	0,9975	4,54E-06
Serbia	2	1,9677	0,98385	0,000346
Seychelles	49	46,6838	0,952731	0,018067
Singapore	478	434,7219	0,90946	0,088594
Slovakia	3	2,8735	0,957833	0,004177
Slovenia	49	47,6491	0,972431	0,020298
Snigapore	3	2,368	0,789333	0,033968
South Africa	23	22,0924	0,960539	0,006901
South Korea	11	9,0733	0,824845	0,25178
Spain	11	10,2812	0,934655	0,006196
Swaziland	1	0,9972	0,9972	#DIV/0!
Sweden	5	4,9722	0,99444	4,32E-05
Switzerland	226	206,4955	0,913697	0,066533
Taiwan	5	4,6087	0,92174	0,02754
Tanzania	3	2,9547	0,9849	2,48E-05
Turkey	1	0,8986	0,8986	#DIV/0!
UK	294	255,7982	0,870062	0,163539
UK	17	16,4822	0,969541	0,012605
Ukraine	17	14,3954	0,846788	0,076986
United Arab Emirates	4	3,9647	0,991175	2,94E-05
United Arab Emirates	56	43,9203	0,784291	0,226955
USA	429	360,2738	0,839799	0,196892

ANALISYS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	21,28648	78	0,272904	2,312553	9,89E-10	1,281791
Inside Groups	381,5251	3233	0,11801			
Total	402,8116	3311				

Now the focus of the ANOVA analyses has shifted to Countries looking for correlations between specific ICOs' countries and the news sentiment related. Firstly it can be seen that there is statistical significance in the results, and looking at the means we can see that there quite high values for the vast majority of countries besides someone that have values below 0,7 and these are: Indonesia with a value of -0,52 showing kind of strong aversion to ICOs; the second lowest is Portugal with a value of 0,26; the third is Malaysia with a value of 0,39; the fourth is Saint Vincent and the Grenadines with a value of 0,44

and the last is Liechtenstein with a value of 0,51. Talking about our country Italy we see a very high value of sentiment of 0,998.

5.1.10 ANOVA BETWEEN COUNTRIES AND POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Afghanistan	2	7,42	3,71	0,9522
Anguilla	5	15,82	3,164	0,05823
Argentina	13	39,24	3,018462	1,969997
Australia	74	228,21	3,083919	1,277038
Austria	5	27,41	5,482	7,68197
Belgium	10	27,17	2,717	0,36689
Belize	52	178,89	3,440192	2,796473
Bermuda	1	2,88	2,88	#DIV/0!
Brazil	11	29,95	2,722727	0,476222
British Virgin Islands	50	169,5	3,39	1,645257
Bulgaria	12	33,88	2,823333	0,564497
Cambodia	10	36,1	3,61	0,644711
Canada	59	194,14	3,290508	2,68325
Cayman Islands	174	570,35	3,277874	1,855493
China	56	179,43	3,204107	1,534363
Colombia	1	4,17	4,17	#DIV/0!
Costa Rica	1	2,97	2,97	#DIV/0!
Croatia	1	1,51	1,51	#DIV/0!
Cyprus	18	63,2	3,511111	0,624963
Czech Republic	8	24,64	3,08	1,211686
Denmark	3	10,48	3,493333	0,566533
Estonia	183	585,31	3,198415	1,340413
Estonia	5	16,75	3,35	0,96725
France	80	306,35	3,829375	2,565054
Georgia	3	8,23	2,743333	0,168533
Germany	49	148,75	3,035714	0,94405
Gibraltar	107	334,32	3,124486	1,418217
Hong Kong	85	283,57	3,336118	3,739638
India	6	20,01	3,335	0,31379
Indonesia	12	20,12	1,676667	1,15777
Indonesia	2	0,24	0,12	0,0288
Ireland	7	20,16	2,88	0,909067
Isle of Man	10	38,89	3,889	0,844366
Israel	36	141,91	3,941944	1,418388
Italy	5	12,71	2,542	0,51562
Japan	27	94,9	3,514815	2,749687

Laos	10	24,49	2,449	1,174543
Liechtenstein	12	21,32	1,776667	1,223206
Lithuania	18	47,46	2,636667	1,693612
Luxembourg	14	45,4	3,242857	0,84133
Malaysia	23	27,61	1,200435	1,55515
Malta	53	180,77	3,410755	2,009492
Marshall Islands	6	23,19	3,865	1,27791
Mexico	11	35,5	3,227273	0,884822
Netherlands	48	163,32	3,4025	1,563074
New Zealand	11	35,43	3,220909	2,731349
Nigeria	4	10,75	2,6875	2,294425
Norway	5	11,34	2,268	0,13052
Philippines	4	10,4	2,6	1,134333
Poland	21	80,97	3,855714	1,552846
Portugal	3	6,37	2,123333	0,026133
Romania	6	27,49	4,581667	3,831777
Russia	162	511,18	3,155432	1,749559
Russia	1	3,22	3,22	#DIV/0!
Saint Kitts and Nevis	16	45,79	2,861875	0,626536
Saint Vincent and the Grenadines	3	1	0,333333	0,018433
Samoa	6	22,38	3,73	0,64732
San Marino	5	18,27	3,654	0,19513
Serbia	2	5,78	2,89	2,2898
Seychelles	49	152,22	3,106531	2,432986
Singapore	478	1489,89	3,116925	2,247601
Slovakia	3	7,65	2,55	0,6111
Slovenia	49	169,27	3,45449	1,32555
Snigapore	3	7,99	2,663333	0,120933
South Africa	23	66,04	2,871304	0,842157
South Korea	11	31,18	2,834545	1,855127
Spain	11	24,56	2,232727	1,534042
Swaziland	1	2,51	2,51	#DIV/0!
Sweden	5	14,22	2,844	0,61893
Switzerland	226	708,72	3,135929	1,670666
Taiwan	5	16,91	3,382	0,91337
Tanzania	3	5,05	1,683333	0,539233
Turkey	1	0,17	0,17	#DIV/0!
UK	294	961,59	3,270714	2,024514
UK	17	64,66	3,803529	2,202437
Ukraine	17	49,61	2,918235	3,75209
United Arab Emirates	4	10,54	2,635	0,271233
United Arab Emirates	56	133,06	2,376071	1,980577
USA	429	1382,85	3,223427	1,874551

ANALISYS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	445,7457	78	5,714688	3,05306	2,85E-17	1,281791
Inside Groups	6051,498	3233	1,87179			
Total	6497,244	3311				

In this case the ANOVA is useful to correlate the countries with the Positive Emotions factor. The results show statistical significance checking the F test and the countries with the highest values are: Austria with a mean of 5,48 but it also has the highest variance of 7,7 so the result is not so significative; the second is Romania with a mean of 4,58 and a variance of 3,83; the third is Colombia with 4,17, but this country has only one event so the result should be excluded; and fourth Israel with mean 3,94 and variance of 1,41, much lower than the others.

The lowest values instead are attributed to Indonesia, Turkey and Saint Vincent and the Grenadines, all with means below 0,34

5.1.11 ANOVA BETWEEN COUNTRIES AND FOCUSFUTURE

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Afghanistan	2	4,89	2,445	0,66125
Anguilla	5	10,09	2,018	0,66202
Argentina	13	17,01	1,308462	0,592264
Australia	74	99,28	1,341622	0,484589
Austria	5	9,38	1,876	0,21293
Belgium	10	14,34	1,434	0,15616
Belize	52	42,7	0,821154	0,480022
Bermuda	1	1,48	1,48	#DIV/0!
Brazil	11	18,01	1,637273	0,786242
British Virgin Islands	50	76,34	1,5268	0,683843
Bulgaria	12	17,59	1,465833	0,911899
Cambodia	10	13,9	1,39	0,042778
Canada	59	77,59	1,315085	0,582229
Cayman Islands	174	227,24	1,305977	0,60655
China	56	64,31	1,148393	0,406083
Colombia	1	2,54	2,54	#DIV/0!
Costa Rica	1	1,67	1,67	#DIV/0!

Croatia	1	2,51	2,51	#DIV/0!
Cyprus	18	34,97	1,942778	0,619892
Czech Republic	8	13,27	1,65875	0,458527
Denmark	3	4,38	1,46	0,2487
Estonia	183	241,6	1,320219	0,6595
Estonia	5	8,28	1,656	0,33208
France	80	116,11	1,451375	0,815425
Georgia	3	4,88	1,626667	0,025733
Germany	49	66,17	1,350408	0,80965
Gibraltar	107	153,42	1,433832	0,604397
Hong Kong	85	122,28	1,438588	0,595827
India	6	12,62	2,103333	1,369387
Indonesia	12	16,09	1,340833	0,352027
Indonesia	2	0,5	0,25	0,0002
Ireland	7	8,95	1,278571	0,108848
Isle of Man	10	23,92	2,392	3,26244
Israel	36	53,64	1,49	0,790766
Italy	5	5,44	1,088	0,71042
Japan	27	42,32	1,567407	0,826666
Laos	10	8,94	0,894	0,224093
Liechtenstein	12	17,5	1,458333	0,780633
Lithuania	18	17,35	0,963889	0,408955
Luxembourg	14	14,97	1,069286	0,589684
Malaysia	23	23,31	1,013478	1,082278
Malta	53	74,73	1,41	0,612027
Marshall Islands	6	12,25	2,041667	0,490137
Mexico	11	17,36	1,578182	0,526136
Netherlands	48	63,32	1,319167	0,646935
New Zealand	11	13,28	1,207273	0,414442
Nigeria	4	6,68	1,67	0,720133
Norway	5	5,24	1,048	0,25992
Philippines	4	13,77	3,4425	0,959425
Poland	21	29,59	1,409048	0,484769
Portugal	3	2,84	0,946667	0,070233
Romania	6	9,67	1,611667	0,255857
Russia	162	258,91	1,59821	1,116564
Russia	1	0,48	0,48	#DIV/0!
Saint Kitts and Nevis	16	27,42	1,71375	0,508052
Saint Vincent and the Grenadines	3	0,75	0,25	0,0001
Samoa	6	6,9	1,15	0,27692
San Marino	5	6,54	1,308	0,01332
Serbia	2	4,99	2,495	6,23045
Seychelles	49	59,97	1,223878	0,568108
Singapore	478	612,97	1,282364	0,69185
Slovakia	3	2,84	0,946667	0,262033

Slovenia	49	70,79	1,444694	0,556663
Snigapore	3	1,88	0,626667	0,131733
South Africa	23	29,1	1,265217	0,147981
South Korea	11	15,31	1,391818	0,646236
Spain	11	16,94	1,54	1,0758
Swaziland	1	1,48	1,48	#DIV/0!
Sweden	5	6,43	1,286	0,32458
Switzerland	226	321,16	1,421062	0,732594
Taiwan	5	7,15	1,43	0,99145
Tanzania	3	4,16	1,386667	0,548433
Turkey	1	0,17	0,17	#DIV/0!
UK	294	411	1,397959	0,804548
UK	17	22,15	1,302941	0,944572
Ukraine	17	28,84	1,696471	1,332624
United Arab Emirates	4	3,56	0,89	0,065467
United Arab Emirates	56	62,74	1,120357	0,653298
USA	429	495,09	1,154056	0,517696

ANALISYS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	140,7476	78	1,804456	2,684865	2,15E-13	1,281791
Inside Groups	2172,849	3233	0,672085			
Total	2313,597	3311				

This analysis demonstrates the correlation between countries and presence in the news related of a focus on future time horizon. The results are statistically significant checking the F test. The highest values are referred to: Philippines with a mean of 3,44 and variance of 0,96; Colombia and Croatia with means of 2,54 and 2,51, but both have only one event; Serbia with a mean of 2,495 but variance of 6,23 that makes the result too doubtful; and Afghanistan with value of 2,445 and variance of 0,66.

The lowest values in biggest presence (counting events) are related to Lithuania, Laos and Belize.

5.1.12 ANOVA BETWEEN COUNTRIES AND FOCUSPRESENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
Afghanistan	2	12,14	6,07	0,3528
Anguilla	5	33,17	6,634	0,50043

Argentina	13	94,32	7,255385	2,467594
Australia	74	556,97	7,526622	4,511609
Austria	5	29,66	5,932	1,56347
Belgium	10	65,54	6,554	0,694382
Belize	52	341,9	6,575	4,044116
Bermuda	1	7,15	7,15	#DIV/0!
Brazil	11	71,5	6,5	1,60282
British Virgin Islands	50	329,26	6,5852	2,525389
Bulgaria	12	101,1	8,425	1,6155
Cambodia	10	79,18	7,918	3,516551
Canada	59	384,95	6,524576	3,85448
Cayman Islands	174	1146,24	6,587586	3,607755
China	56	334,41	5,971607	3,201476
Colombia	1	12,5	12,5	#DIV/0!
Costa Rica	1	9,48	9,48	#DIV/0!
Croatia	1	7,71	7,71	#DIV/0!
Cyprus	18	133,25	7,402778	4,193233
Czech Republic	8	60,34	7,5425	2,632593
Denmark	3	16,78	5,593333	4,165733
Estonia	183	1223,95	6,688251	3,91363
Estonia	5	29,7	5,94	4,11135
France	80	503,88	6,2985	2,576605
Georgia	3	21,38	7,126667	0,766033
Germany	49	338,41	6,906327	5,14607
Gibraltar	107	748,14	6,991963	3,861335
Hong Kong	85	612,07	7,200824	5,300272
India	6	42,81	7,135	0,76927
Indonesia	12	66,17	5,514167	3,896099
Indonesia	2	14,7	7,35	0,0512
Ireland	7	49,17	7,024286	3,406662
Isle of Man	10	68,78	6,878	1,885062
Israel	36	268,78	7,466111	5,07707
Italy	5	25,29	5,058	2,26657
Japan	27	175,34	6,494074	3,490702
Laos	10	88,23	8,823	5,136357
Liechtenstein	12	70,47	5,8725	3,950893
Lithuania	18	128,76	7,153333	3,570847
Luxembourg	14	97,67	6,976429	3,795332
Malaysia	23	147,92	6,431304	1,658248
Malta	53	354,12	6,681509	2,729028
Marshall Islands	6	35,98	5,996667	0,761947
Mexico	11	85,05	7,731818	5,392916
Netherlands	48	348,47	7,259792	3,366368
New Zealand	11	80,22	7,292727	4,857762
Nigeria	4	25,73	6,4325	4,432625
Norway	5	41,59	8,318	2,63727

Philippines	4	26,9	6,725	0,267833
Poland	21	151,11	7,195714	6,710936
Portugal	3	15,51	5,17	0,6604
Romania	6	35,64	5,94	5,9598
Russia	162	1114,24	6,878025	4,398266
Russia	1	7,73	7,73	#DIV/0!
Saint Kitts and Nevis	16	123,16	7,6975	4,641273
Saint Vincent and the Grenadines	3	20,28	6,76	0,0259
Samoa	6	34,18	5,696667	0,814587
San Marino	5	28,06	5,612	1,48312
Serbia	2	13,9	6,95	0,2312
Seychelles	49	335,33	6,843469	3,945044
Singapore	478	3232,79	6,763159	3,544603
Slovakia	3	21,56	7,186667	0,896633
Slovenia	49	344,69	7,03449	4,684829
Snigapore	3	19,38	6,46	1,6996
South Africa	23	140,42	6,105217	4,052744
South Korea	11	61,64	5,603636	3,470005
Spain	11	66,71	6,064545	3,337427
Swaziland	1	7,18	7,18	#DIV/0!
Sweden	5	46,34	9,268	11,55702
Switzerland	226	1577,95	6,98208	3,606893
Taiwan	5	26,47	5,294	2,31963
Tanzania	3	15,52	5,173333	0,227433
Turkey	1	5,89	5,89	#DIV/0!
UK	294	2029,65	6,903571	3,28736
UK	17	108,52	6,383529	3,295549
Ukraine	17	91,07	5,357059	2,734697
United Arab Emirates	4	25,27	6,3175	3,337292
United Arab Emirates	56	380,89	6,801607	2,048534
USA	429	2742,66	6,393147	5,163868

ANALISYS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	646,7279	78	8,291384	2,155479	2,76E-08	1,281791
Inside Groups	12436,23	3233	3,846654			
Total	13082,96	3311				

In a similar way as before this analysis demonstrates the correlation between countries and presence in the news related of a focus on present time horizon. The results show statistical significance due to the F test.

The highest values are referred to Colombia and Costa Rica but they have only one registration event. The other higher values more significant are Sweden, Laos and Bulgaria with means of 9,26 but variance of 11,55, 8,82 but variance of 5,13 and 8,42 and a much lower variance of 1,61.

The lowest mean values are related to Italy with a result of 5,058 and variance of 2,26; Portugal with mean of 5,17 and variance of 0,66; and Tanzania with mean of 5,173 and variance 0,22.

5.1.13 ANOVA BETWEEN SOURCE CODE AND SENTIMENT AND POSEMO

In the following analyses it has been chosen to value the relation of some ICO dummy variables to the variables Sentiment and Posemo because are the ones which better explain the investors feelings in the quality perception of the business given by the presence of the most important ICO characteristics detected by the literature.

SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	1571	1407,187	0,895727	0,109929
1	1796	1598,163	0,889846	0,12855

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	0,028986	1	0,028986	0,241831	0,622919	3,844224
Inside Groups	403,3362	3365	0,119862			
Total	403,3652	3366				

POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	1571	4933,41	3,140299	1,823513
1	1796	5799,34	3,229031	2,072687

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	6,597813	1	6,597813	3,372373	0,066386	3,844224
Inside Groups	6583,389	3365	1,956431			
Total	6589,986	3366				

The presence of source code is a dummy variable and is related to a 1 if present and 0 is absent.

The aim of these ANOVA is to check for a correlation between specific factors of the ICOs, like in this case the availability of source code, and the sentiment and positive emotions that this factor might affect the news. In this case though both the ANOVA analyses show that there is not statistical significance, meaning that the two groups' mean are about the same, so the presence or absence of source code does not affect positively or negatively the sentiment or positive emotions in the news.

5.1.14 ANOVA BETWEEN TEAM DISCLOSURE AND SENTIMENT AND POSEMO

SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	117	111,2129	0,950538	0,035502
1	3250	2894,138	0,890504	0,122758

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	0,407021	1	0,407021	3,398927	0,065327	3,844224
Inside Groups	402,9582	3365	0,11975			
Total	403,3652	3366				

POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	117	373,3	3,190598	1,426535
1	3250	10359,45	3,187523	1,97738

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	0,001068	1	0,001068	0,000545	0,98137	3,844224
Inside Groups	6589,985	3365	1,958391			
Total	6589,986	3366				

Also in this case the presence of team disclosure is a dummy variable and is related to a 1 if present and 0 is absent. And again, as before the results are both not statistically significant because both the F test are lower than their respective F critical. This means that also the Team Disclosure presence or absence does not affect the news' sentiment and positive emotions.

5.1.15 ANOVA BETWEEN PRE-ICO PHASE AND SENTIMENT AND POSEMO

SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	1706	1513,952	0,887428	0,128315
1	1661	1491,398	0,897892	0,111142

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	0,092148	1	0,092148	0,7689	0,38062	3,844224
Inside Groups	403,273	3365	0,119843			
Total	403,3652	3366				

POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	1706	5315,97	3,116043	1,711234
1	1661	5416,78	3,261156	2,201573

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	17,72211	1	17,72211	9,073724	0,002612	3,844224
Inside Groups	6572,264	3365	1,953125			
Total	6589,986	3366				

In these ANOVA analyses it is studied the relation between the presence of a Pre-ICO phase, recorded as a dummy variable, and the news' sentiment and positive emotions factors. The first analysis related to the sentiment does not show statistical significance, resulting in a non-difference between groups' mean and so not affecting the sentiment. While for the second one on positive emotions the results are statistically significant. More precisely, the presence of a Pre-ICO phase increases the mean of Positive Emotions factor in the news by 0,15 which is a 5% more than the mean of ICOs without this characteristic.

5.1.16 ANOVA BETWEEN BONUS PROGRAM AND SENTIMENT AND POSEMO

SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	1908	1667,731	0,874073	0,14715
1	1459	1337,62	0,916806	0,083155

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	1,509815	1	1,509815	12,64268	0,000382	3,844224
Inside Groups	401,8554	3365	0,119422			
Total	403,3652	3366				

POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	1908	5898,37	3,091389	1,822528
1	1459	4834,38	3,313489	2,108121

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	40,78372	1	40,78372	20,9548	4,87E-06	3,844224
Inside Groups	6549,203	3365	1,946271			
Total	6589,986	3366				

Consequently, this ANOVA shows the impact of presence or absence of a Bonus Program in the ICO phase on the news' sentiment and positive emotions factors. Both the results show statistical significance leading to the conclusion that the Bonus Program affects the news' factors. In particular, the presence of a Bonus Program increases the mean of Sentiment by about 0,04 (4,6% of total) and the mean of Posemo by about 0,22 (7,1% of total).

5.1.17 ANOVA BETWEEN MVP PRESENCE AND SENTIMENT AND POSEMO

SENTIMENT

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	2629	2327,661	0,885379	0,129118
1	738	677,6894	0,918278	0,086052

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	0,623707	1	0,623707	5,211219	0,022504	3,844224
Inside Groups	402,7415	3365	0,119685			
Total	403,3652	3366				

POSEMO

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Mean</i>	<i>Variance</i>
0	2629	8193,17	3,116459	1,756886
1	738	2539,58	3,441165	2,594482

ANALYSIS OF VARIANCE

<i>Variation Origin</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Significant Value</i>	<i>F crit</i>
Between Groups	60,75556	1	60,75556	31,31187	2,37E-08	3,844224
Inside Groups	6529,231	3365	1,940336			
Total	6589,986	3366				

Very similarly to before also this ANOVA shows statistical significance in the results and quite the same impact on the news' factors.

The objective was to check for correlations between presence of an MVP, recorded as a dummy variable, in the ICO phase and the sentiment and positive emotions factors in the news related.

The first result shows that the presence of an MVP increases the mean of Sentiment by 0,033 (3,7% of the total) and the second result shows that the presence of a MVP increases the mean of Posemo by 0,325 (10% of the total).

5.2 CORRELATIONS BETWEEN NEWS FACTORS AND CAPITAL RAISED

In this paragraph it is shown the correlation between the most relevant news' factors and the dependent variable chosen as proxy of ICOs "success" that is Capital Raised. It has been chosen to proceed with correlations and not ANOVA because the latter has only meaning with groups of categorical variables like the ones presented in the paragraphs before.

A first measure to be implemented when looking for link between two continuous variables is to perform the correlation analysis.

Correlation analysis is a statistical evaluation method used to study the strength of a relationship between two continuous variables, measured numerically. This analysis is useful when wanting to establish if there are possible connections between variables.

Generally, if a correlation is found between two variables, it means that when there is a systematic variation in one variable, there is also a systematic variation in the other; the variables change together over a period of time.

5.2.1 PEARSON CORRELATION

A very common correlation is that of Pearson, to measure the degree of the relationship between linearly correlated variables. In this case, the correlation between the variables can assume a value included in the interval -1 and +1. A correlation coefficient equal to 1 indicates that for each positive increase in a variable, there is a positive increase of a fixed proportion in the other. A correlation coefficient of -1 means that for each positive increase in a variable, there is a negative decrease of a fixed proportion in the other.

Zero means that for each increase there is no positive or negative increase. The two variables are therefore not linked.

The formula for calculating the Pearson correlation coefficient is the following:

$$r = \frac{\sum_i (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Equation 2 - Pearson correlation

In order to do this kind of analysis Excel has been used, in particular the formula CORRELATION(). This method as though some drawbacks, mainly the fact that can be done only between 2 set of variables, so if wanting to look for the correlation of several factors on a dependent variable, in this way they are not considered as mutually existing in the correlation, so the parameters of correlation only reflect the univocal existence of the variable taken in consideration.

5.2.2 CORRELATION PARAMETERS

This paragraph shows the correlation parameters between the dependent variable Capital Raised and a few news' factors chosen to be the most explicative and relevant.

The ones picked are:

- WordCountOriginal
- Lexicaldiversity
- Posemo
- Negemo
- FocusPast
- FocusPresent
- FocusFuture
- ComplexityTFIDFavg

Following it is presented the table of factors related to their respective correlation parameter on the Capital Raised and a graph better showing the impact.

	Capital Raised
WordCountOriginal	-0,005752092
Lexicaldiversity	0,04066283
Posemo	0,027071697
Negemo	0,009139108
FocusPast	0,069570735
FocusPresent	-0,080164558
FocusFuture	-0,048017146
ComplexityTFIDFavg	0,01003918

Table 1 - Correlation parameters

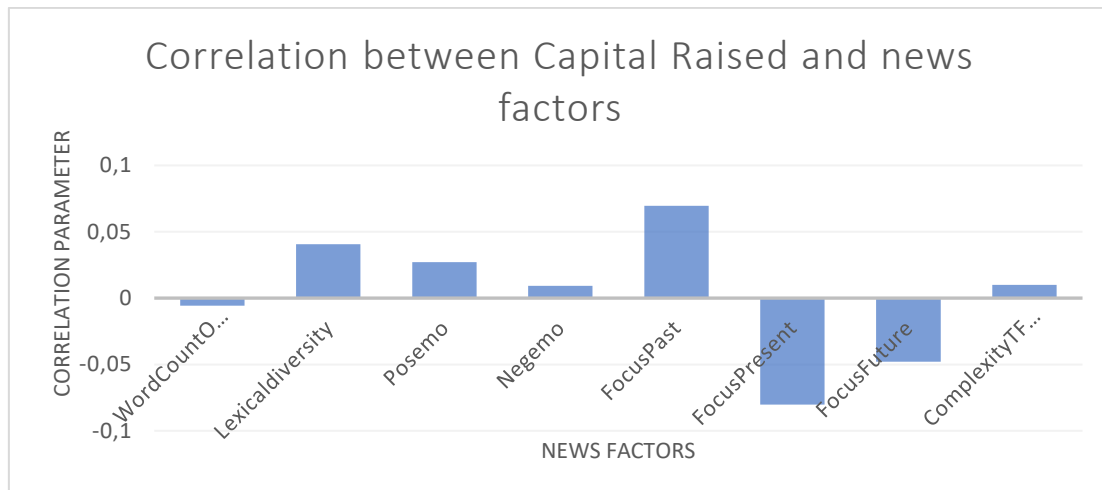


Figure 6 Correlation parameters

From the results it is understandable the impact of each factor on the “success” of an ICO without considering the mutual existence of the other factors in the correlation model.

The first factor taken in consideration is the number of words counted in the text and this negatively related to the “success” of a token sale. The impact is though quite limited since the correlation parameter is -0,00575, meaning that a unitarian increase in the Word-Count leads to a decrease of the amplitude of the parameter on the Capital Raised.

The second one is the lexical diversity in the news text, this is positively related to the dependent variable instead. Looking at the graph it is possible to see that the impact is one of the strongest, reflecting in a correlation parameter of 0,0406. This result induce to think that more sophisticated news lead the investors to think about the related ICO as a better quality one.

The third one is the positive emotions that the news’ text make emerge to a reader. Here we can see a positive impact but not so prominent, reflected in a correlation parameter of 0,027. The outcome is that the more positive emotions a text make emerge the more capital the project is likely to raise, this is understandable thanks to the pattern, seen in the first chapter, that ICO investors tend to show, investing in projects that follow a wave of good expectations, made by the other investors.

The fourth parameter in consideration is the one related to negative emotions that a news' text make emerge. This is still positively correlated with the Capital Raised but in a much lower magnitude. The positiveness of the correlation factor could be explained to a possible investors' behavior driven by buying something "un-wanted" by the market at a discount because the other players in the market do not see the quality of the asset that, instead, the buyer sees.

The fifth variable is the Focus on the Past, this is strongly positively correlated with the capital raised and a potential explanation could be that investors assess the quality of a business relying in the highest percentage on past performances, maybe related to the ICO team members, so if there is historical track of the past the investors are more likely to invest in that business.

Following the same conceptual scheme, the 6th and 7th variables and strongly negatively correlated with the capital raised, this because they both are based on the focus of present and future, that by definition is related to much more uncertainty, getting away the interest of the investors perceiving a bigger presence of risk related to that business.

The last variable is the complexity of the news texts, and this is slightly positively correlated with the capital raised. This follows the path already recognized on whitepapers by previous authors (Samieifair & Baur, 2020) who showed that the complexity of whitepapers, in this case, lead to higher amount raised, telling us that investors perceive projects with more complex whitepaper as more successful. It could be said that the same is happening for the news' complexity.

5.2.3 CONCLUSIONS

The analysis performed in this chapter have been carried out in order to detect which were the most significant differences in the news' variables means given by the different ICO categories, countries of origin, and other dummy variables, explaining a statistically significant difference between the groups.

The results are multiple, firstly it has been conducted a research on the ICO categorical variables Categories and Countries to look for significant differences in the news variables in different ICO categories and countries.

Then the analyses have been carried on the best ICO “success” indicators recognized by the literature. The results show that the presence of a Pre-ICO phase, MVP availability and Bonus Program are related to higher values of news indicators like positive emotions and sentiment explaining a better perception of the quality of the business. On the other hand, others ICO “success” factors detected by the literature like Source code availability and Team disclosure do not show any statistically significant result in relation to news variables.

Lastly, a correlation analysis on the most important “success” factor, which is the Capital raised, has been conducted on the news’ indicators showing interesting relations that induce to think that some news’ characteristics influence the investors behaviour and lead to invest more in token sales with specific related news factors, like the Focus on the past, the Lexical diversity and the Positive emotions.

6 AVENUES FOR FUTURE RESEARCH

The Initial Coin Offerings and related blockchain system inevitably are a very powerful financial tool to raise funds in a fast and in a secure way, but the literature about these topics is at a very early stage and many more aspects of it must be studied.

Given, as said before, the great strength of ICOs, there are still a substantial number of issues starting from the very low protection investors receive, this can lead to future research about regulation, answering the question if there could be a type of ICO regulation that doesn't affect negatively the development and spreading of this financing channel technology.

An interesting question is how and to what degree should regulation adapt to this changing entrepreneurial finance environment?

About this another hint for future study could be to understand if the introduction of a 'crypto underwriter' can help. Or analyse if the development of standards could bring more security and stability to the system.

Another quite big problem is the very low amount of information disclosed by token sellers, these firms are not required to disclose any document about their business and founders, all the information investors have is because ICO issuers voluntary do so and buyers have to rely on them.

This leads sellers to spread exaggerated or false information, it would be interesting to understand how can be possible to reduce as much as possible information asymmetry and moral hazard behaviours.

Moreover, the correlation between daily ICO returns and the lag of daily ICO returns is negative, suggesting that ICOs are highly volatile and that they are speculative, high-risk investments.

So interesting questions, source of future analysis, could be: How do ICO investors value firms? How do they choose which firm to invest in? Confronting them with investors in traditional financing system, how do they differ?

It might be relevant to address future studies to understand which firms really succeeded in developing what promised, not only raising requested capital, and understand what factors determine this.

Related to this, a possible factor to consider is to analyse if firms with high technological capabilities are more able in developing their product. While most ventures intend to develop a product, ICO ventures often do not have a working product at the time of the ICO so it is interesting to see how the funding predictors and the funding raised in ICOs relate to future performance. Are ICOs that receive high amounts of funding going to outperform those that receive less?

Subsequently, in my opinion it is important to understand what fraction of ICOs survives in the long run and how their token prices evolve, so future research could be on detecting which are the ICOs that survive and comprehend the common factors that characterize them.

Finally, another crucial point can be, given ICOs similarities with crowdfunding, seeing how investors and entrepreneurs can relate and communicate, eventually obtaining information from the crowd and use it to improve products or innovation.

7 BIBLIOGRAPHY

- academy.binance.com*. (s.d.). Tratto da <https://academy.binance.com/en/articles?page=1&tags=bitcoin>
- Adhami, S. (2018). Why do businesses go crypto? An empirical analysis of initial coin.
- Adhami, S., Giudici, G., & Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of initial coin.
- Alpers, G. W., Winzelberg, A., C., C., & H.D., R. (2005). . Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior*.
- AU10TIX*. (s.d.). Tratto da <https://www.au10tix.com/>
- Bellavitis, C. (2020). Ban, Boom, and Echo! Entrepreneurship and initial coin offerings.
- Bellavitis, C., Cumming, D., & Vanacker, T. (2020). Ban, Boom, and Echo! Entrepreneurship and Initial Coin Offerings.
- Block, J., Groh, A., Hornuf, L., & Vismara, S. (2020). The entrepreneurial finance markets of the future:.
- Boreiko, D., & Risteski, D. (2020). Serial and large investors in initial coin offerings.
- Buterin, V. (2013). A next generation smart contract & decentralized application platform. *Ethereum Whitepaper*.
- Domingo, R.-S., Pineiro-Chousa, J., & Lopez-cabarcos, A. (2020). What factors drive returns on initial coin offerings?
- Fisch, C. (2019). Initial coin offerings (ICOs) to finance new ventures.
- Fisch, C., Masiak, C., Vismara, S., & Block, J. (2020). Motives and profiles of ICO investors.
- Florisyak, D., & Schandlbauer, A. (2020). The Information Content of ICO White Papers.
- Huang, W. (2019). The geography of initial coin offerings.
- Huang, W., Meoli, M., & Vismara, S. (2020). The geography of initial coin offerings.
- IcoBench*. (s.d.). Tratto da www.lobench.com
- Kaal, W. (2018). Initial Coin Offerings: Emerging Practices, Risk Factors, and Red Flags.
- Kahn, Tobin, R., Massey, A., & J.A., A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count.
- Kher, R., Terjese, S., & Liu, C. (2019). Blockchain, Bitcoin, and ICOs: a review and research.
- Masiak, C. (2020). Initial coin offerings (ICOs): market cycles and relationship.
- Masiak, C., & Block, J. (2020). Initial coin offerings (ICOs): market cycles and relationship.
- Matthias Lischke, B. F. (2016). Analyzing the Bitcoin Network: the first four years.

- Momtaz, P. (2020). CEO Emotions and Firm Valuation in Initial Coin Offerings: An Artificial Emotional Intelligence Approach.
- Momtaz, P. (2020). Entrepreneurial Finance and Moral Hazard: Evidence from Token.
- Momtaz, P. (2020). Initial coin offerings.
- Nakamoto, S. (2008). Bitcoin: a peer-to-peer electronic cash system.
- Pennebaker, J., & Tausczik, Y. (s.d.). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.
- Perez, C., Sokolova, K., & Konate, M. (2020). Digital social capital and performance of initial coin offerings.
- Russell, M. (2013). Russell, M. A. (2013). Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More.
- Saiedi, E., Brostrom, A., & Ruiz, F. (2020). Global drivers of cryptocurrency infrastructure adoption.
- Sameeh, T. (2018). ICO basics - the difference between security tokens and utility tokens. *Cointelligence*.
- Samieifair, S., & Baur, D. (2020). Read me if you can! An analysis of ICO white papers.
- Schuckes, M., & Gutmann, T. (2020). Why do startups pursue initial coin offerings (ICOs)? *Unlock*. (s.d.). Tratto da unlock-bc.com
- Vaezi, S., & Kafshgar, S. (2012). Learner characteristics and syntactic and lexical complexity of written products. *International Journal of Linguistics*.
- www.coindesk.com. (s.d.).
- www.investopedia.com. (s.d.).
- www.kdnuggets.com. (s.d.). Tratto da www.kdnuggets.com.
- www.kinsta.com. (s.d.).
- www.lexisnexis.com. (s.d.).
- www.medium.com. (s.d.).

