POLITECNICO DI TORINO

STEPHANIE ESCORCIO FRANKE

Modelling startup success: The role of venture capital financing and business dynamism

Torino

2022

ABSTRACT

Startups are important innovation engines that have your importance greatly recognized due to its role in digital transformation. Understanding the factors that influence their success is of interest to founders, investors and policymakers. Being a type of investment focused on high growth and potential profitability, venture capitalists invest in startups believing in their ability to grow and generate returns. In order to map the ecosystem and improve business understanding, semi-structured interviews were conducted with different players from the venture capital industry. A logistic regression modelling has been developed by the use of data collected and pre-processed from Crunchbase for 15 countries. After variable selection and exploratory data analysis, a model has been fitted to the response variable of startup success. The role of business dynamism and its influence across countries has been assessed too. The main conclusion is that countries with higher business dynamism present a more competitive landscape for venture capital-backed startups, startup raising a series D investment round increases the odds of startup's success, but raised amount did not reveal statistically significant.

Keywords: Venture capital. Startups. Business dynamism. Logistic regression

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

VC	Venture Capital
CVC	Corporate Venture Capital
ICT	Information and Communication Technology
PE	Private Equity
AIFI	Associazione Italiana del Private Equity, Venture Capital e Private Debt
VeM	Venture Capital Monitor
PCA	Principal Component Analysis
MVA	Multivariate Analysis
EDA	Exploratory Data Analysis
IQR	Inter-Quartile Range
HW	Higher Whisker
LW	Lower Whisker

SUMMARY

1 INTRODUCTION

According to Blank (2010), a startup is a new and temporarily established venture founded to search for a repeatable and scalable business model, based on the pillars of invention, innovation and interaction, within a scenario of extreme uncertainty. It can reframe an existing market, exploring a niche that incumbents are not properly supplying, or it can promote a lower cost alternative to an existing solution, or it can even create a new market from scratch. They have become even more important due to digital transformation, i.e., business change triggered and shaped by the wide-spread adoption of digital technologies. It can either create new businesses or modify existing ones, capitalizing on new opportunities for spawning new forms of entrepreneurship and requires organization and cultural change in order to meet changes in behavior and market demands as a whole (NAMBISAN, WRIGHT e FELDMAN, 2019). Also, the COVID-19 pandemic has brought challenges that increased awareness of organizations towards urgent digital transformation (MCKINSEY, 2020). In this context, startups play an important role in the search for solutions that speed up that process, both by creating disruptive innovations or even making traditional organizations rethink their business models in order to take advantage of new opportunities.

However, startups are subjected to uncertainties derived from its dynamic and high-risk environment (BONAVENTURA, CIOTTI, et al., 2020). They may face difficulties, especially in high-technology industries, due to information asymmetry between startup founders and investors, either because entrepreneurs hold information regarding the company's future prospects (SHANE e STUART, 2002), or the limited access entrepreneurs have about the decision process taken by potential investors (DESSEIN, 2005), which allows opportunistic behavior by both sides.

In consonance with the obstacles aforementioned, the startups' access to conventional financing sources is limited due to banks risk-aversion. Additionally, due to the lack of asset tangibility, data to assess its operational history and stable cash flows, innovative startups have to rely mostly on external equity financing (HALL e LERNER, 2010) (MEGGINSON, MELES, et al., 2016). These financial difficulties may be even harsher when there are innovation activities involved as reinforce information asymmetry between entrepreneurs and investors, making funding decisions riskier (SAVIGNAC, 2008).

One of the most significant sources of equity financing is venture capital (VC), which comprises financial resources supplied to startup and small businesses that present potential for long-term growth. It can solve the financial constraints faced by those young firms, that can reach noticeable size and market position, by fulfilling its fund requirement (GOMPERS e LERNER, 2004). In exchange for the high risk investors have to bear in this situation, as only a few of the VC-backed companies evolve to successful and highly profitable businesses, they usually get equity stake and participation rights in those firms (HALL e LERNER, 2010). Indeed, this type of funding provides VCs decision-making power and managerial influence over the founder, coaching the startup through their lives while targeting exits with positive returns (HELLMANN e PURI, 2000).

Startups have drawn extensive attention in the past decade from both private investors and policy makers. Governments in many countries have implemented policies and further initiatives aimed at promoting entrepreneurial growth (CUMMING, SAPIENZA, *et al.*, 2009) in order to attract and retain promising startups that may contribute to economic growth, job creation and technological advancements (BUSSBANG, MONTUORI e BRAH, 2019). For example, several countries and regions have invested in the creation and maintenance of business incubators, in the offering of subsidized VC and in the implementation of attractive tax policies (CUMMING e FISCHER, 2012). Believing that VC stimulates innovation activities of the companies they invest on, a considerable number of policy-makers attempt to create and expand its local VC industry (HIRUKAWA e UEDA, 2011).

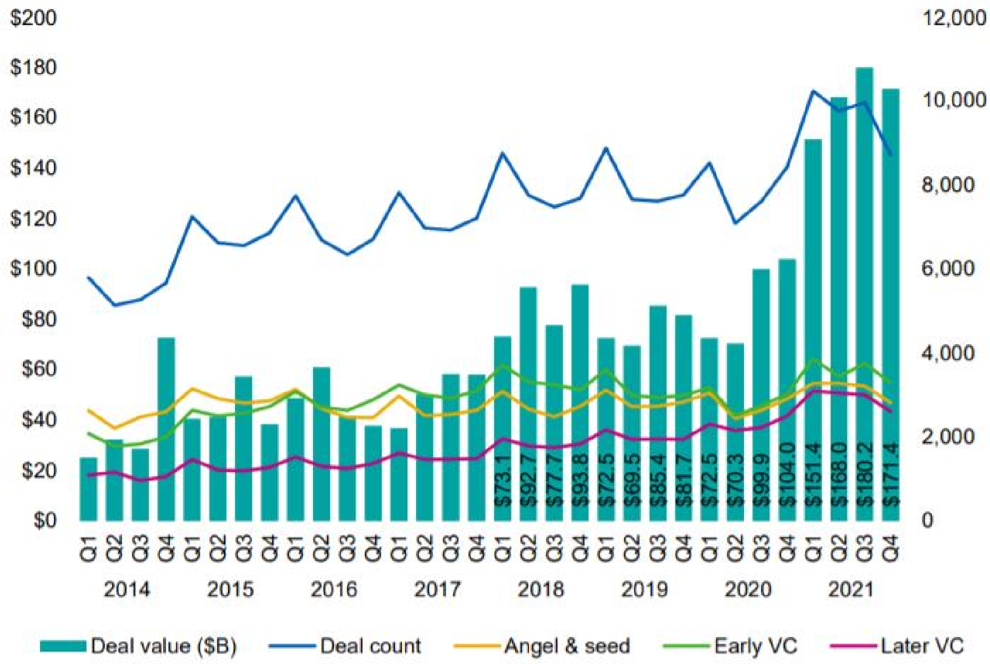
1.1 CONTEXTUALIZATION

Contrary to what was claimed by several business leaders in 2020, who feared Covid-19 would undermine startup ecosystems, new policies and financial support have been provided by governments. Furthermore, if venture capital flows were hesitant at first, they soared as a result of the new opportunities the pandemic has brought (STARTUP GENOME, 2021).

Taking into account both the investment volume and VC deals numbers seen globally, 2021 was the year in which the VC industry registered records, in the US, Canada, Brazil, the UK, Germany, Israel, Ireland, India, and the Nordic region. In accordance with the Venture Pulse, a quarterly report that analyzes the latest global trends in venture capital investment data, released by KPMG (2021), the total volume of investment received worldwide by startups throughout the full year reached US$ 671 billion across 38644 deals, surpassing the same period result of 2020 by 94% (Figure 1), with 7021 more deals.

As stated in the aforementioned report, within the “Angel & seed” stage, rounds classified as angel are the ones either without any private equity (PE) or VC firms participating or only mentioned as individuals making investments in a financing. Therefore, when it comes to the VC industry, only seed rounds are relevant.

Figure 1 – Global venture financing over time (2014-2021)



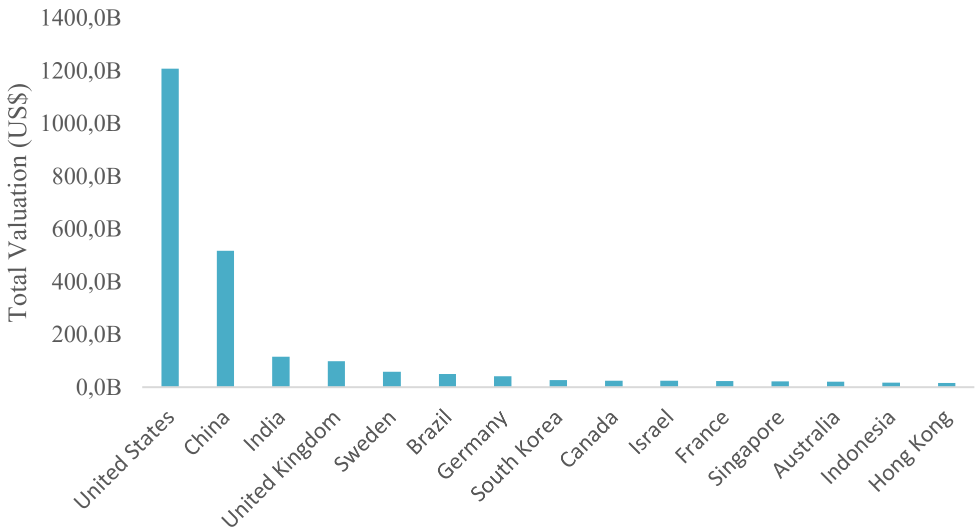
Source: Venture Pulse, Q4’21. Global Analysis of Venture Funding, KPMG.

Globally, a diverse number of market segments have attracted VC investors in the last year, ranging from fintechs, B2B services, healthtechs, biotechs, mobility, cybersecurity, to autotech. According to Bain (2021), technology has been a driving force behind the boom in venture investments from 2010 through 2021, consistently receiving the majority of venture funding provided and nearly doubling in the first quarter of 2021, period in which tech startups accounted for approximately 70% of total investment.

In alignment with the growth of the VC’s industry, new startups are emerging all the time with disruptive solutions and gaining capital to grow with scalability. In addition to that, the number of so-called “unicorns” has dramatically increased, turning 2021 into a record year for companies receiving this status when valued at US$1 billion or more. Between October 2020 and June 2021 the number of unicorns rose 43%, reaching the result of more than 800 startups around the world with valuation above this threshold in August 2021. Based on data collected from CB Insights, the total worldwide startups' valuation is illustrated across countries (Figure 2) and industries (Figure 3).

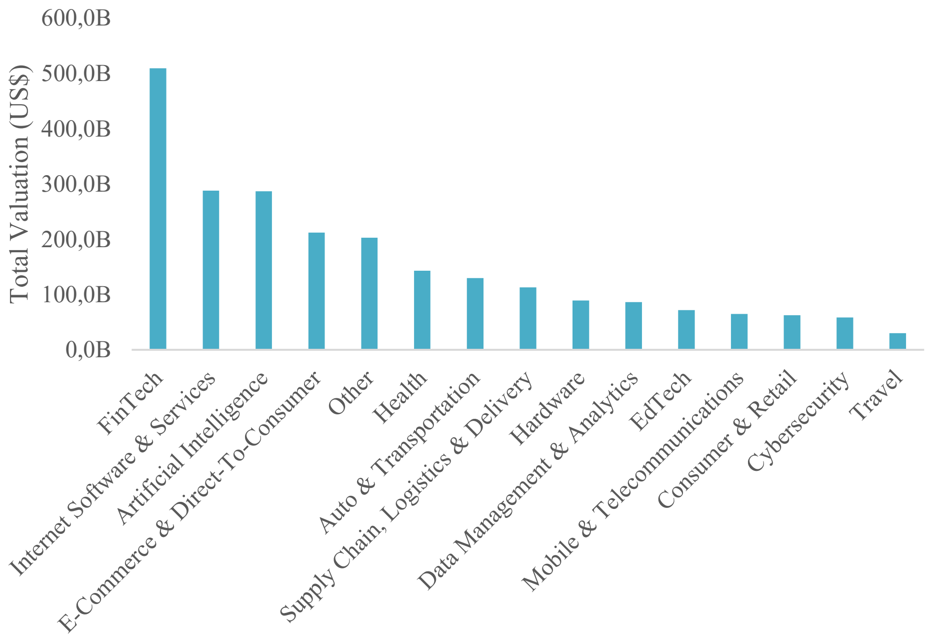
In pandemic years, there were some changes in the degree of maturity observed in some industries, measured by Startup Genome through the count of series A deals and global exits achieved, classifying them into declining, maturity or growth phases. For instance, Edtech and Gaming were both on the edge of decline phase in 2019, and now they have experienced a rise in series A funding, what caused them to return to the mature phase. Similarly, Fintech has received a higher number of series A investments in recent years, going from mature to growth phase.

Figure 2 – Startups valuation by country as of June-2021



Source: adapted from CB Insights, 2021.

Figure 3 – Industry-wise startup total valuation as of June-2021



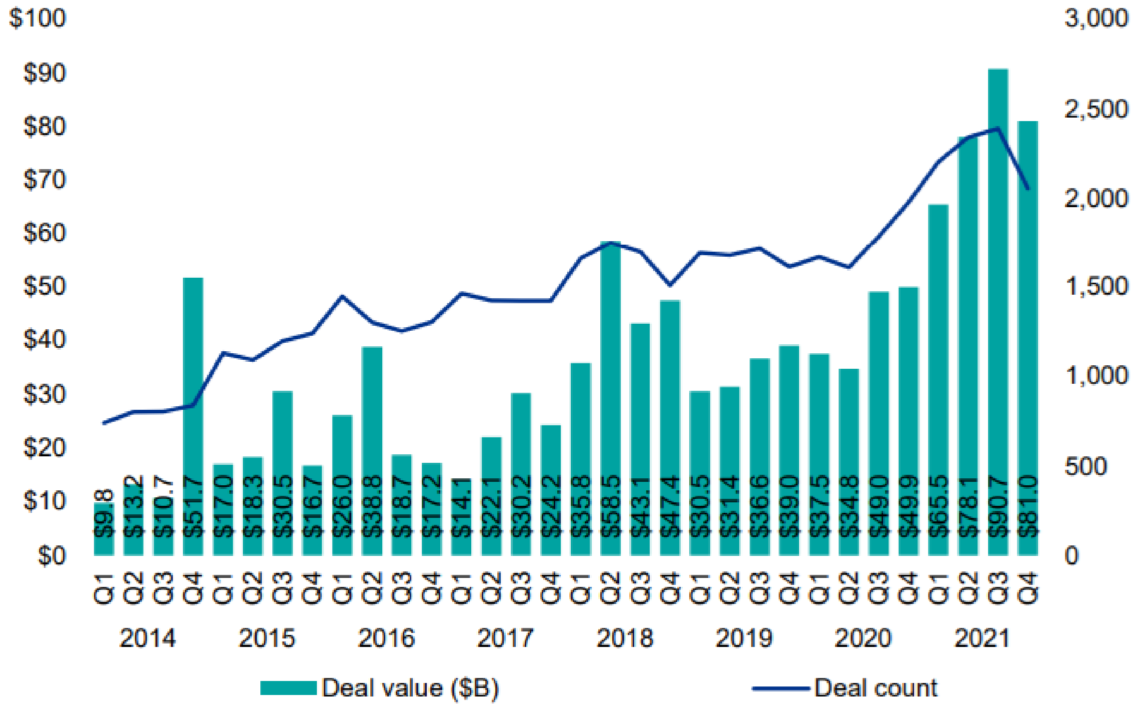
Source: adapted from CB Insights, 2021.

As a result of this study, it was found that industries stated in the growth phase are increasing in early-stage funding deals at the rate of 107% over five years. Meanwhile, those in the mature phase are growing at 33% and decline phase industries are dropping by 28%, over the same period. It is also highlighted that the deep tech startups (AI & Big Data, Advanced Manufaturing & Robotics, Blockchain, Agtech & New Food) have been the fastest growing ones and that Adtech and Digital Media industries underperformed compared to other ones.

Large companies are also taking advantage of this situation, as they understand the strategic and financial benefits for them of investing in startups, proving to be a solution for those companies looking to keep up the pace of digital transformation. According to KPMG, the sum of all the round values in which corporate venture capital (CVC) investors participated reached a total of US$315 billion in 2021, accounting for 29.4% of all VC deals globally, and surpassing the same period result of 2020 by 84% (Figure 4).

Corporate venture capital relies on the proceeding of directly investing corporate funds into privately-held new ventures (GOMPERS e LERNER, 2000). In this relationship, the corporate investor takes advantage of a financial return, access to the startup's innovation, people and any additional resources derived from that, while the new venture benefits from a financial injection and corporate investor's expertise, reputation and facilities (GOMPERS e LERNER, 2004).

Figure 4 – Global corporate venture financing over time (2014-2021)



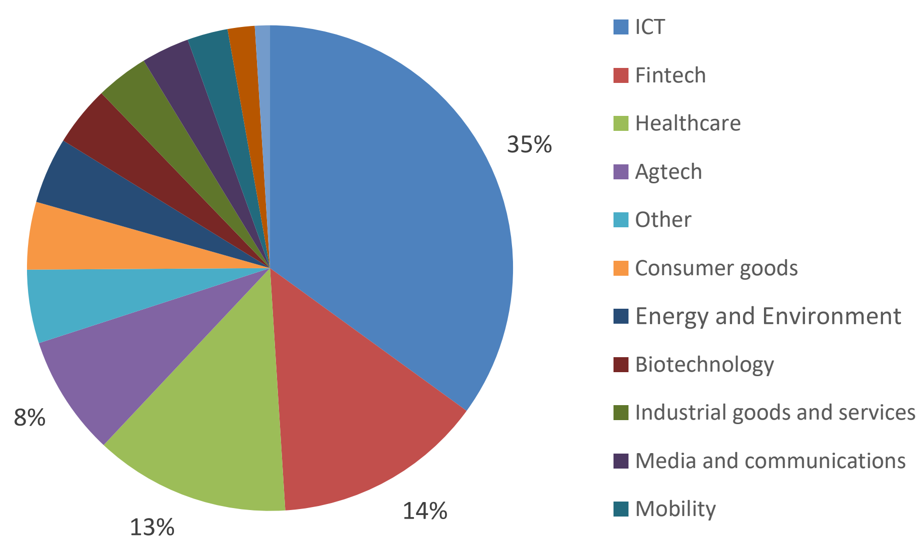
Source: Venture Pulse, Q4'21. Global Analysis of Venture Funding, KPMG.

According to KPMG (2021), an intense investment environment alongside the continued drive for digitalization, especially with the rise of the Omicron variant and the consequent postponement of returning to offices in some countries, in addition to the ongoing evolution of the VC industry in less developed jurisdictions, like South America and Africa, will presumably cooperate with keeping VC industry high heading during 2022. Another factor that may contribute for this to happen is the diversifying and still-expanding base of startups worldwide exerting significant demand for capital.

As a result of the collaboration between the Associazione Italiana del Private Equity, Venture Capital e Private Debt (AIFI) and LIUC Business School, a study of the Italian venture capital market is carried out regularly by the Venture Capital Monitor (VeM), mainly based on public sources. According to a report released in 2021 taking into account both the investment volume and the number of VC deals seen in Italy, the total volume of investment received by Italian startups throughout 2021 reached €992 million across 291 deals, surpassing the same period result of 2020 by 83%, with 68 more deals. At the level of geographical areas, investments in Northern Italy amounted to 64% of the total.

From a sectoral point of view, Information and Communication Technology (ICT) has largely attracted the interest of VC investors (35%), below what was achieved in previous years. Fintech (14%) sector appears in second place, followed by Healthcare (13%) and Agtech (8%), which is depicted in Figure 5.

Figure 5 – Distribution of VC-backed Italian startups by market segment (2021)



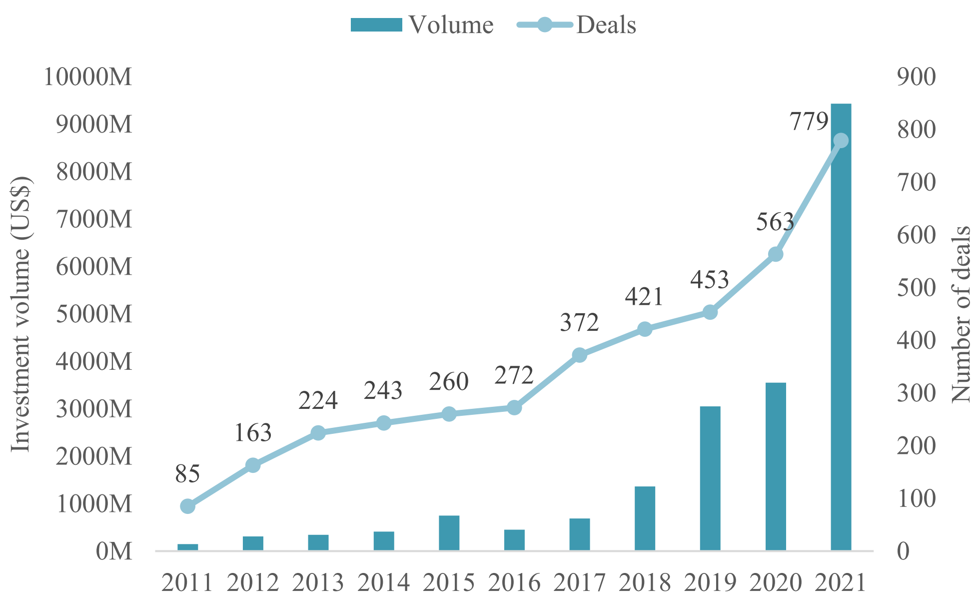
Source: adapted from *VeM Venture Capital Monitor Rapporto Italia 2021*, AIFI and LIUC.

The monitoring report of Italian startups updated by Camera di Commercio D'Italia on February 21, 2022 accounted for a total of 14201 startups in the country, 27% of which are located in the Lombardy region. However, Italy has seen just 2 startups achieve unicorn status (Yoox and MutuiOnline).

Distrito, an open innovation platform aimed at accelerating innovation in corporations, leveraging startups and its access to investments, released a report at the end of the first half of 2021 regarding the Brazilian venture capital industry. The analyzed startups should follow a series of restrictions in order to be considered in the study. They must have innovation as part of the core business, active status at the time of the study, Brazilian nationality and current operation in the country.

According to this study, the total volume of investment received by Brazilian startups during 2021 reached US$ 9.4 billion – only 1.4% of the global volume invested –, in 779 deals – only 2% of the total number of deals in the world –, surpassing the same period result of 2020 by 166% (Figure 6), with 216 more deals.

Figure 6 – Total investment received by Brazilian startups over time (2011-2021)

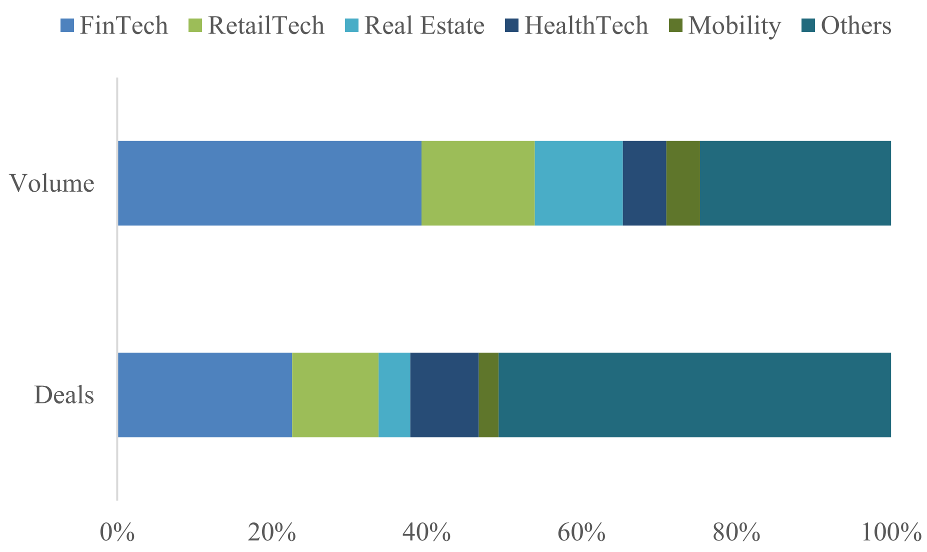


Source: adapted from *Report Retrospectiva 2021*, Distrito.

Nevertheless, it is highlighted by the report the potential distortion in data caused by 31 mega rounds that took place throughout the year with amount invested exceeding US$100 million, representing 74% of the total amount invested in the country. In addition to that, the total volume is highly concentrated in the South and Southeast regions of the country, totaling about 92% of the investment rounds that came off in the period, 64% only in São Paulo.

Considering the different sectors in which startups are inserted in, fintechs lead the list of the most invested ones (39%) in 2021, followed by retailtechs (15%) and Real Estate (11%), as depicted in Figure 7. The first two are, historically, the most prosperous sectors of the Brazilian startup ecosystem, with many representatives between unicorns and scale-ups. Meanwhile, the Real Estate sector appears well positioned due to its outliers that had mega rounds during this period, but there were just 32 deals in total.

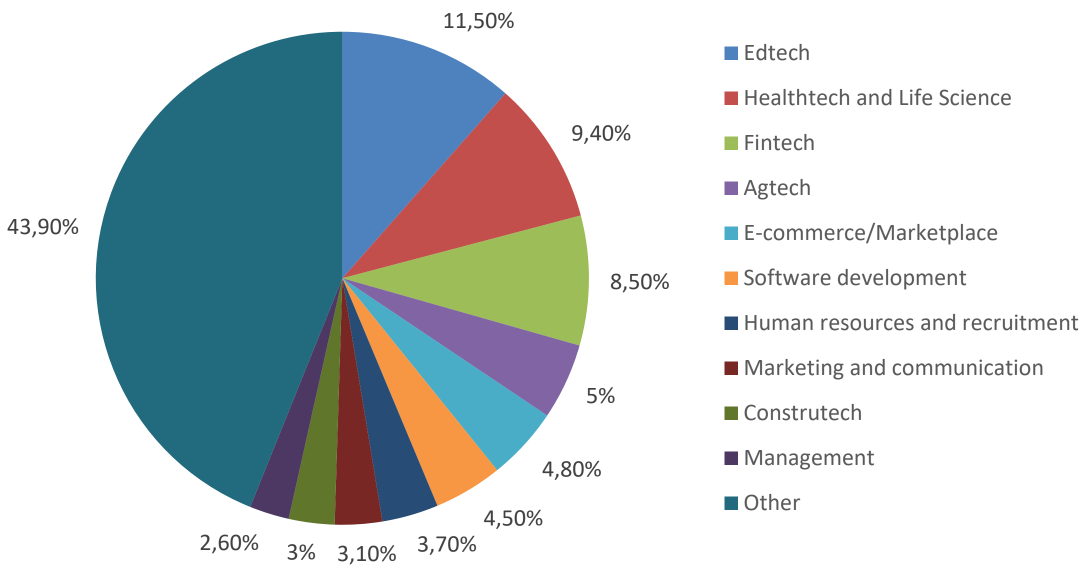
Figure 7 – Total investment received by each sector in Brazil (2021)



Source: adapted from *Report Retrospectiva 2021*, Distrito.

Taking into account the total number of nearly 13700 active Brazilian startups at the end of 2021, which represents 20 times more than a decade ago, a study conducted by Abstartups and Deloitte (2021), based on the mapping of startups in 314 Brazilian cities, identified São Paulo as the city with the highest number of startups in the country (32.5%). With most of them following the SaaS business model (40.8%), more than 11% of Brazilian startups are focused on education, representing the most significant market segment nowadays (Figure 8). During the last year, 11 new startups were awarded the title of Brazilian unicorns, making the country end the year with a total of 25 startups valued at US$1 billion or more.

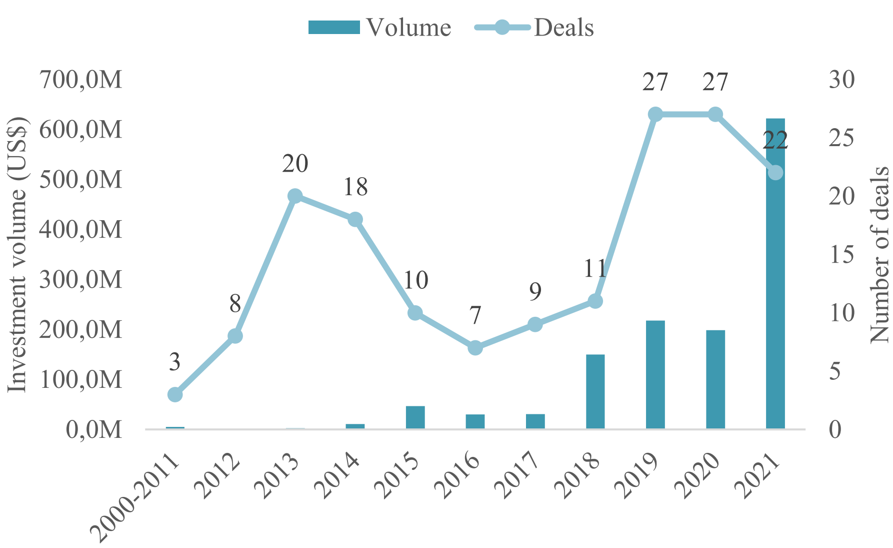
Figure 8 – Distribution of Brazilian startups by market segment (2021)



Source: adapted from *Mapeamento do ecossistema brasileiro de startups 2021*, Abstartups.

In the Brazilian scenario, there were 212 investment rounds involving CVCs in the country, a number reduced to 162 if only considering rounds with disclosed amounts, totaling US$1.3 billion invested over the last 20 years of this modality (Figure 9).

Figure 9 – Total investment with the participation of CVCs over time (2011-2021)

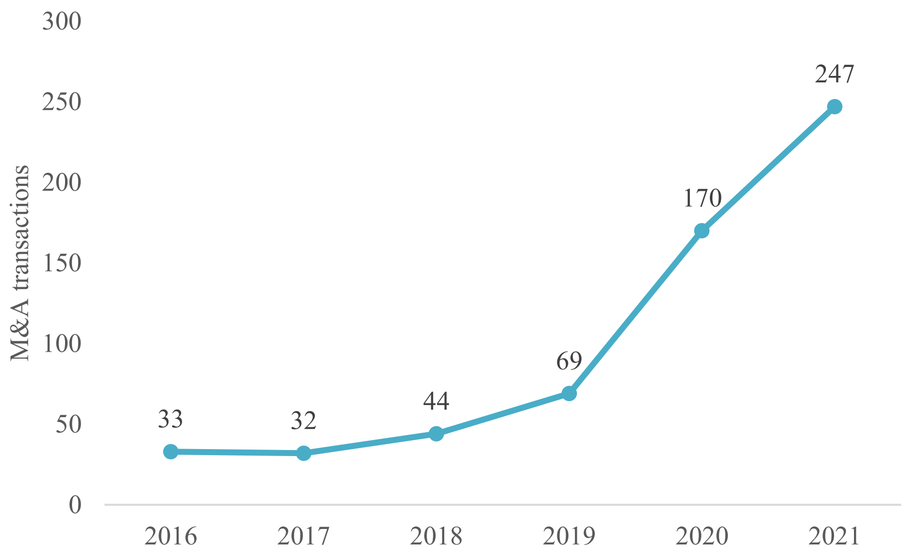


Source: adapted from Corporate Venture Capital Report 2021, Distrito.

The investment volume and number of deals became more relevant from 2013 onwards, with around 70% of these investments concentrated in the early stages, demonstrating its priority in startups at the beginning of their development trajectory.

Including both transactions in which the startup was acquired by a traditional company or by another startup, a total of 247 M&A transactions were identified in Brazil in 2021, the highest recorded number in history for the country (Figure 10). Among the benefits of this strategy, three of them are highlighted by Distrito's report: accelerating digital transformation, adding technology teams and creating new product lines.

Figure 10 – Total number of M&A transactions in Brazil over time (2016-2021)



Source: adapted from Distrito (2021).

1.2 MOTIVATION

The choice of the theme of the present work was primarily due to the author's great interest in entrepreneurship and innovation initiatives that have a socioeconomic impact on society. Moreover, the author did an internship at a multinational technological corporation that presents several initiatives fostering open innovation:

- A program designed especially for startups, offering mentorship, technical guidance, special discounts and learning content;

- An innovation hub connecting developers and startups, with the objective of sharing knowledge and promoting networking, either through virtual events or meetings in a physical space;
- A corporate venture capital fund focused on investments in women-led technology and innovation startups.

From direct contact with innovation projects led by several startups and being aware of their difficulty in accessing capital, the author not only approached the aforementioned themes, but also realized the importance of the subject for the improvement of the economy, especially in times that require a quick response to uncertainties.

The changes brought about by technological advances happen at an increasingly accelerated pace and, therefore, companies must be able to keep up with the technological gap that tends to increase over time. Nevertheless, this is not an effortless process. A survey carried out by Harvard University revealed that in 2019, even before the global pandemic, digital transformation was the main factor of concern for CEOs and executives of large companies but, despite this, 70% of efforts in digital transformation did not reach the desired results.

All in all, digital transformation consists in a complete paradigm shift, a new business, product, customer and value perspective. Meanwhile, startups are new ventures that, by definition, operate with a high level of uncertainty in search of a scalable model, which reinforces the necessity to be highly agile, adaptable and responsive to changes. That is why the startup culture has gained more and more importance in the market.

In addition to that, as stated previously, the importance of venture capital has increased over time. Being a type of investment focused on companies with high growth and profitability potential, venture capitalists invest their money in startups and scale-ups, believing in their ability to grow and generate expressive returns, typically targeting innovative companies. This work is aimed at contributing to the knowledge on such themes.

1.3 PROBLEM DEFINITION

In the first quarter of 2020, startups' prospects looked bleak due to restrictions related to the start of the global pandemic status. Global VC spending decreased 17% when compared to the preceding quarter, according to Crunchbase (2020). Then Covid-19 changed everything in terms of how people once lived and how they will live, conferring entrepreneurs a unique position to lead the way back to economic vitality for several economies and considerably boosting technology-related businesses (STARTUP GENOME, 2021).

According to Endeavor (2021), due to a new generation of startups that have attracted the attention of the largest funds in the world, VC investments are no longer concentrated in the United States and are becoming global. VC to non-US companies gre from 20% to 62% in the last 20 years. Possible reasons for that might be the qualification of their entrepreneurs and the growth potential of their companies in underserved and underexplored markets. This trend has been called “capital-geography decoupling”.

Generally, VC is viewed as a key determinant of startup growth and success, which may explain the increasing number of policy measures at national or subnational level intended to secure a greater supply of this financing stream. However, given the uncertain environment startups are inserted into and the particularly high risks faced by venture capitalists, alongside the considerable attention that startups have been drawing from private investors and policy makers, an in-depth study of which criteria lead to its success has become a meaningful issue to be explored.

1.4 OBJECTIVES

The present thesis has the objective of answering some research questions:

Q1: Are venture capital supply data statistically significant to model startup success?

Q2: Reaching a certain milestone in venture capital funding can be statistically significant to model startup success?

Q3: Countries with higher business dynamism present a more competitive landscape for venture capital-backed startups?

Q4: Are there significant differences in venture capital supply among countries?

Although several articles studied the relationship between VC investment and startup success, few are those that broaden the scope to encompass comparisons and discussions based on data from different countries.

More specifically, it will identify which explanatory variables are statistically significant for startup success, and then explore the role of business dynamism in this context.

1.5 TEXT STRUCTURE

For this purpose, the present document is structured as follows. Section 1 covers the work introduction, providing the contextualization of venture capital industry, author motivation with the issue, the problem description, objectives to be answered throughout this work and how it will be guided until the conclusion.

In Chapter 2, the literature review will be explored in order to provide all necessary theory to posterior comprehension and work development. Some of the subjects covered are the relationship between startups and venture capital funds, addressing all financing and statistical understanding requirements to follow the rest of the document.

Chapter 3 will present the method to be followed to answer the research questions raised through a seven-step approach from business understanding to model evaluation.

Then, chapter 4 provides all results and statistical analysis required to cover the discussion, passing through business understanding, data collection, data pre-processing, variable selection process, exploratory data analysis, model fit and model evaluation.

Finally, chapter 5 will conclude the document highlighting the benefits brought by the study and next steps that could be covered to improve the venture capital industry understanding.

2 LITERATURE REVIEW

2.1 INNOVATION

Novelty itself does not constitute innovation. It is actually defined with respect to how the introduction of a new feature modifies previously established practices. Indeed, “an innovation can be considered as a routine that purports to be new and potentially superior with regard to the accepted way of dealing with a given problem” (VANBERG, 1992).

The innovation process is driven by risks and uncertainties built into itself, making it as selective and competitive as to create incentives for the most promising solution to arise. As a result, it promotes a knowledge spillover effect and increases its potential returns (CANTAMESSA e MONTAGNA, 2016). Recognition, development and assessment of an opportunity, in addition to the risks and uncertainties that surround it, are essential steps in achieving a successful solution (SANZ-VELASCO, 2006) and supporting the number of new ventures launched (LANZA e PASSARELLI, 2014).

Uncertainty is interpreted as the individual's inability to precisely predict something due to lack of understanding about a certain condition (MILLIKEN, 1987). Nevertheless, the improved evaluation of the environment and perception of potential changes on that have a vast impact into the acknowledgement of feasible opportunities, which might be reached by means of developing resources and capabilities to manage and reduce uncertainties (ZACH, 2001). Additionally, the impact of those in appraising uncertainties may differ across geographical locations (HOPP e STEPHAN, 2012).

The firm’s innovation capacity refers to a continuous improvement of its resources and capabilities in order to generate innovation to meet market needs and deliver superior value (SZETO, 2000), which is commonly related to the achievement of a competitive advantage and the long-term success in businesses (HAN, KIM e SRIVASTAVA, 1998). Whereas resources are defined as competitive assets that are owned or controlled by a company, capabilities refer to its capacity to competently perform internal activity through the deployment of these resources (AMIT e SCHOEMAKER, 1993).

Innovation supports the company in coping with environmental turbulence, i.e. unexpected and unpredictable changes, especially in dynamic markets (BAKER e SINKULA, 2002). Its capacity to innovate also influences the region’s economic performance by virtue of the encouragement of connectivity, creativity and confrontation among distinct perspectives (LOPES, OLIVEIRA, *et al.*, 2021).

A special sort of entrepreneurial venture, i.e. startup, designed around rapid growth, is addressed in this work. Due to its contribution to the creation of new jobs, the variety of related businesses, the intensification of the knowledge spillover effect and incentive to competition, startups are considered a meaningful driver for economic growth and have innovation at their core (STOICA, ROMAN e RUSU, 2020).

The most common way to measure business dynamism among researchers is by gross entry and exit rates of companies. A stronger dynamism is related to higher rates of productivity growth as it influences the reallocation of resources from low-productivity to high-productivity activities, allowing successful firms to grow and the less productive ones to shrink (DAVIS e HALTIWANGER, 2014).

Therefore, the firm's innovation capacity aligned with the business dynamics are fundamental for companies to be able to face the turbulent and rapidly changing environment in which they are inserted and to manage the required entrepreneurship for the sustainable development of regions.

2.2 STARTUPS

According to StartupBlink organization (2021), startup can be defined as any business that applies an innovative solution that validates a scalable economical model, being innovation either a product, service, process or a business model. By this definition, it is noted that innovation can be given in different ways, but that is central for the success of a startup (RIES, 2019).

Startups must be designed to face situations of extreme uncertainty and, therefore, most traditional business management tools and concepts may not be applied in their context. The future of startups is unpredictable and subject to changes at an incredibly fast pace (RIES, 2019).

2.2.1 **Death Valley Curve**

Despite a clearly growing market, creating and scaling startups is not an easy task and several statistics raised over the years prove the high mortality rate of them, and also for the startups invested by VC.

A study carried out by Insper and Spectra Investments (2014) concluded that the mortality rate of portfolio companies of Brazilian funds in the last 30 years (1982 – 2014) was around

30&, that is, for every 100 investments completed, 39 were total losses. According to the authors themselves, this number is in line with what is seen in the rest of the world.

Several studies try to map the reasons for this high mortality rate. One of the most recent was carried out by Wildur Labs (2021), composed pf the testimony of more than 360 startups that failed for some reason, between 2000 and 2020. Some of the reasons mentioned are lack of clear business model, loss of focus, team misalignment, competition, failure to pivot, poor management of financial resources, among others.

Most business startups fail within their first ten years, and most surviving young businesses do not grow but remain small, not achieving enough scalability. However, a small fraction of young firms exhibit very high growth and contribute substantially to job creation and economic growth.

2.2.2 **Determinants of startup success and growth**

For years, several authors have been trying to understand what the success factors of a startup are. However, as entrepreneurship is a complex task susceptible to several variables, it is difficult to reach a conclusion and attribute the success of startups to a few factors.

What constitutes growth for new digital ventures is significantly different from established firms. In fact, young companies have limited financial performance and it is not possible to be assessed through conventional metrics such as customers, revenues, profits and turnover. Therefore, several studies have considered other outcomes like survival rate, funding, resources and capabilities as measures of success (AUDRETSCH, 2012).

Other studies found a list of factors that appears to have a correlation with the success of a startup. One of them have been conducted with 24 potential factors and, at the end, 8 of them were considered to have a positive correlation with startups' success: supply chain integration, market scope, firm age, size of founding team, financial resources, previous marketing experience, previous industry experience and patent protection (SONG, PODOYNITSYNA, *et al.*, 2008).

Additionally, startup success have already been calculated with the contribution of metrics like early stage success (ratio of series B to series A companies), late stage success (ratio of series C to series A companies) and speed to exit calculated through IPO (STARTUP GENOME, 2021).

Another paper refers to the process of merging or acquiring all or parts of other companies' (M&A). A process that look for complementarity between companies in different aspects, for instance, in resource, channel, brand and technology. Companies involved may benefit from synergies, which might result in competitive advantage.

2.2.3 **Funding access**

When a new venture is starting its operations, there are a few options on how to finance itself: bank loans, donations or awards, advances from customers, supplier credits, government financing, retained earnings, or selling company's stakes. Generally, the two sources that may provide large sums of capital to them are either bank loans or the sale of company stakes (PAVANI, 2004).

Loans are well-known mechanisms in which a person or a company receives money from a financial institution and pays that amount in the future with an additional interest. On the other hand, the sale of company shares is another approach most used by startups and it is the investment model carried out by VCs (PAVANI, 2004).

2.3 FINANCING

The innovation process is highly risky, especially in its earlier phases and, as startups are directly related to the innovation process, they are subject to it. Due to the risk involved, innovation is generally funded through equity coming from investors (e.g., VC funds), that, due to that, target a high return on investment (CANTAMESSA e MONTAGNA, 2016).

2.3.1 **Types**

Angels use their own capital to invest and they tend to focus on younger companies than do VCs and make a larger number of smaller investments. Besides the money, they commonly contribute with the startup by providing their professional experience and networks.

Venture capital funds are leaded by professional investors in considerably larger rounds of investment, with the goal of exit after some years, seeking for financial return (CANTAMESSA e MONTAGNA, 2016).

2.3.2 **Investment rounds**

Each financing event is known as a round. There is no rule that a company need to pass for all the stages of financing, meaning that a company might receive several rounds of investment at a specific stage, or it might receive sufficient investment in one round and ignore what were supposed to be the next stages.

It is common for the startup to carry out more than one round of investment during its process of growth, being your first round called Seed, the second Series A, the third Series B, and so on (FELD e MENDELSON, 2011). Sometimes startups have an additional round called Pre-Seed for after a few months raise a more robust round, the seed round. There may be a situation in which the startup was not able to capture a new round of investment with another VC, and then its current investors end up having to make new investments, resulting in an extension of the last round.

2.4 VENTURE CAPITAL

A venture capital company can be summarized by five main components (METRICK e YASUDA, 2011):

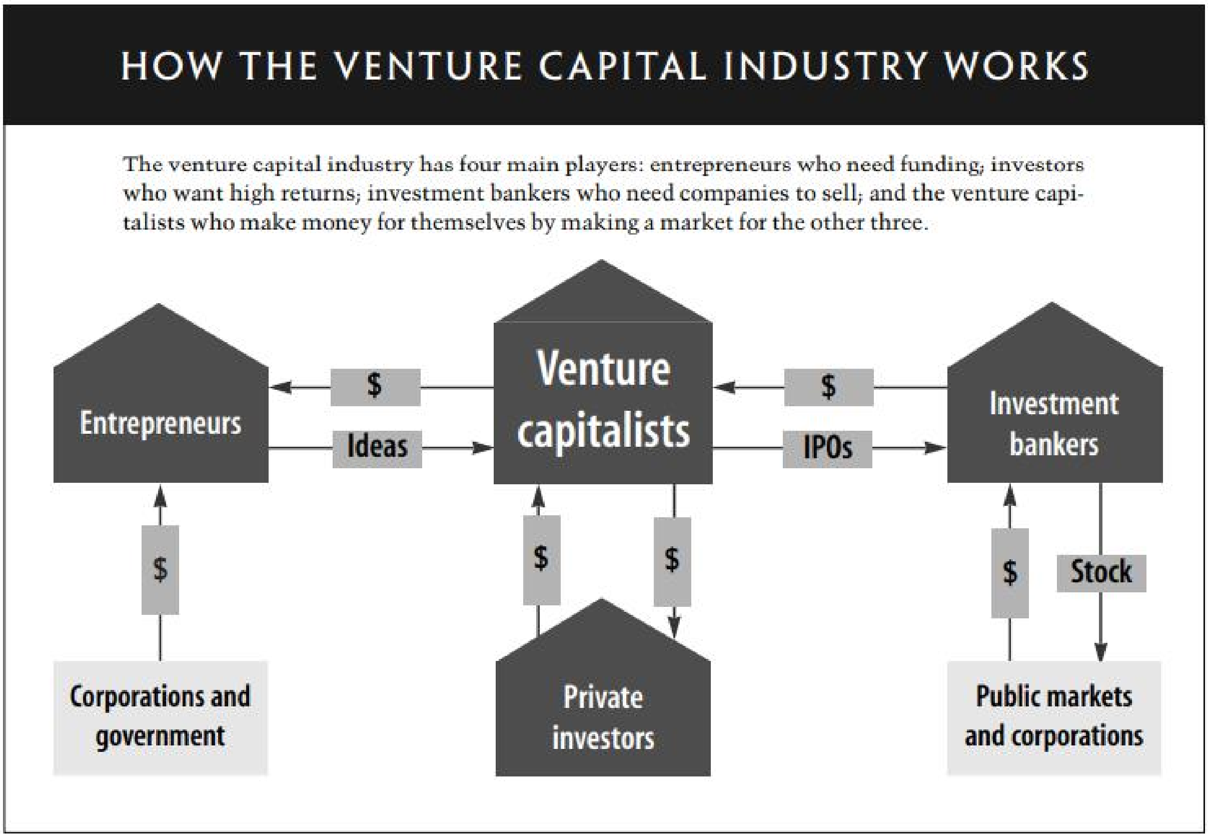
1) Financial intermediary, which means that the VC act as a third-party in a financial transaction, taking the investors' capital to make equity investments in the companies;
2) Investment provided just for privately held companies, being defined as a type of private equity investment;
3) Performs an active role in managing and helping companies in its portfolio;
4) Objective of exiting investments through a sale or an initial public offering (IPO) as a mechanism to return money to their investors;
5) Fund the internal growth of companies, which explains the interest in new businesses.

Figure 11 better explains the VC role as a financial intermediary. Concisely, entrepreneurs receive money from VCs in exchange for a stake in their startups, private investors pump money into VCs expecting high financial returns, and investment bankers sell the startups, returning the VCs' investments.

The active role of VC in the portfolio companies may be, for instance) a position on the board of directors, providing regular advice or being unofficial recruiters as young companies face difficulties in attracting high-quality talent. VC firms can also provide value-added services to its portfolio companies like supervision, governance and network connection (CAMPBELL e FRYE).

Most VC firms specialize their funds by stage, industry, or geography. Some of them may even change focus over time or work with more than one strategy.

Figure 11 – Venture capital as a financial intermediary in the industry



Source: Zider, 1998

2.4.1 Venture capital stages

Venture capital funds may differ from each other by the startup stage they are specialized, ranging from early to late stages.

Early stage companies are directly associated with the initial commercialization of a product, testing its Minimum Viable Product (MVP) or doing a pilot, being the company in the process of organizing itself or in the market for three years or less. Even though VCs can participate in earlier stages of financing (e.g., seed), they are usually more involved with early stage rather than a pre-marketing stage aimed at proving a concept.

Meanwhile, late stage companies are those that already have a proven product and profits or, at least, a direct path towards profitability, with scalable business. This group might include companies with positive cash flow and that are considering IPOs.

2.4.2 Investment decisions

Investment decisions already begin when VCs are prospecting for new opportunities and deciding on companies to invest in, follows a term sheet valuation that has to be presented to the company, then the VC performs due diligence, until all parties reach a final closing.

2.4.3 **Exit strategies**

The VC holds a requirement to exit, focusing on financial return. Since there are plenty of small businesses and in most of them would be difficult to find a real expectation of exit within a set time period, VCs generally just invest on them if they see a realistic chance of this firm growing enough within five to seven years after the initial investment.

Exits can occur in three different ways. It may happen either through an IPO followed by the sale of the VC stake in the open market, through a sale of the portfolio company to another investor, or through the sale of the company to another company (M&A).

2.5 BUSINESS DYNAMISM

Business dynamism can be defined as “the process of firm entry, growth, and exit, and the simultaneous creation and destruction of jobs” (OECD, 2020). Since there is not an official variable associated to it, several authors supposed different ones to measure it: entry rates, job reallocation rates (OECD, 2020), number of new jobs (DECKER, HALTIWANGER, *et al.*, 2014)

It plays a role in reallocating resources from low to higher productivity units. Thus, acquiring an improved perception towards this trend can increase the effectiveness of public policies applied on innovation (DECKER, HALTIWANGER, et al., 2014)

A stronger dynamism is related to higher rates of productivity growth as it influences the reallocation of resources from low-productivity to high-productivity activities, allowing successful firms to grow and the less productive ones to shrink, relying on the continuous process of firm entry and exit.

One way of measuring business dynamism is through the Global Competitiveness Index that, overall, measure national competitiveness. The 11th pillar is the business dynamism one and it assesses the private sector’s capability to engender and adopt new technologies and new forms of working.

2.5.1 **Drivers**

A strong association between business dynamism, market structure and firm heterogeneity was found (OECD, 2020). High industry concentration, i.e., an industry with a small number of large companies, is highly associated with low business dynamism, with high entry barriers to the market (GINEVICIUS e CIRBA, 2007).

Moreover, countries with higher intangibles assets and digital technology intensity, e.g. R&D, patents and software) may indeed reduce incentive for new young firms to enter and innovate as it requires large initial investment (OECD, 2020).

2.6 MULTIVARIATE ANALYSIS

A multivariate data set is composed by data in which values of several measurements (variables) are recorded on each individual or object (unit) in one or more samples. Thus, Multivariate Analysis (MVA) is the analysis of multivariate data sets supported by the application of a collection of methods (RENCHER, 2002). Some of the objectives involving multivariate analysis are described as follows (JOHNSON e WICHERN, 2007).

- Data reduction or structural simplification: phenomenon representation as simply as possible without losing valuable information;
- Sorting and grouping: creation of similar objects depending on certain aspects;
- Investigation of the dependence among variables: determination of the existence and the nature of the relationship among variables of interest;
- Prediction: usage of the relationship among observed variables in order to predict the values of one or more variables of interest;
- Hypothesis construction and testing: assumptions validation or reinforcement of prior conviction through the test of statistical hypotheses formulated in terms of determined parameters.

2.6.1 **Principal Component Analysis**

The Principal Component Analysis (PCA) is a dimension reduction technique for data (LUONG, DELIGIANNIS, *et al.*, 2018) and its main idea is to reduce a complex data set in which there are a large number of interrelated variables to a lower dimension, while retaining as much information as possible of the data set. This reduction is achieved by mapping a new set of orthogonal dimensional features, the principal components (ZHAO, ZHENG, et al., 2019), which are aggregates of the correlated variables.

Then, these principal components are ordered so that the first few retain most of the variance in the observed variables (ANDERSON, 2003). Accordingly, the first principal component accounts for the largest possible variance, while the second component, under the condition of no correlation to the first one, will account for the second largest variance, and so on.

In order to decide on how many principal components should be retained to summarize the data, four criteria can be used (RENCHER, 2002). The first criterion is to keep sufficient components to explain a particular percentage of the total variance, since the fraction of variance explained by a principal component is the ratio between the variance of that principal component and the total one.

The second criterion is based on retaining those components that explain more variance than the average variance of the variables, i.e. those with eigenvalue greater than 1 in the case of a correlation matrix. The eigenvector represents the direction of the maximum variance in the data set, i.e. the principal component, while eigenvalues are coefficients applied to eigenvectors that give the vectors their magnitude, indicating how much of the data's variability is explained by its eigenvector.

The third criterion uses the scree graph, which is the plotting of the component's eigenvalue versus its position in the descending list of the total variance explained. The goal is to look for a natural break between a steep curve, characterized by larger eigenvalues, and a following straight line. The eigenvalues that are usually retained are those before the straight line (RENCHER, 2002).

Finally, the fourth criterion is based on testing the hypothesis that the last k eigenvalues are small and identical. It is helpful to test the significance of the components with larger eigenvalues as suggesting those are the ones that capture all the fundamental dimensions, whereas the last components reflect noise.

A PCA involves conducting the following steps: (1) selecting and measuring a set of variables, (2) understanding the correlation matrix, (3) extracting and determining a number of factors from that, (4) rotating the factors to improve interpretability and (5) describing the obtained results. Overall, a factor is more easily explained when it presents high correlation with some observed variables which do not correlate with other factors, so that each variable loads on as few factors as possible (CATTELL, 1973).

2.7 HYPOTHESES TESTING

2.7.1 **Kruskal-Wallis H test**

Kruskal-Wallis test is a rank-based non-parametric H method, for three or more groups, to test whether samples are from the same distribution. It is an alternative to the ANOVA test for parametric measures and an extension of the Mann-Whitney U test, allowing the comparison of more than two independent groups. The null hypothesis is that the distribution functions of the samples are equal (KRUSKAL e WALLIS, 1952).

In order to apply the Kruskal-Wallis H test, four assumptions have to be met:

- Independent variable with two or more categorical independent groups;
- Dependent variable measurable at the ordinal or continuous level;
- Independence of observations (i.e., no relationship between the members in each group or between groups);
- All groups with the same shape distributions.

However, this test just provides information that at least two groups are statistically significantly different from each other, but does not point out which specific groups are they. Then, it is possible to use a post hoc test in order to discover it.

If the null hypothesis is rejected by Kruskal-Wallis H test, it is desirable to proceed with paired multiple comparisons in order to assess which pairs of independent groups present different distribution functions.

2.8 OVERSAMPLING

The challenge of working with imbalanced datasets is that most of the models might ignore the minority class, resulting in a model with poor performance. One way of addressing this issue is oversampling the minority class. In most cases, oversampling is preferred over undersampling because the second may lose information on data when removing instances.

2.8.1 **Synthetic Minority Over-sampling Technique (SMOTE)**

Synthetic Minority Over-sampling Technique is a statistical technique aimed at increasing the number of cases in a dataset in a balanced way by generating new instances from existing observations of the minority class.

The algorithm try to overcome the potential overfitting problem by random oversampling. It works by combining features of the target case with features of its k nearest neighbors, that are chosen to interpolate new synthetic instances (CHAWLA, BOWYER, *et al.*, 2002).

2.9 REGRESSION MODELS

Regression methods are usually applied to describe the relationship between a response (or dependent) variable and one or more explanatory (or independent) variables. Accordingly, regression analysis traces the conditional distribution of a response variable as a function of one or several explanatory ones. This relationship is of great interest when considering the possibility of the former being affected by the latter, or when the independent variables may be used to predict the dependent one (FOX, 2016).

The most common example of modeling is the linear regression, in which one the outcome variable is assumed to be continuous. Meanwhile, the outcome variable in logistic regression is binary or dichotomous one, differentiating itself from a linear regression by the form the model is provisioned and its assumptions (HOSMER e LEMESHOW, 2000).

Models should not perfectly either predict or explain the mean response, which means there is error involved in the process. However, the regression modeling might be able to understand and control the uncertainty involved in estimating the parameters of the distribution. In the case of linear regression, the error is normally distributed, whereas the binomial distribution describes the distribution of errors when it comes to the logistic regression (HILBE, 2009).

2.9.1 **Logistic regression**

Logistic regression is the foremost method used to model binary responses following a Bernoulli probability distribution, which consists of a distribution of 1s and 0s, with 1 indicating a success or the occurrence of an event. Its primary feature is that the fitted value is the probability of a particular outcome to occur is equal to p ranging from 0 to 1 (HILBE, 2009). Its function is based on S-curve graph involving the probability p and the natural logarithm e, following the general form expressed in (1), where $\beta_0, \dots, \beta_p$ are the unknown parameters that will be estimated and $x_0, \dots, x_p$ are the independent variables (CIABURRO, 2018).

$$p(\mathrm{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}} \tag{1}$$

Logistic regression is about modeling the odds rather than a probability. The odds describe the expected number of successes (positive events) per failures (negative events). Additionally, a link function defines how the linear combination of the predictors is related to the mean of the response, which is reached by applying the logit transform to $p(\mathrm{x})$, resulting in the called log odds, i.e. the log of the odds (DALPIAZ, 2016).

$$\log \frac{p(\mathrm{x})}{1 - p(\mathrm{x})} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \tag{2}$$

It is possible to find the magnitude of the change in the odds for a unit increase in the regressor by exponentiating the associated coefficient, resulting in the odds ratio. Thus, if odds ratio is greater than one it means that the event presents higher odds of occurring in the first group.

In order to run the logistic regression model, some assumptions need to be met:

- There should be no multicollinearity between predictor variables, which means that there should not be a high correlation between independent variables inserted in the model;
- The continuous independent variables should be linearly related to the log odds

Logistic regression models are estimated with the use of an estimation technique called maximum likelihood. Defining the probability P(y) associated with the total amount of ones and zeros cases encountered in a sample:

$$P(y) = \prod P^y \, (1 - P)^{1-y} \tag{3}$$

This function is called the likelihood function for the logistic regression model. Being P in function of the parameters and their effects βs, the idea of the maximum likelihood estimation is to take those βs that maximize the likelihood function, which is achieve through an iterative process (DEMARIS e SELMAN, 2013).

There are a variety of ways of assessing the model's fit, for example using Pearson Chi-square statistic, Hosmer-Lemeshow test and confusion matrix (HOSMER e LEMESHOW, 2000). However, this work will focus on confusion matrix and in the receiver operator characteristic curve (ROC).

The ROC curve displays the optimal relationship of the model sensitivity by one minus the specificity, and it can be used to determine the predictive power of the model. A ROC statistic of 0,5 represents a model with no predictive power, values ranging from 0,5 to 0,65 have little predictive power, from 0,65 to 0,80 have moderate predictive power, while values greater than 0,8 have strong predictive power. This ROC statistic is called Area Under the Curve (AUC).

Sensitivity can be defined as the probability of detecting true signal, measuring how many true positives were identified out of all the positives, while specificity as the probability of detecting false signal. The better threshold is the one that maximizes both sensitivity and specificity (HILBE, 2015).

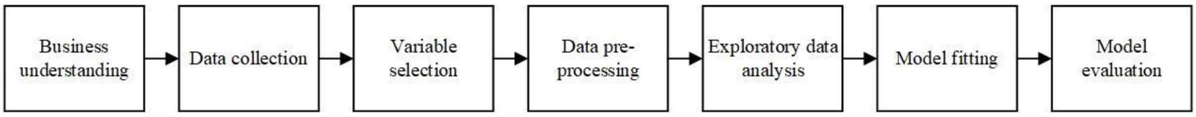
The confusion matrix can be understood as a table of correctly and incorrectly predicted fitted values, indicating how well the model predicts group membership. To build this, it is necessary to define a threshold, so that estimated probabilities higher than that will predict the output variable as 1, otherwise it is equal to 0.

From the confusion matrix, it is possible to derive three main metrics to assess the model: accuracy, sensitivity and specificity. The former is calculated through the ratio of correct predictions by the total predictions (HILBE, 2015).

3 METHOD

Based on the proposed objectives to identify the most suitable model to explain the startup's success in a global perspective, the problem will be analyzed following the method proposed in Figure 12, and the results obtained will be examined in Chapter 4. It comprises seven steps, starting from the business understanding to the model evaluation.

Figure 12 – Research methodology



Source: elaborated by the author

First and foremost, understanding the business environment and its context is determinant to establish the aim of the model. It provides clarity when selecting which technique to use, influences the choice of the dependent variable and what independent ones would be appropriate to consider.

In-depth understanding of the business is based on data obtained from semi-structured interviews. By combining the results acquired from interviews with secondary data and literature review, it is feasible to enrich the business understanding.

Data collection is the process of gathering observations to be explored. It provides the first understanding about the dataset that will be handled and how information is spread across different sources and structures.

As the variables to be incorporated in the model must be related to our problem, the next step is to select, adapt and compute those that are the most relevant to improve the performance of the model.

The fourth step consists of systematizing available data on one database, and pre-processing it, to be further statistically analyzed. The Exploratory Data Analysis (EDA) is crucial in order to execute preliminary investigations on data to provide pertinent statistical measurements, detect anomalies, verify assumptions and comprehend the relationship among its elements.

The fifth step is aimed at finding the best possible fit to data, which is performed using the likelihood ratio test, through the running of a logistic regression model. Finally, the last step is to evaluate the model concerning improvements in its accuracy, recall and precision metrics.

3.1 RESEARCH QUESTIONS

In order to deal with the problem exposed above, it will initially establish research questions to be subsequently studied. Therefore, the aim of this section is to define the questions to be answered by data analysis through application of the methodology explained.

Firstly, the logistic regression model, alongside distinct statistical techniques, would be useful to understand and fit the model that explains startup success, answering to Q1-Q2. Afterwards, a nonparametric method named Kruskal-Wallis test would be conducive to answering Q3-Q4.

Q1: Are venture capital supply data statistically significant to model startup success?

Q2: Reaching a certain milestone in venture capital funding can be statistically significant to model startup success?

Q3: Countries with higher business dynamism present a more competitive landscape for venture capital-backed startups?

Q4: Are there significant differences in venture capital supply among countries?

Answering these questions allows further understanding on the panorama of VC industry, its association with startup success and the role of country's business dynamism.

4 RESULTS

This chapter develops the proposed method, in order to answer research questions Q1-Q4. For this purpose, the seven-step approach will be followed and, after model evaluation, some hypotheses will be tested using the Kruskal-Wallis H test.

4.1 BUSINESS UNDERSTANDING

In addition to the material encountered in literature review that facilitates the understanding of the business environment, it is relevant to comprehend what is the meaning of success for the startup and the VC involved in the funding round. Furthermore, how the selection and evaluation of startup occurs in order to receive an investment, the different roles and which variables can be used to characterize the event.

Three semi-structured interviews (informants A, B and C) were conducted following the script depicted in Table 1. Interviewees' profiles are presented in Table 2. The objective was to develop critical sense in order to explore future analyses.

Table 1 – Interview structure

Industry overview
What is, under your perspective, the reason for the rise in VC market during last year?
Investment process
How works the selection process of startups to be invested in? What are the most important variables for the VC?
What are the main objectives of a VC investment? What signals that the deal was a success?
How is defined the amount invested in a startup? How do investment rounds work?
Dependent variable selection
Given the difficulty of accessing startup data as they are privately held, what is your perception in considering its success as having either an IPO or M&A transaction?
What are the benefits for the startup when reaching this stage?

Table 2 – Interviewee's profile

	A	B	C
Position	VC fund CEO	VC fund CEO	Startup employee
Years of experience	10	6	5
Interview time	30 minutes	1 hour	1 hour

Source: elaborated by the author

The results from the interviews are presented based on the three main divisions: industry overview, investment process and dependent variable selection.

4.1.1 **Industry overview**

According to B, the global rise in VC industry might be explained by the maturity of the entrepreneurial ecosystem outside the United States, more entrepreneurs, more structured actions, maturate technology application to solve problems in massive scale and macroeconomic factor such as inflation and interest rates that attract capital.

Respondent A believes that national policies and legislation may influence investors to look abroad, seeking to increase exposure to the international market and pursue attractive results within adequate risk-return ratios.

4.1.2 **Investment process**

According to B, the criteria to select a startup to be invested varies across VC companies, as there are several specialized investment funds, for instance in the investment stage (e.g., early or late stage) or in cultural aspects (e.g., diversity related). However, all respondents agree on some principal aspects to be analyzed:

- Financial health (more relevant for late stage startups already generating cash);
- Technological base and distribution channel;
- Team execution ability;
- Potential scalability;
- Cultural alignment.

Respondent C added some other aspects, such as the size of startup's target market, as it needs to be at least big enough for the startup to grow, alignment to market trends, the value proposition and its business model, highlighting features on how the startup works and how it plans to monetize its business.

For respondent A, the main objective of a VC investment depends on the fund's thesis, which is based on the VC company expertise. Its thesis may be aligned with more financial dynamics, a return profile and a risk-return distribution for portfolio organizations, socioeconomic impact criteria, among others.

Regarding the amount invested in a startup, respondent C affirmed that it might be related to the resources needed for startup to grow from one point to another. The startup defines how much cash to request for strategic planning and then has to demonstrate what are the growth assumptions, the milestones and how to make them tangible in the business plan.

According to B, the startup defines the value to be invested, but they have to demonstrate it is the necessary fund for its growth plan to materialize. Moreover, investment rounds vary according to the startup's maturity level and it is directly related to the risk involved.

4.1.3 **Dependent variable selection**

According to B, the main objective of an investor is to achieve success at divestment, which may be related to generating value and impact, the desirable startup valuation increase during investment process, the occurrence of its next investment round or reaching a M&A or IPO at the end.

Respondent A reinforced that both VC and startup are looking for liquidity. While the VC is pursuing a profitable way out, the startup may see it as an advantageous monetization.

Besides that, a M&A causes the startup to become part of a more robust, integrated ecosystem, resulting in greater value generation with the acquisitor. On the other hand, reaching an IPO means the startup has reached high levels of maturity, transparency and relevance, being able to finance itself by going public. Therefore, all respondents agree with the dependent variable selection.

4.2 DATA COLLECTION

Following the approach used by several papers (ARROYO, COREA, *et al.*, 2019) (ZBIKOWSKI e ANTOSIUK, 2020) (GASTAUD, CARNIEL e DALLE, 2019), the present work uses data from the Crunchbase platform, a provider of business information of public and private companies, made available either by the companies themselves or through news announcements. It includes the funding round history of startups, ranging from early-stage rounds to late-stage ones.

A sample of 81460 startups was gathered with founding dates ranging from 1992 to 2017. 1992 is the year in which the popularity of the internet started to grew, and new business models flourished during that period based on the internet (LITAN e RIVLIN, 2001). Following previous works (BENTO, 2017) (ZBIKOWSKI e ANTOSIUK, 2020), companies founded between 2017 and 2022 were excluded from the analysis because they are at its initial stage of operation and the mean age at which a startup raises it first round of VC investment is three years. The startups included in the sample received at least one round of VC financing between 1992 and 2021.

Moreover, the oldest samples, which are mainly from United States, may be biased towards companies that persisted through economic downturns, but it is supposed that this bias was reduced over time as Crunchbase became popular. The data used in the research was obtained on April 10^{th}, 2022.

In order to reach the desired extent of data to be explored, some filters were activated when looking for information on the platform. Moreover, the selection of the countries to be further explored was taken based on the global rankings of startup ecosystems comprising 1000 cities and 100 countries named *Global Startup Ecosystem Index* (STARTUPBLINK, 2021).

This ranking stipulates a total score for means of comparison across countries, taking into account subscores as quantity, quality and business environment. The former checks the activity level of an ecosystem through the number of startups, coworking spaces and accelerators, among others indicators. The quality score encompasses both the quality of startups and other supporting organizations. Finally, the third factor focuses on attributes at the country level, for instance national infrastructure, policies and legislation (STARTUPBLINK, 2021).

Therefore, for the purpose of the present work, the first thirteen countries with higher overall score and with sufficient information available on Crunchbase, plus Brazil and Italy that are object of the author’s personal interest, were selected to be explored. The filters applied on the database are listed in Table 3.

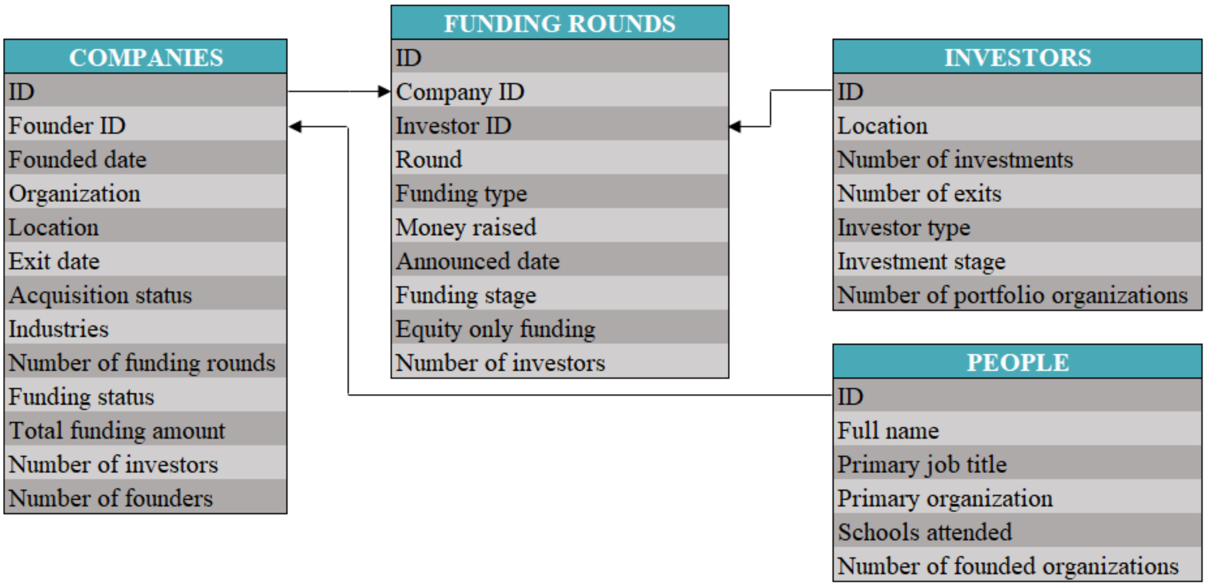
Table 3 – Database specificities

Number of entries	81460 - 29448 null raised amount entries = 52012 valid entries
Filters	Investor Type: Venture Capital Founded date: 1992-2017 Equity Only Funding Announced date: before 2022
Location	Australia, Brazil, Canada, China, France, Germany, India, Israel, Italy, The Netherlands, Singapore, Sweden, Switzerland, United Kingdom and United States

Source: elaborated by the author

Before selecting the variables to be applied for the logistic regression, the fundamental ones provided by Crunchbase were identified (Figure 13). The *funding rounds* table includes information on each transaction taking place in VC industry. The table is composed by the investment round type (e.g., pre-seed, seed, series A to J), the funding type (i.e., venture capital), the amount of money raised by the startup, the transaction's announced date, the funding stage it belongs to, if it is an equity only funding and the number of investors involved in the specific transaction.

Figure 13 – Crunchbase's data structure



Source: elaborated by the author

The *investors* table summarize data regarding the investor type (i.e., venture capital), in which country it is located, the number of investments they take part in, the number of exits (i.e., M&A or IPO) reached by the companies invested by them, the investment stage they usually participate and the number of portfolio organizations.

The *companies* table holds basic information such as when the organization was founded, its full name, what is the headquarters location, the exit date – in case it had one –, its acquisition status by another firm, to which industries the company belongs to, the number of raised funding rounds, its current funding status, the total funding amount until the moment data were collected, besides the total number of investors and founders involved with the company.

The *people* table describes individuals who are founders, including the person's name, primary job title and organization, its education level defined by the number of schools attended and the number of founded organizations. Although not used in the research, additional tables are available on the platform.

4.3 VARIABLE SELECTION

Defining the dependent variable requires a reasonable understanding on what success in a new business venture means. Undertaking an IPO is a commonly adopted measure, and several studies tend to portray a positive contribution of venture capital finance into the company's probability of survival until the IPO stage (SHANE e STUART, 2002) (BER e YAFEH, 2007) (WANG e WANG, 2011). The acquisition of the new business is also viewed as a signal of its success. High-potential startups often receive high premiums to be acquired by established companies, leading to a desirable result to the founders, representing a "win-win" outcome for both parties (COTEI e FARHAT, 2017) (GRAEBNER e EISENHARDT, 2004). Therefore, the target variable adopted in the logistic regression is 1 if the startup went to an IPO or was being acquired, and the null value encompassing all other companies.

The actual dataset used was created by combining information from the *funding rounds*, *investors*, *companies* and *people* tables from Crunchbase. Table 4 provides an overview of the variables and its respective definitions used to compose the logistic regression model.

Startup's headquarters location and *industry group* the startup is referenced to in the database were defined as categorical variable, considering that fast growing industries may attract more reputable investors (HOENIG e HENKEL, 2015).

In addition to the general objective of studying the venture capital industry in a variety of countries, the location variable is important as some of them might benefit from better access to VC funding, alongside the national infrastructure and legislation offered (RÖHM, KÖHN, *et al.*, 2017). Aligned with that, a categorical variable with positive value for those cases in which both venture capital investor and startup are located in the same country during early stage investment has been added (ARROYO, COREA, *et al.*, 2019).

Due to the fact that the first 3-year period of startup survival is the average at which the company usually receives its first round of investment, this point in time became of interest for several researchers. In the following model, the *last venture capital funding stage* (i.e., seed, early or late stage) and the *raised amount* within such period are considered (BROWN e ROCHA, 2020)(GASTAUD, CARNIEL e DALLE, 2019).

Table 4 – Selected variables

Concept	Variable	Type	Description	Reference
Success	IPO/M&A	Binary	1 if the company was either acquired by other or went public	(ZBIKOWSKI and ANTOSIUK, 2020)
Company	Startup location	Categorical	Country in which the startup is located	(ARROYO, COREA, *et al.* , 2019)
	Startup industry	Categorical	1 if startup is referenced to that industry	(HOENIG and HENKEL, 2014)
	Location similarity	Categorical	1 if both venture capital investor and startup funded are located in the same country (early stage)	Adapted from (ARROYO, COREA, *et al.* , 2019)
	VC stage	Categorical	Last venture capital funding stage received within the first three-year period since the firm was founded	(BROWN and ROCHA, 2020)
	Raised amount 3-year	Numerical	Total venture capital funding in US dollars within the first three-year period since the firm was founded	(GASTAUD, CARNIEL and DALLE, 2019)
	Total raised amount	Numerical	Total venture capital funding in US dollars	(ARROYO, COREA, *et al.* , 2019)
	Number of funding rounds	Numerical	Total number of funding rounds	(ARROYO, COREA, *et al.* , 2019)
Founder	Number of founders	Numerical	Total number of founders	(GASTAUD, CARNIEL and DALLE, 2019)
	Academic experience	Numerical	Number of superior degrees concluded by the founder	(ZBIKOWSKI and ANTOSIUK, 2020)
	Entrepreneurial experience	Categorical	1 if any of the founders had company founding experience prior to founding the focal venture	(KIM and PARK, 2017)
Funding round	Funding round type	Categorical	1 if the company completed a Pre-Seed, Seed, A, B, C, D, or E-J round	(GASTAUD, CARNIEL and DALLE, 2019)
	Raised amount in funding round	Numerical	Raised amount in Pre-Seed, Seed, A, B, C, D, or E-J round	Adapted from (GASTAUD, CARNIEL and DALLE, 2019)
	Startup age in funding round	Numerical	Company's age when announced the funding round	Adapted from (GASTAUD, CARNIEL and DALLE, 2019)
	First round age	Numerical	Startup's age at first funding round in years	(GASTAUD, CARNIEL and DALLE, 2019)
Investor	Number of investors	Numerical	Total number of investors for each round	(PAHNKE, KATILA and EISENHARDT, 2015; ARROYO, COREA, et al., 2019)
	VC reputation	Numerical	Total number of portfolio ventures, backed by the investor, acquired or taken public	(KIM and PARK, 2017; RÖHM, KÖHN, *et al.* , 2017)
	VC experience - industry	Categorical	1 if at least one investor had previous experience in the focal startup's industry	(BAUM and SILVERMAN, 2004)
	VC experience - investments	Numerical	Number of VC's performed investments	(KIM and PARK, 2017)
	VC size	Numerical	Amount of total funds under management of the VC firm	(CHANG, WONG and HO, 2016)
Environment	VC competition	Numerical	Market concentration adapted from Herfindahl-Hirshman Index (HHI)	(HONG, SERFES and THIELE, 2019)
	Startup competition	Numerical	Market concentration adapted from Herfindahl-Hirshman Index (HHI)	Adapted from (HONG, SERFES and THIELE, 2019)

Source: elaborated by the author

If not clearly stated in Crunchbase, the VC stage category was coded to those startups that received venture capital finance within the first three-year period since its foundation, according to the funding rounds participated. This may be helpful for the overall understanding of the problem because late stage investment is generally characterized by more stable future payoffs and a better access to information.

In order to complete information regarding the company itself, the variables *total raised amount* until the end of the studied period interval, *total number of founders* and *number of funding rounds* are computed (ARROYO, COREA, *et al.*, 2019) (GASTAUD, CARNIEL e DALLE, 2019).

Additionally, the variable for *founder's academic experience* was calculated by the maximum number of superior degrees concluded by any founder (ZBIKOWSKI e ANTOSIUK, 2020), whereas *founder's entrepreneurial experience* indicates if any of the founders had previous founding experience (KIM e PARK, 2017).

Regarding funding rounds, following (BENTO, 2017), binary variables were designated in case the startup completed a specific investment round, as well as the amount raised in each one of them and the company's age when passing through this milestone (GASTAUD, CARNIEL e DALLE, 2019).

All dates encountered were transformed into time ranges, since using it just as provided could result in introducing a bias. Hence, in order to define *startups's age when received a specific funding round*, the variable considered is calculated by the number of years between the respective event and company's foundation (ZBIKOWSKI e ANTOSIUK, 2020).

As a proxy for *venture capital reputation*, the variable has been built considering the aggregated number of startups, backed by at least one of the investors, that either went public or was acquired up until December 2021 (KIM e PARK, 2017) (RÖHM, KÖHN, *et al.*, 2017) . Another factor included is the total number of investors for each investment round (PAHNKE, KATILA e EISENHARDT, 2015).

In order to complete information regarding the investor, three variables are considered as relevant to the study: the *investor previous experience in the respective industry* and with overall investments, taking into account the *total number of performed investments* and *total funds under its management* (KIM e PARK, 2017).

To measure the role of venture capital competition on startups' success, the Herfindahl-Hirshman Index (HHI) is calculated as a proxy. The markets where the investors are allocated are defined according to their geographical location (HONG, SERFES e THIELE, 2019) and, in this case, the computation of HHI is done based on the investment amount for a given market and year. Adapting this approach to calculate competitiveness in private startup markets, the index will measure the consolidation of venture capital dollars among competing startups in an industry.

The main difference between the variables startup and VC competition is that the former is calculated based on the share of the investments received by each startup in a given market, whereas the latter is composed by the share of investment made by each VC company in a given market.

4.4 DATA PRE-PROCESSING

Initially, only the variables related to a raised amount were exposed to outlier detection and elimination study as the venture capital supply is the variable of greatest interest and the platform accepts data insertion from the people involved in the transaction, making the data subject to human error. It was conducted by applying the Inter-Quartile Range (IQR) technique, which relies on the elimination of values above the *higher whisker (HW)* and below the *lower whisker (LW)*.

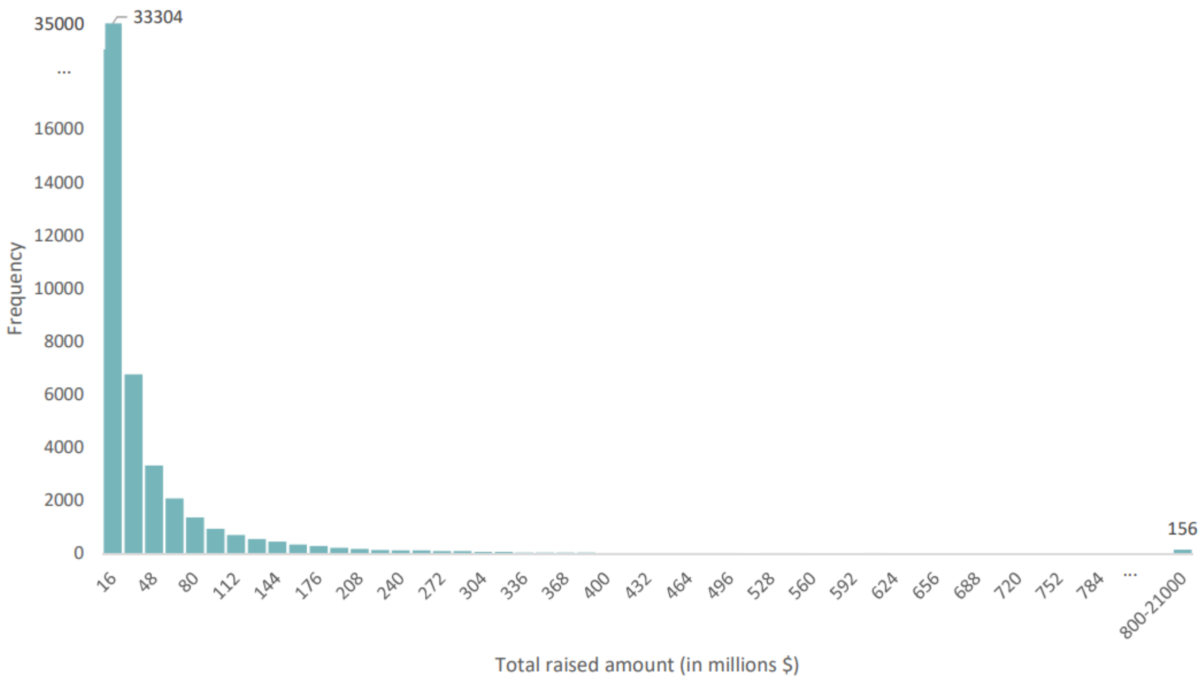
$$IQR = Q3 - Q1 \tag{4}$$

$$HW = Q3 + 1.5 * IQR \tag{5}$$

$$LW = Q1 - 1.5 * IQR \tag{6}$$

Initiating the analysis, Figure 14 provides a panorama on the total raised amount in millions of dollars as it presents the frequency of observations for some intervals of data.

Figure 14 – Histogram on Total Raised Amount



The results of descriptive statistics and the application of IQR technique are summarized in the following tables 5 and 6.

Table 5 – Descriptive Statistics

Descriptive Statistics		
Average	$	34.600.534,40
Median	$	7.994.308,00
Model	$	132.216,00
Minimum	$	321,73
Maximum	$	20.535.850.000,00
# Entries		52012

Table 6 – IQR technique

Inter-Quartile Range		
Q1	$	2.071.595,68
Median	$	7.994.308,00
Q3	$	28.482.957,46
IQR	$	26.411.361,79
HW	$	68.100.000,14
LW	$	-37.545.447,00

Based on these data, it is evident how median and average present distinct values. That can be explained by the asymmetric characteristic of the distribution function. Due to that, it is more adequate to use the median as basis for comparison between categories, which is why a nonparametric method will be applied to test hypotheses (Kruskal-Wallis H test).

Analyzing the tables, the maximum raised amount value is much larger when compared to any other point, which indicates that it is clearly an outlier. Since not all of the companies present in the dataset had time to have more than one round of investment and some of them may have failed during this path, two events that may explain the mode value so inferior to the others, not all values higher than HW were considered as outliers.

Possible alternatives for outliers detection would be analyzing data either according to its location or its specific round of investment. However, as the objective of the study is to study the specificities and compare the VC industry between different countries, treating outliers like the former alternative could cause critical data loss and increase imbalance between categories.

Meanwhile, analyzing data according to the rounds of investment would not be feasible as the classification in round type is not accurate. There are some rounds of investment raised without a clear classification stated, which would add another source of subjectivity to data.

The available data refer to different countries, cover a wide variety of variables and almost 30000 cases were already removed due to null entries. Moreover, as the dataset is subject to model imbalance since almost half of the data corresponds to the United States and there are three times more cases of companies that neither had an M&A nor an IPO, outliers were hardly considered an error in data acquisition. Accordingly, only 156 (0.3%) cases were eliminated from the database. Then, the final dataset consisted of 51856 companies.

4.4.1 **Principal component analysis**

In Crunchbase, company profiles can belong to multiple industries, which became an impediment to use the industry variable as categorical, as the categories would not be independent from each other and there are more than 70 industry variables in the database. Since the industry groups' variables are highly correlated, a PCA was applied to them, in order to find a smaller number of principal components that may lead to more stable estimates of the regression coefficients (RENCHER, 2002).

In order to achieve the final desirable variables, 12 iterations were performed. In each one of them, the data were analyzed, following the steps:

1) Correlation Matrix in order to verify if any variable presents correlation $\leqslant 0.3$ when compared to all other ones, meaning that it should be considered to cut them off if so;
2) Kaiser-Meyer-Olkin (KMO) and Bartlett's Test in order to verify if value expected was achieved;
3) Anti-image Matrix in order to verify if values of the main diagonal are $\geqslant 0.6$, otherwise the variables should be considered to be eliminated from the analysis;
4) Total Variance Explained in order to verify how many components explain a lot of variance in a first moment, and the result reached was that 30 components would be the first trial;
5) Rotated Component Matrix in order to verify which original industry group could be aggregated for each considered component.

Some variables presented correlation $\leqslant 0.3$ in few iterations, resulting in the elimination of a number of variables and others to be considered as already independent from the others and then kept apart from the process to be added again to dataset at the end. There were also variables with low observed frequencies that have been eliminated from the study.

Regarding the KMO test of sampling adequacy, it was valued at 0.574, which is acceptable, even though a value higher than 0.7 would be desirable. It could indicate that there are not sufficient items for each factor. Meanwhile, the Bartlett's test was statistically significant, indicating that the correlation matrix is different from an identity one, in which correlations between variables would be all zero.

With the aim of rotating the component matrix to ease the problem interpretability, the varimax rotation had been chosen in SPSS as this is an orthogonal rotation technique which maximizes the variances of loadings on the new axes. Thus, the rotated component matrix contains all the loadings for each component. The items with high loadings for each component will be aggregated to be used in the research methodology. Table 7 displays the items and component loadings for the rotated components

After rotation, the iteration chosen to be representative consists of 19 variables, 3 of them were eliminated from PCA (Education, Real Estate and Other) to be included on its own to the final dataset. The remaining 16 components resulted from the PCA explain 59% of the variance. Besides that, the scree plot (**Erro! Fonte de referência não encontrada.**) shows that after approximately the 16th component, the curve flattens because the differences between the eigenvalues decline.

Figure 15 – Scree plot

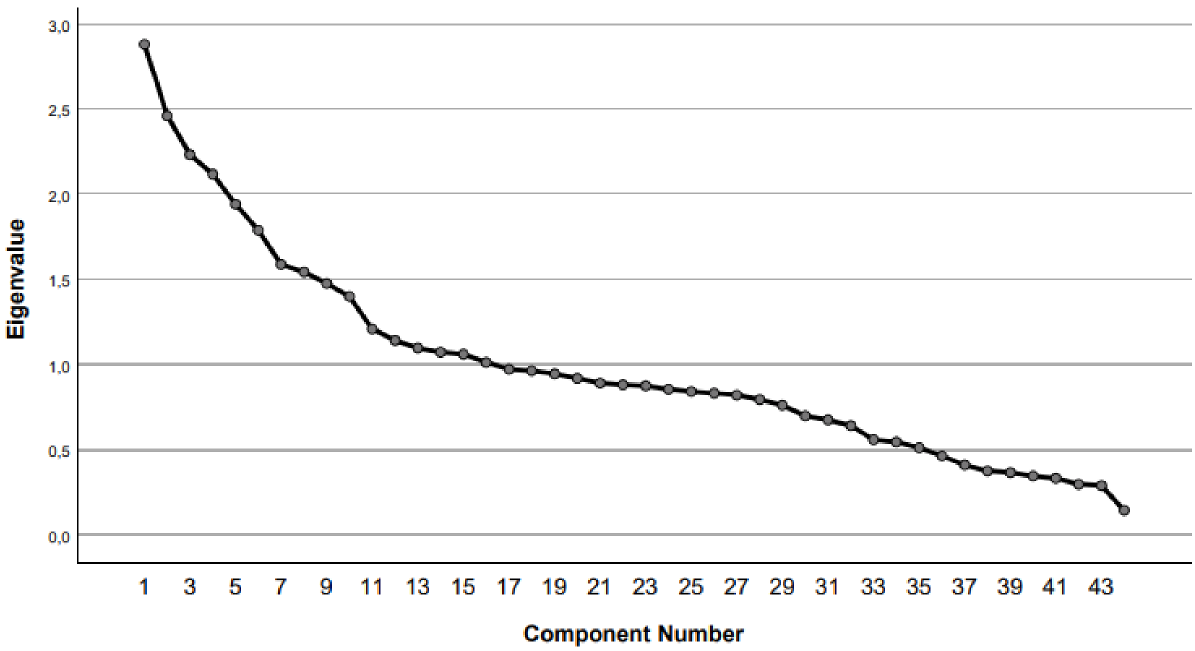


Table 7 – Rotated component matrix

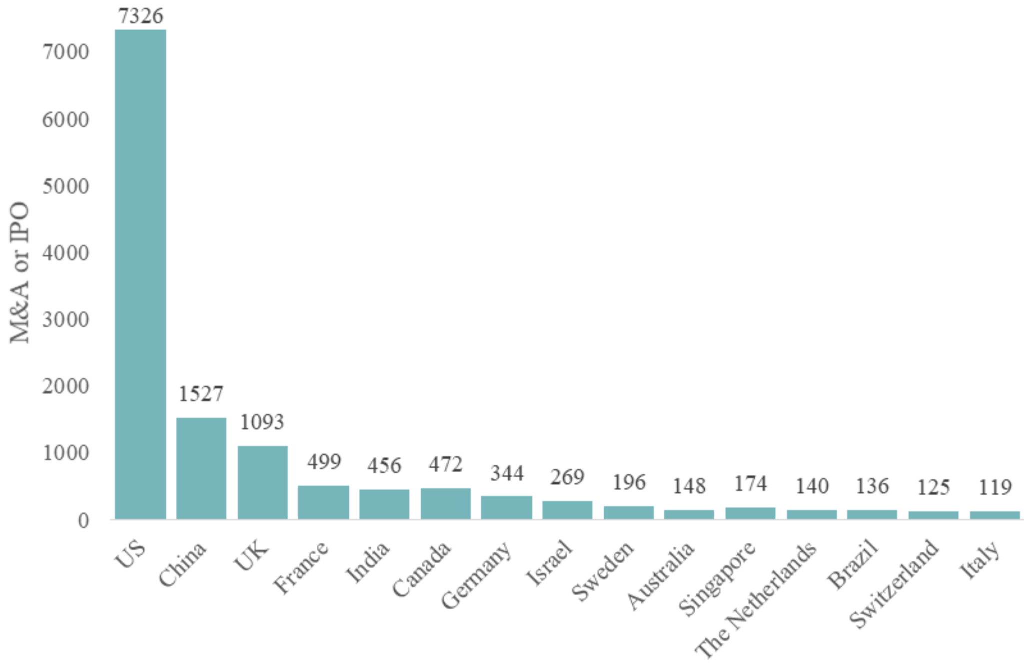
Industry Goups	Component Loading															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Data and Analytics	0,860															
Artificial Intelligence	0,893															
Software	0,526															
Energy		0,863														
Natural Resources		0,781														
Sustainability		0,831														
Biotechnology			-0,815													
Health Care			-0,710													
Transportation			0,436													
Science and Engineering			-0,636													
Music and Audio				0,464												
Video				0,715												
Content and Publishing				0,678												
Media and Entertainment				0,847												
Design					0,800											
Commerce and Shopping					0,407											
Clothing and Apparel					0,849											
Consumer Goods					0,388											
Lending and Investments						0,679										
Financial Services						0,855										
Payments						0,680										
Consumer Electronics							0,837									
Hardware							0,849									
Manufacturing							0,341									
Advertising								0,885								
Sales and Marketing								0,883								
Apps									0,734							
Mobile									0,761							
Platforms									0,584							
Internet Services										0,679						
Messaging and Telecommunications										0,752						
Information Technology											0,585					
Government and Military											0,509					
Privacy and Security											0,740					
Sports												-0,691				
Gaming												-0,680				
Administrative Services													-0,684			
Professional Services													-0,727			
Events														0,711		
Community and Lifestyle														0,446		
Travel and Tourism														0,556		
Food and Beverage															0,608	
Agriculture and Farming															0,768	
Navigation and Mapping																0,828

Source: elaborated by the author

4.5 EXPLORATORY DATA ANALYSIS

To perform initial investigations in our response (dependent) variable, the data was stratified according to the startup's location, considering the frequency of M&As and IPOs observed in the period studied (Figure 16).

Figure 16 – Number of M&As or IPOs observed per country



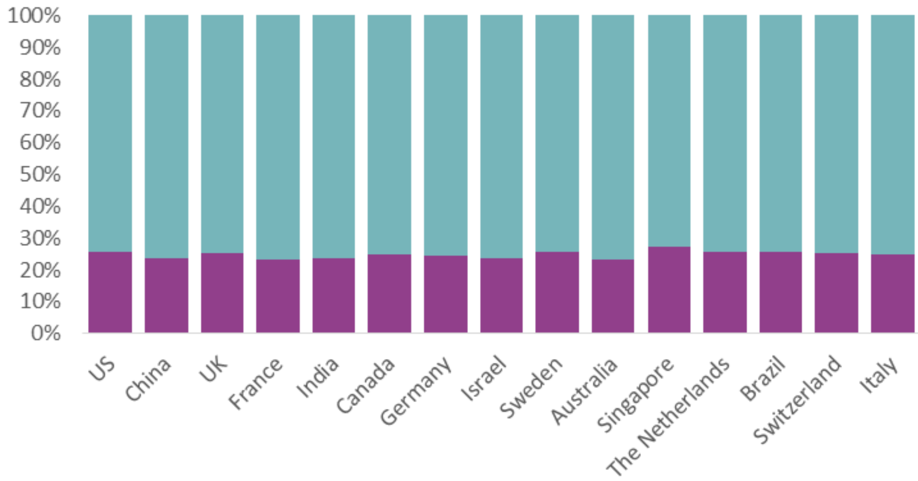
Source: elaborated by the author

The results are exactly as expected showing that the United States presents many more cases of success when compared to the others. It may be due to the fact that Silicon Valley has been the birthplace of many innovation and high-tech companies, forging an extremely welcoming culture for innovative solutions in the country, and becoming a place to look for startups and investors from all over the world since the 1990s, the beginning of our study period (RAO, 2013).

It is curious that, even though the United States presents the greatest number of success cases, the proportion to the total number of startups appears to be similar in all countries, around 25% of total number of observations, as depicted in Figure 17.

This means that the United States also has the highest number of observed failures, which makes sense considering the country as a pioneer that faced several trials and errors encountered until finding its maturity and best practices to be replicated.

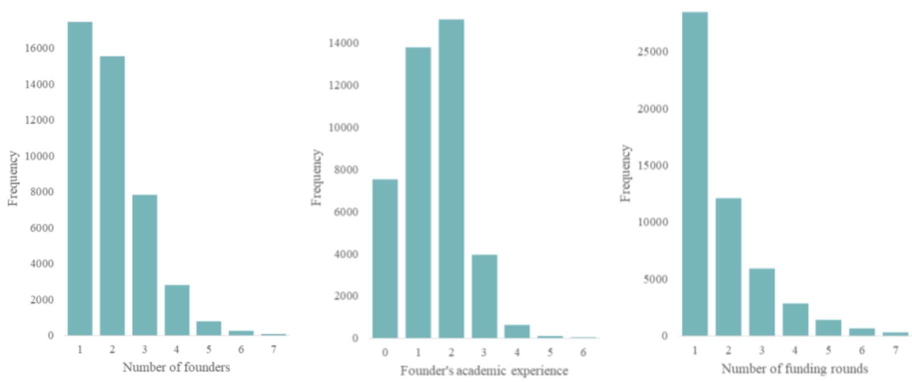
Figure 17 – Proportion of M&As or IPOs per country



Source: elaborated by the author

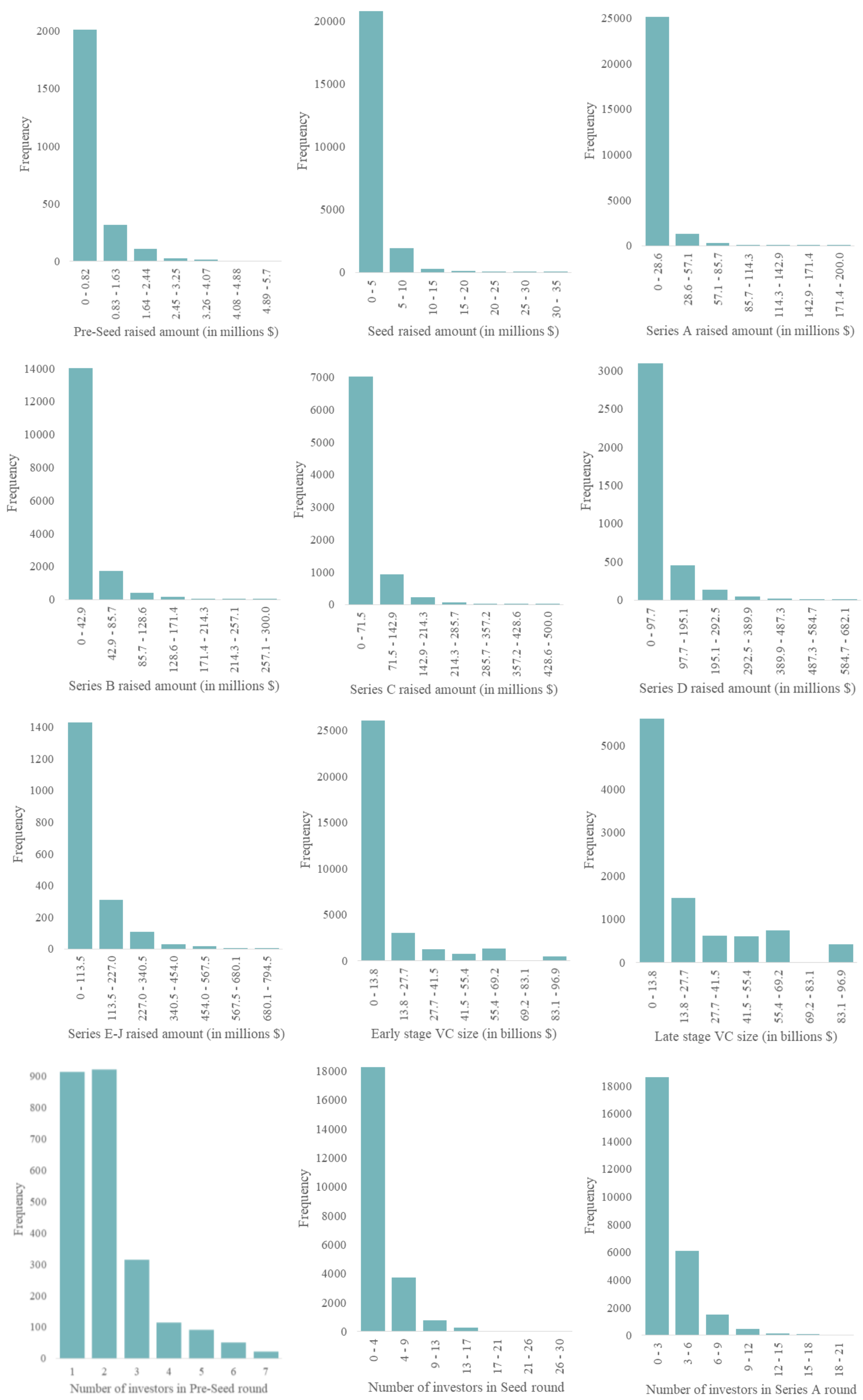
Observing the histogram of all the continuous variables, there is no clear pattern about their distribution. Applying the Kolmogorov-Smirnov test for testing if each variable follows the normal distribution, it has showed a large deviation from the normal curve for all the cases, resulting in a p-value < 0.05 at 5% level of significance, thus rejecting the null hypothesis of the variable following the given distribution.

Figure 18 – Histogram of variables distribution (1/4)



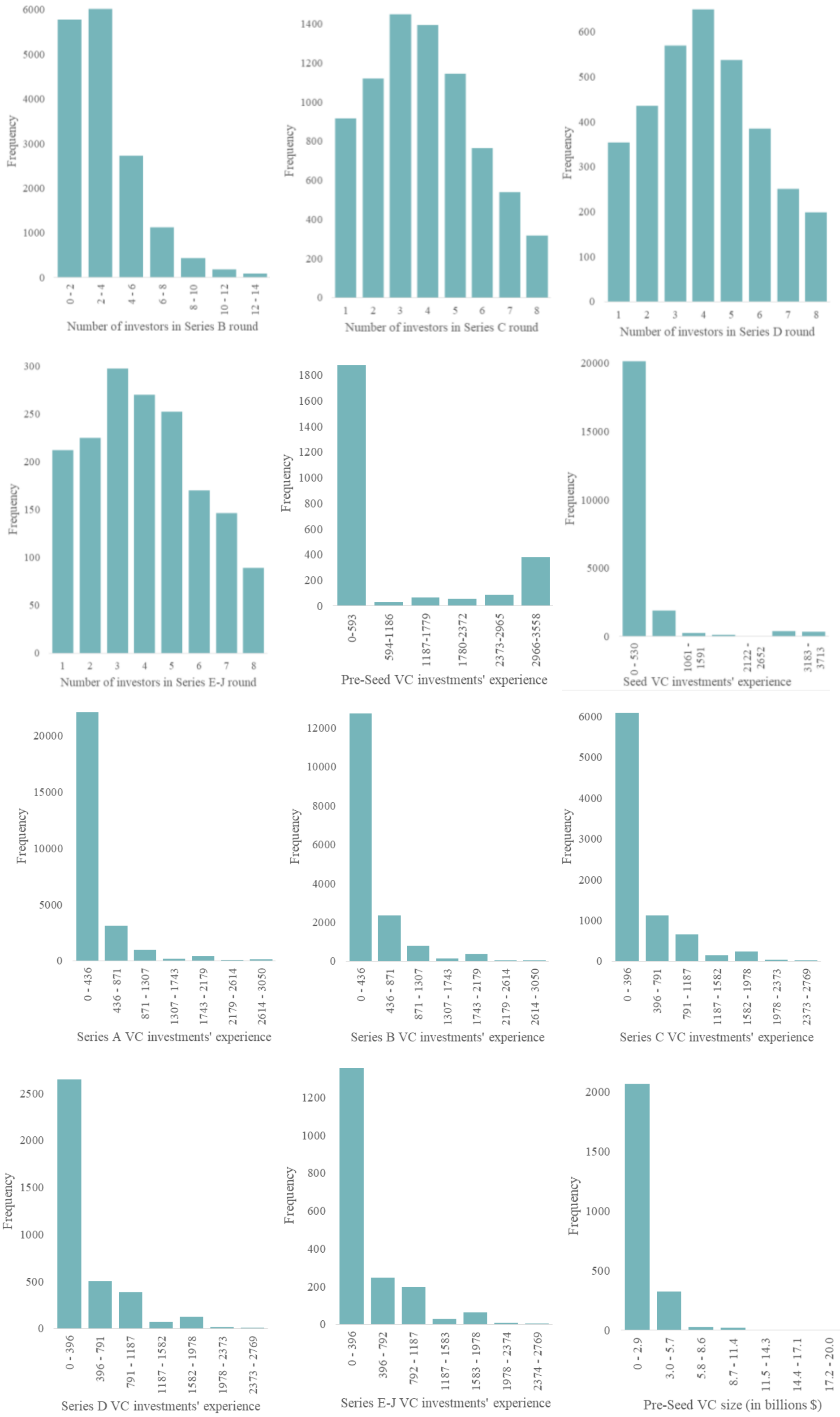
Source: elaborated by the author

Figure 19 – Histogram of variables distribution (2/4)



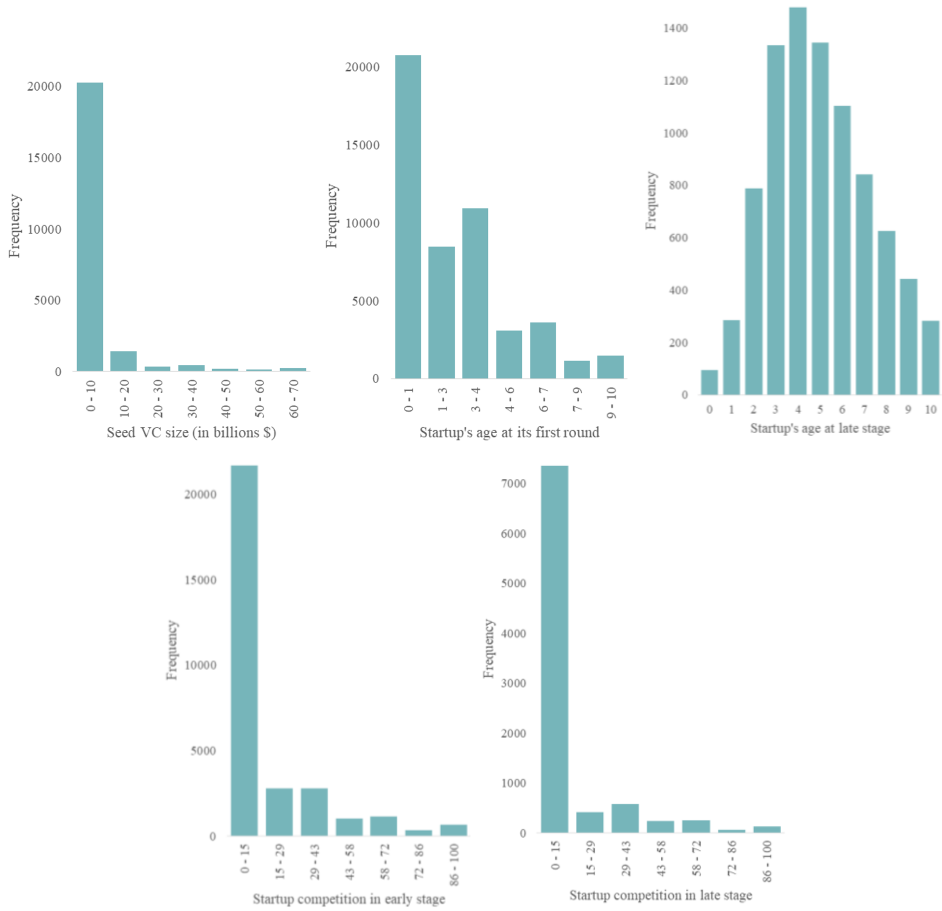
Source: elaborated by the author

Figure 20 – Histogram of variables distribution (3/4)



Source: elaborated by the author

Figure 21 – Histogram of variables distribution (4/4)



Source: elaborated by the author

After that, it is relevant to investigate the proportion of successes and failures for each one of the categorical independent variables (Figure 22). Investor industry's experience seems to have an influence over startup's success as, for instance, 82,8% of the startups that had a M&A or IPO and received a late stage VC investment in series C have been invested by an experienced person in the industry. Meanwhile, for the startups that did not receive and M&A or IPO, just 15,6% of the investors had any experience in the industry, in series C.

For the remaining variables, there is no visual concerns about its representativeness nor its distribution. The logistic regression will determine if they may influence or not.

Figure 22 – M&A or IPO versus categorical variables

Source: elaborated by the author

Despite the initial appearance, it is fundamental to take a closer look at what is the role and representativeness of each variable in the final model. As next step, the non-parametric statistics of Spearman correlation matrix is calculated (table 8).

The variables related to raised amount in each round of investment demonstrates a strong level of correlation ($\geq 0,7$) with the total raised amount. For series B, C and D the value is 0,72, and for series E-J the value is 0,74. This correlation may indicate that the same variable has been taken two times to be explored. Therefore, one of them will be removed when fitting the model, as they can mislead the results.

As the final variables have been selected and analyzed, it is now possible to find the best fit for logistic regression model and interpret its parameters significance and log odds involved.

Table 8 – Spearman correlation

	N_Founders	F_academy	N_FR	RA_PreSeed	N_invpreseed	VC_expinv_preseed	VC_size_preseed	RA_Seed	N_invseed	VC_expinv_seed	VC_size_seed	RA_SeriesA	N_invA	VC_expinv_A	RA_SeriesB	N_invB	VC_expinv_B	RA_SeriesC	N_invC	VC_expinv_C	RA_SeriesD	N_invD	VC_expinv_D	RA_SeriesEJ	N_invEJ	VC_expinv_EJ	A_FirstRound	Startup_comp_Early	Startup_comp_Late	VCsize_Early	VCsize_Late	RA_Total
N_Founders	**1,00**	0,05	0,12	0,03	0,06	-0,04	0,02	0,11	0,19	0,07	0,13	0,08	0,17	0,07	0,06	0,12	0,06	0,08	0,12	0,08	0,06	0,08	0,03	0,06	0,03	0,04	-0,13	-0,11	-0,09	0,11	0,13	0,08
F_academy	0,05	**1,00**	0,02	-0,01	0,02	-0,01	0,03	0,01	0,02	0,03	0,04	-0,02	-0,03	0,04	-0,02	0,01	0,03	-0,02	0,00	0,04	-0,05	-0,02	0,03	-0,06	-0,03	0,02	0,00	-0,03	-0,05	0,05	0,03	0,06
N_FR	0,12	0,02	**1,00**	0,23	0,03	-0,05	-0,05	0,33	0,22	0,18	0,29	0,17	0,23	0,22	0,19	0,28	0,24	0,15	0,26	0,25	0,19	0,25	0,23	0,14	0,30	0,28	-0,24	-0,07	-0,04	0,27	0,31	0,59
RA_PreSeed	0,03	-0,01	0,23	**1,00**	0,28	-0,11	0,11	0,38	0,16	0,01	0,18	0,26	0,18	0,05	0,13	0,18	0,00	-0,23	-0,14	-0,24	0,08	-0,18	-0,28	0,49	0,15	-0,58	0,00	-0,20	-0,08	0,14	-0,16	0,52
N_invpreseed	0,06	0,02	0,03	0,28	**1,00**	0,11	0,34	0,26	0,32	0,08	0,26	0,19	0,25	0,14	0,21	0,38	0,17	0,10	0,16	0,10	0,62	0,52	-0,03	0,25	-0,06	-0,42	-0,02	-0,25	-0,44	0,24	0,19	0,14
VC_expinv_preseed	-0,04	-0,01	-0,05	-0,11	0,11	**1,00**	0,34	0,06	0,18	0,23	0,24	0,04	0,15	0,24	0,15	0,24	0,22	0,19	0,29	0,26	0,28	0,42	0,52	0,23	0,52	0,64	-0,04	-0,37	-0,37	0,27	0,24	-0,13
VC_size_preseed	0,02	0,03	-0,05	0,11	0,34	0,34	**1,00**	0,15	0,21	0,12	0,34	0,14	0,25	0,18	0,23	0,30	0,17	-0,29	-0,13	-0,11	-0,21	-0,16	-0,07	-0,05	0,02	0,19	-0,05	-0,19	0,02	0,31	-0,22	-0,02
RA_Seed	0,11	0,01	0,33	0,38	0,26	0,06	0,15	**1,00**	0,43	0,10	0,43	0,51	0,35	0,10	0,34	0,26	0,06	0,30	0,22	0,07	0,28	0,21	0,04	0,13	0,08	-0,05	0,03	-0,31	-0,29	0,26	0,18	0,69
N_invseed	0,19	0,02	0,22	0,16	0,32	0,18	0,21	0,43	**1,00**	0,22	0,48	0,30	0,49	0,18	0,19	0,30	0,09	0,22	0,29	0,14	0,20	0,19	0,07	0,23	0,11	0,02	-0,11	-0,34	-0,32	0,32	0,25	0,34
VC_expinv_seed	0,07	0,03	0,18	0,01	0,08	0,23	0,12	0,10	0,22	**1,00**	0,63	0,14	0,20	0,34	0,11	0,14	0,22	0,13	0,14	0,20	0,16	0,11	0,07	0,20	0,09	0,03	-0,21	-0,19	-0,16	0,34	0,22	0,08
VC_size_seed	0,13	0,04	0,29	0,18	0,26	0,24	0,34	0,43	0,48	0,63	**1,00**	0,33	0,33	0,36	0,25	0,24	0,27	0,25	0,21	0,28	0,30	0,19	0,15	0,23	0,16	0,07	-0,23	-0,31	-0,29	0,56	0,39	0,40
RA_SeriesA	0,08	-0,02	0,17	0,26	0,19	0,04	0,14	0,51	0,30	0,14	0,33	**1,00**	0,38	0,24	0,48	0,29	0,12	0,37	0,25	0,07	0,22	0,14	0,04	0,13	0,07	0,02	-0,07	-0,20	-0,23	0,39	0,19	0,62
N_invA	0,17	-0,03	0,23	0,18	0,25	0,15	0,25	0,35	0,49	0,20	0,33	0,38	**1,00**	0,13	0,21	0,39	0,04	0,25	0,36	0,06	0,18	0,18	0,03	0,17	0,11	0,06	-0,19	-0,19	-0,20	0,29	0,16	0,25
VC_expinv_A	0,07	0,04	0,22	0,05	0,14	0,24	0,18	0,10	0,18	0,34	0,36	0,24	0,13	**1,00**	0,17	0,12	0,44	0,10	0,08	0,36	0,08	0,15	0,25	0,12	0,15	0,25	-0,21	-0,08	-0,11	0,69	0,37	0,25
RA_SeriesB	0,06	-0,02	0,19	0,13	0,21	0,15	0,23	0,34	0,19	0,11	0,25	0,48	0,21	0,17	**1,00**	0,52	0,26	0,54	0,33	0,09	0,29	0,17	0,02	0,11	0,08	0,04	-0,11	-0,14	-0,13	0,34	0,29	0,72
N_invB	0,12	0,01	0,28	0,18	0,38	0,24	0,30	0,26	0,30	0,14	0,24	0,29	0,39	0,12	0,52	**1,00**	0,16	0,27	0,48	0,05	0,09	0,19	0,01	0,05	0,13	0,03	-0,20	-0,14	-0,13	0,26	0,16	0,43
VC_expinv_B	0,06	0,03	0,24	0,00	0,17	0,22	0,17	0,06	0,09	0,22	0,27	0,12	0,04	0,44	0,26	0,16	**1,00**	0,08	0,05	0,47	0,07	0,08	0,30	0,10	0,10	0,22	-0,19	-0,09	-0,07	0,54	0,44	0,28
RA_SeriesC	0,08	-0,02	0,15	-0,23	0,10	0,19	-0,29	0,30	0,22	0,13	0,25	0,37	0,25	0,10	0,54	0,27	0,08	**1,00**	0,46	0,12	0,46	0,20	0,02	0,33	0,12	0,02	-0,11	0,00	-0,11	0,22	0,36	0,72
N_invC	0,12	0,00	0,26	-0,14	0,16	0,29	-0,13	0,22	0,29	0,14	0,21	0,25	0,36	0,08	0,33	0,48	0,05	0,46	**1,00**	0,11	0,18	0,28	0,02	0,13	0,18	0,06	-0,21	-0,08	-0,17	0,18	0,33	0,39
VC_expinv_C	0,08	0,04	0,25	-0,24	0,10	0,26	-0,11	0,07	0,14	0,20	0,28	0,07	0,06	0,36	0,09	0,05	0,47	0,12	0,11	**1,00**	0,06	0,08	0,42	0,09	0,11	0,31	-0,21	-0,13	-0,04	0,38	0,57	0,18
RA_SeriesD	0,06	-0,05	0,19	0,08	0,62	0,28	-0,21	0,28	0,20	0,16	0,30	0,22	0,18	0,08	0,29	0,09	0,07	0,46	0,18	0,06	**1,00**	0,48	0,14	0,60	0,32	0,11	-0,12	0,10	0,06	0,15	0,27	0,72
N_invD	0,08	-0,02	0,25	-0,18	0,52	0,42	-0,16	0,21	0,19	0,11	0,19	0,14	0,18	0,15	0,17	0,19	0,08	0,20	0,28	0,08	0,48	**1,00**	0,23	0,29	0,40	0,14	-0,20	-0,06	-0,07	0,17	0,24	0,37
VC_expinv_D	0,03	0,03	0,23	-0,28	-0,03	0,52	-0,07	0,04	0,07	0,07	0,15	0,04	0,03	0,25	0,02	0,01	0,30	0,02	0,02	0,42	0,14	0,23	**1,00**	0,11	0,21	0,42	-0,19	-0,12	-0,03	0,30	0,42	0,13
RA_SeriesEJ	0,06	-0,06	0,14	0,49	0,25	0,23	-0,05	0,13	0,23	0,20	0,23	0,13	0,17	0,12	0,11	0,05	0,10	0,33	0,13	0,09	0,60	0,29	0,11	**1,00**	0,42	0,14	-0,10	0,11	0,15	0,15	0,21	0,74
N_invEJ	0,03	-0,03	0,30	0,15	-0,06	0,52	0,02	0,08	0,11	0,09	0,16	0,07	0,11	0,15	0,08	0,13	0,10	0,12	0,18	0,11	0,32	0,40	0,21	0,42	**1,00**	0,24	-0,20	-0,05	-0,02	0,18	0,23	0,33
VC_expinv_EJ	0,04	0,02	0,28	-0,58	-0,42	0,64	0,19	-0,05	0,02	0,03	0,07	0,02	0,06	0,25	0,04	0,03	0,22	0,02	0,06	0,31	0,11	0,14	0,42	0,14	0,24	**1,00**	-0,19	-0,12	-0,06	0,23	0,35	0,14
A_FirstRound	-0,13	0,00	-0,24	0,00	-0,02	-0,04	-0,05	0,03	-0,11	-0,21	-0,23	-0,07	-0,19	-0,21	-0,11	-0,20	-0,19	-0,11	-0,21	-0,21	-0,12	-0,20	-0,19	-0,10	-0,20	-0,19	**1,00**	0,10	0,07	-0,25	-0,28	0,06
Startup_comp_Early	-0,11	-0,03	-0,07	-0,20	-0,25	-0,37	-0,19	-0,31	-0,34	-0,19	-0,31	-0,20	-0,19	-0,08	-0,14	-0,14	-0,09	0,00	-0,08	-0,13	0,10	-0,06	-0,12	0,11	-0,05	-0,12	0,10	**1,00**	0,62	-0,17	-0,16	-0,17
Startup_comp_Late	-0,09	-0,05	-0,04	-0,08	-0,44	-0,37	0,02	-0,29	-0,32	-0,16	-0,29	-0,23	-0,20	-0,11	-0,13	-0,13	-0,07	-0,11	-0,17	-0,04	0,06	-0,07	-0,03	0,15	-0,02	-0,06	0,07	0,62	**1,00**	-0,17	-0,14	-0,07
VCsize_Early	0,11	0,05	0,27	0,14	0,24	0,27	0,31	0,26	0,32	0,34	0,56	0,39	0,29	0,69	0,34	0,26	0,54	0,22	0,18	0,38	0,15	0,17	0,30	0,15	0,18	0,23	-0,25	-0,17	-0,17	**1,00**	0,58	0,45
VCsize_Late	0,13	0,03	0,31	-0,16	0,19	0,24	-0,22	0,18	0,25	0,22	0,39	0,19	0,16	0,37	0,29	0,16	0,44	0,36	0,33	0,57	0,27	0,24	0,42	0,21	0,23	0,35	-0,28	-0,16	-0,14	0,58	**1,00**	0,45
RA_Total	0,08	0,06	0,59	0,52	0,14	-0,13	-0,02	0,69	0,34	0,08	0,40	0,62	0,25	0,25	0,72	0,43	0,28	0,72	0,39	0,18	0,72	0,37	0,13	0,74	0,33	0,14	0,06	-0,17	-0,07	0,45	0,45	**1,00**

Source: elaborated by the author

4.6 MODEL FITTING

To fit the model, the categorical variables will be transformed into dummy ones. For a categorical variable with k classes, k-1 dummy variables will be created, and the remaining one will be the reference (HOSMER e LEMESHOW, 2000), with the binary value of 0 (table 9).

Table 9 – Binary transformation

Binary value	IPO/M&A	Location similarity	Founder's entrepreneurial experience	X' round of investment	VC industry experience
1	Yes	Yes	Yes	Yes	Yes
0	No	No	No	No	No

Source: elaborated by the author

The same idea will be applied to the variables ‘Location’ and ‘VC stage’ (Table 10). In the former, the United States will be chosen as reference since it is the country with the majority of studies already done about it and corresponds to almost 50% of the dataset.

Table 10 – Dummy variables for the ‘location’ categorical one

	Location													
US (reference)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Australia	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Brazil	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Canada	0	0	1	0	0	0	0	0	0	0	0	0	0	0
China	0	0	0	1	0	0	0	0	0	0	0	0	0	0
France	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Germany	0	0	0	0	0	1	0	0	0	0	0	0	0	0
India	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Italy	0	0	0	0	0	0	0	0	1	0	0	0	0	0
The Netherlands	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Singapore	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Sweden	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Switzerland	0	0	0	0	0	0	0	0	0	0	0	0	1	0
United Kingdom	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Source: elaborated by the author

In the case of the ‘VC stage’ variable, the reference decision has been selected as seed, while the model will test the influence of early and late stages (Table 11).

Table 11 – Dummy variables for the 'VC stage' categorical one

	VC stage	
Seed (reference)	0	0
Early	1	0
Late	0	1

Source: elaborated by the author

As just 25% of the sample consists of observed successes, characterizing an imbalanced dataset, to improve the model performance and prevent it from being more sensitive to the majority class and reaching a biased output, it is necessary to address this issue. In this case, a re-sampling technique of oversampling the minority class (i.e., successes), named SMOTE, has been applied.

Even though it is not a work focused on machine learning, the platform called Azure Machine Learning has been used just to get the results after the SMOTE algorithm application to the original dataset, following the structure depicted in Figure 23. The present work is only interested in the resulted dataset after SMOTE component.

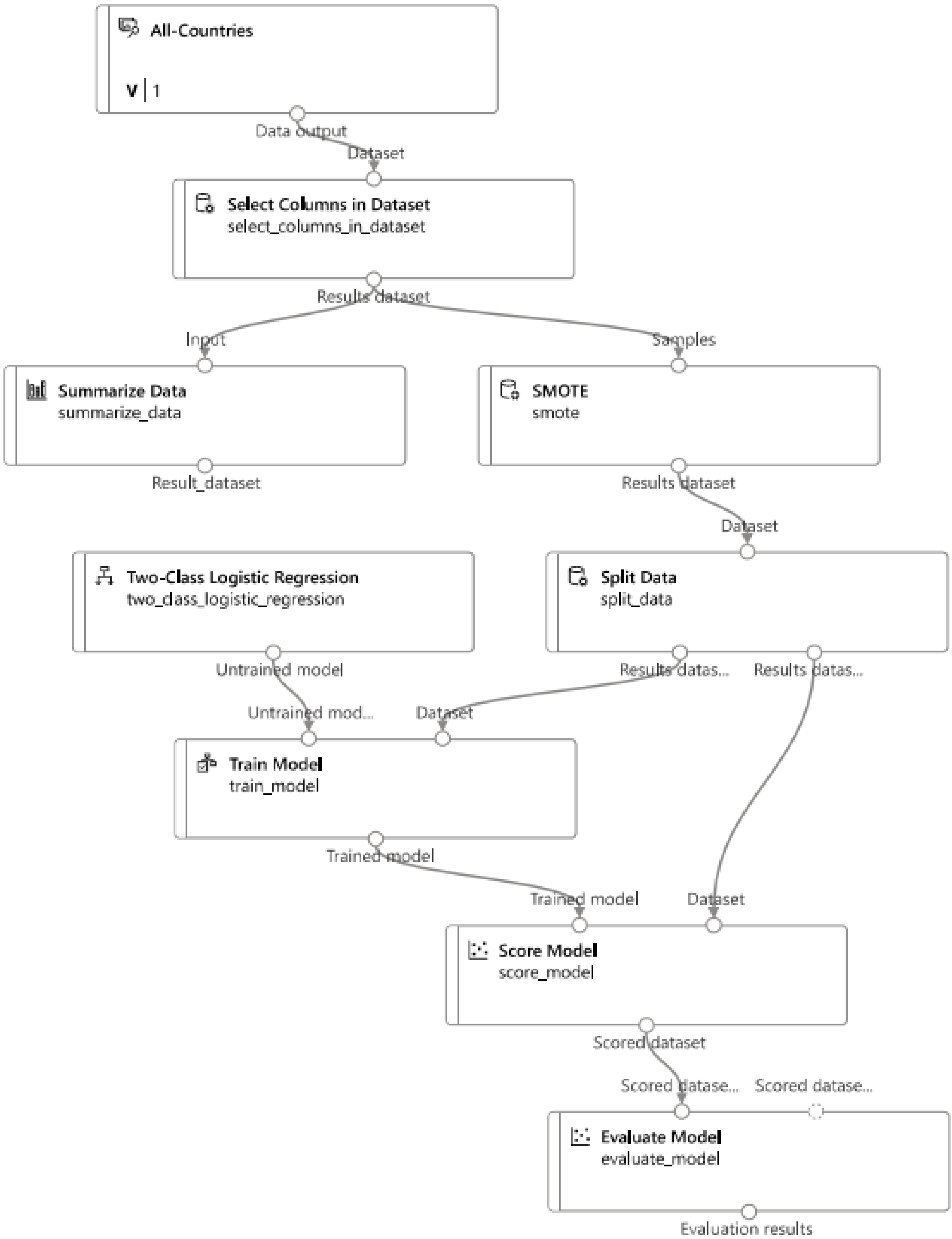
In order to avoid the oversampling not being applied to all the location categories, the SMOTE algorithm was applied to each one of them separately. Afterwards, the final sample is composed by 79552 observations, with 50% of them representing the previous minority class of successes.

The model fitting process is an iterative one. The model is tested until the parameters are statistically significant to be considered in it. In this case, it will happen when p-value < 0,05 at 5% level of significance.

To test the assumption of existing a linear relationship between any continuous independent variables and the logit transformation of the dependent variable, the Box-Tidwell Test was applied using SPSS. For it to be done, all independent variables are added to the model alongside its interaction with their logs. If the interaction term is statistically significant (p-value < 0,05), the assumption is not confirmed. Then, it is usual applying a transformation to the independent variable until the assumption is met.

Two models were evaluated, one following the stepwise forward process for variable selection, and the other following the stepwise backward process. Since the latter presented a lower AIC, it was chosen. The results are summarized in table 12.

Figure 23 – Example of machine learning modelling with SMOTE algorithm



Source: elaborated by the author

Table 12 – Backward stepwise method – coefficient estimates

Method: Backward stepwise				
Variable	**Estimate**	**Std error**	**P**	**Odds ratio**
Intercept	-0,787	0,25	0,000	0,455
Australia	0,191	0,070	0,006	1,211
Brazil	0,292	0,076	0,000	1,339
Canada	0,247	0,034	0,000	1,280
China	0,066	0,026	0,011	1,069
France	0,207	0,040	0,000	1,230
Germany	0,218	0,048	0,000	1,243
India	0,153	0,042	0,000	1,166
Israel	0,174	0,053	0,001	1,190
Italy	0,283	0,079	0,000	1,327
Netherlands	0,406	0,074	0,000	1,501
Singapore	0,425	0,069	0,000	1,529
Sweden	0,322	0,063	0,000	1,379
Switzerland	0,324	0,078	0,000	1,382
UK	0,281	0,029	0,000	1,324
Education	0,569	0,031	0,000	1,767
Other	0,412	0,020	0,000	1,509
Real Estate	0,418	0,034	0,000	1,519
Founder's entrepreneurial experience	0,575	0,019	0,000	1,778
Pre-Seed funding round	0,232	0,053	0,000	1,261
Pre-Seed raised amount	0,000	0,000	0,000	1,000
Seed funding round	0,166	0,018	0,000	1,181
Series A funding round	0,105	0,019	0,000	1,111
Series B funding round	0,087	0,019	0,000	1,090
Series C funding round	0,126	0,042	0,003	1,134
Series C number of investors	-0,047	0,007	0,000	0,954
Series C VC investments' experience	0,000	0,000	0,008	1,000
Series D funding round	0,451	0,041	0,000	1,570
Series D VC investments' experience	-0,001	0,000	0,000	0,999
Series EJ funding round	0,094	0,044	0,034	1,099
VC industry's experience	0,508	0,017	0,000	1,662
Late stage startup competition	-0,001	0,000	0,000	0,999
VC early stage 3-year	0,416	0,020	0,000	1,516
VC late stage 3-year	0,398	0,053	0,000	1,489
$Tourism\_Lifestyle^2$	-0,015	0,002	0,000	0,985
$Entertainment^2$	-0,023	0,002	0,000	0,977
$Adm\_Prof\_Services^2$	-0,033	0,002	0,000	0,967
$Financial\_Services^2$	-0,039	0,002	0,000	0,962

Source: elaborated by the author

The model provides 37 significant variables, with p-value < 0,05. Interpreting the odds ratio, it means that startup being from Singapore increases the odds of success by 1,529, while being from China almost does not increase the odds of success.

It is clearly unexpected seeing the odds of success for a startup being in any country higher than in United States. It may have happened because part of the time window (1992-1999) considered in the study was before venture capital industry starts appearing in the other countries, results being affect also by the internet bubble time in which one a lot of companies broke. Therefore, after balancing the dataset, it may have biased the results due to the great amount of losses in the period mentioned for United States.

Moreover, startup being from education industry increases the odds of success by 1,767 and founder having previous entrepreneurial experience increases the odds of success by 1,778. Startup raising a series D investment round increases the odds of success by 1,570, answering to Q2. Additionally, it is possibly to confirm the question raised in EDA that the investor having previous experience in the industry increases the odds of success by 1,662.

4.7 MODEL EVALUATION

By using the model parameters encountered in the last chapter, it is possible to build the success probability of a startup, based on the intercept and the remaining variables statistically significant found during backward stepwise method. Through the application of this equation, it is viable to estimate the probability of each startup obtaining success and compare it with the original output.

Considering the threshold as 50%, which means that probabilities above this will be predicted as success in the model, it is possible to build the confusion matrix to explore the mistakes found using the model (table 13)

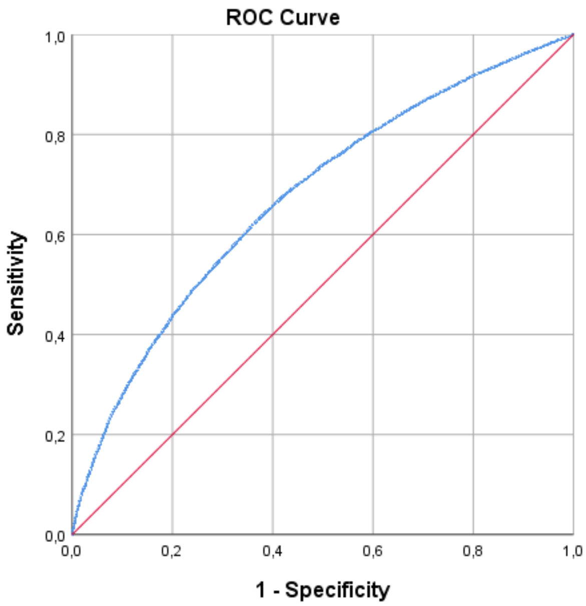
Table 13 – Confusion matrix for threshold of 50%

		Predicted	
		No success	success
Observed	No success	24706	15057
	Success	14350	25439

Source: elaborated by the author

The overall accuracy of the model is 63%, while the null model classified all cases as success and provided accuracy of 50%. The sensitivity is 64%, which indicates the amount of successes predicted correctly over the total amount of successes. Meanwhile, the specificity is 62%, which indicates the amount of failures predicted correctly over the total amount of failures.

Figure 24 – ROC curve plotting 1-specificity versus sensitivity



Source: elaborated by the author

Since the value of area under the curve is higher than 65% (AUC = 67,5%), it is a model of moderate predictive power. However, this value is near the lower limit of 65%, which would indicate a little predictive power. Therefore, the model is an acceptable one, but on the verge of a poorly fitted model.

4.8 KRUSKAL-WALLIS TEST

Since the hypothesis of the available data following a normal distribution was rejected at 5% of significance verified by the Kolmogorov-Smirnov test, hypothesis testing to answer the research questions Q3-Q4 was done through the nonparametric method Kruskal-Wallis H Test.

In order to answer Q3 if countries with higher business dynamism present a more competitive landscape for venture capital-backed startups, firstly we need to define this new variable. Business dynamism has been calculated as the number of new companies' entries in the market in a given year (OECD, 2020). After collecting data for the period considered in our study (1992-2017), two Kruskal-Wallis H test have been conducted.

The aim of the first one is to test the medians of the countries considering its business dynamism during the adopted period in order to group them according to its values.

$$H_0 = All\ countries\ have\ the\ same\ median\ in\ business\ dynamism$$
$$H_1 = At\ least\ one\ country\ presents\ a\ different\ median\ in\ business\ dynamism$$

Kruskal-Wallis H test results in table 14 indicated no statistical equality among all the countries at 5% level of significance (p-value < 0,05), meaning that at least one country does not present the same distribution and median as the others, rejecting the null hypothesis.

Table 14 – Median geographic comparison in business dynamism

Location	**Business Dynamism Median**
Australia	936
Brazil	407
Canada	1189
China	514
France	971
Germany	1206
India	822
Israel	211
Italy	370
The Nethe	692
Singapore	181
Sweden	326
Switzerlan	542
United Kir	2925
United Sta	15165
p-value	0,000
	Reject

Source: elaborated by the author

By performing a post hoc test comparing each pair of countries, it is possible to group them according to its level of business dynamism as depicted in table 15.

Table 15 – Countries grouping by business dynamism

Low	**Low - Medium**	**Medium - High**	**High**
Israel	Brazil	Australia	Canada
Italy	China	France	Germany
Singapore	Switzerland	India	United Kingdom
Sweden		Netherlands	United States

Source: elaborated by the author

Thus, the next step is to test the medians of the levels of business dynamism considering the calculated variable startup competition that is adapted as a measure of market concentration, in early and late stages.

$$H_0 = All\ levels\ of\ business\ dynamism\ have\ the\ same\ median\ in\ startup\ competition$$

$$H_1 = At\ least\ one\ level\ of\ business\ dynamism\ presents\ a\ different\ median\ in\ startup\ competition$$

Kruskal-Wallis H test results in table 16 indicated no statistical equality among all the levels of business dynamism at 5% level of significance (p-value < 0,05), meaning that at least one group does not present the same distribution and median as the others, rejecting the null hypothesis.

Table 16 – Median business dynamism comparison in startup competition

Business Dynamism Level		Startup Competition Early Stage	Startup Competition Late Stage
Low		71,94	60,69
Low/Medium		22,61	20,40
Medium/High		44,09	37,71
High		3,30	2,95
	p-value	0,000	0,000
		Reject	**Reject**

Source: elaborated by the author

By performing a post hoc test comparing each pair of business dynamism level, all pairs of business dynamism levels considering startup competition in early stage are different, only Medium/High and Low pairs do not reject the null hypothesis of presenting the same median considering startup competition in late stage at 5% level of significance (p-value > 0,05).

Table 17 – Business dynamism levels multiple comparisons

Pair	Startup Competition Early Stage	Startup Competition Late Stage
High - Low/Medium	Different	Different
High - Medium/High	Different	Different
High - Low	Different	Different
Low/Medium - Medium/High	Different	Different
Low/Medium - Low	Different	Different
Medium/High - Low	Different	Equal

Source: elaborated by the author

Therefore, as countries with high business dynamism presented lower measure of startup concentration in venture capital market, it may be reasonable to affirm that those countries are characterized by a higher level of startup competition, confirming the research question Q3.

To answer Q4, if there are significant differences in venture capital supply among countries, the following test hypotheses have been defined:

$$H_0 = All\ countries\ have\ the\ same\ median\ in\ raised\ amount$$
$$H_1 = At\ least\ one\ country\ presents\ a\ different\ median\ in\ raised\ amount$$

Therefore, the medians of the countries were tested considering total raised amount and all rounds raised amount. Table 18 presents the results indicating that, in all cases, at least one country has a different median considering the raised amount at 5% level of significance (p-value < 0,05), confirming the research question Q4.

Table 18 – Median geographic comparison in raised amount

	Raised Amount Medians							
Location	**Pre-Seed**	**Seed**	**Series A**	**Series B**	**Series C**	**Series D**	**Series E-J**	**Total**
Australia	$ 138.832,52	$ 1.026.834,30	$ 5.678.500,00	$ 11.396.180,04	$ 26.533.015,44	$ 63.067.292,61	$ 81.493.320,38	$ 3.407.100,00
Brazil	$ 157.082,47	$ 864.089,28	$ 4.542.800,00	$ 12.181.950,00	$ 28.609.188,25	$ 44.924.000,00	$ 191.133.172,81	$ 2.102.245,00
Canada	$ 393.750,00	$ 1.260.715,86	$ 6.144.895,43	$ 14.320.661,18	$ 26.657.095,47	$ 36.961.518,33	$ 82.575.700,00	$ 3.885.162,11
China	$ 360.922,00	$ 946.906,84	$ 3.773.481,86	$ 14.921.819,92	$ 32.251.504,43	$ 55.129.661,72	$ 77.126.000,00	$ 8.498.024,42
France	$ 398.491,40	$ 1.549.349,53	$ 4.915.792,51	$ 10.480.735,73	$ 24.193.376,85	$ 18.828.564,37	$ 78.645.199,11	$ 4.298.702,84
Germany	$ 165.270,00	$ 1.585.775,25	$ 6.550.796,71	$ 12.459.579,85	$ 22.506.849,50	$ 37.119.171,22	$ 38.598.591,01	$ 6.918.471,92
India	$ 238.244,67	$ 1.061.081,75	$ 5.266.031,79	$ 13.755.159,98	$ 28.612.234,63	$ 43.207.243,77	$ 73.610.918,54	$ 4.365.004,11
Israel	$ 211.646,90	$ 1.904.500,00	$ 6.732.122,40	$ 14.192.050,00	$ 27.691.069,31	$ 28.595.542,00	$ 25.554.356,30	$ 9.187.800,00
Italy	$ 117.372,71	$ 774.543,35	$ 3.864.858,76	$ 6.868.897,65	$ 12.199.977,70	$ 8.382.470,01	-	$ 1.718.070,15
The Netherlands	$ 123.541,00	$ 1.187.389,47	$ 6.252.331,81	$ 12.416.861,63	$ 31.865.838,16	$ 44.203.400,00	$ 59.188.206,65	$ 3.956.026,66
Singapore	$ 215.480,00	$ 1.025.324,19	$ 5.615.500,00	$ 13.315.413,97	$ 30.956.300,44	$ 28.595.542,00	$ 93.037.437,67	$ 3.578.311,29
Sweden	$ 139.933,43	$ 1.147.502,95	$ 5.036.850,00	$ 10.042.750,00	$ 19.822.046,10	$ 38.703.301,11	$ 26.482.881,21	$ 2.490.796,00
Switzerland	$ 232.569,00	$ 2.061.519,23	$ 7.144.865,08	$ 15.703.500,00	$ 29.708.750,59	$ 33.576.330,67	$ 96.330.967,41	$ 8.742.282,49
United Kingdom	$ 215.561,66	$ 1.347.720,00	$ 6.675.900,00	$ 14.802.521,37	$ 31.024.151,40	$ 27.669.168,87	$ 52.241.000,00	$ 3.923.762,85
United States	$ 134.772,00	$ 2.746.380,00	$ 16.344.000,00	$ 46.328.200,00	$ 101.079.000,00	$ 160.252.500,00	$ 209.380.000,00	$ 374.082.000,00
p-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Reject	**Reject**	**Reject**	**Reject**	**Reject**	**Reject**	**Reject**	**Reject**

Source: elaborated by the author

5 CONCLUSION

Statistical analysis and insights obtained from the present study might be of help for investors, policymakers and founders as the scope of analysis includes different locations and startups from numerous industries.

As next steps, regarding the peculiarities that each country may be subject to, it would be interesting to develop a model for each one of them and add a variety of variables that might influence the venture capital industry in the specific country of interest.

6 REFERENCES

ABSTARTUPS; DELOITTE. MAPEAMENTO DO ECOSSISTEMA BRASILEIRO DE STARTUPS. [S.L.], P. 29. 2021.

AMIT, R.; SCHOEMAKER, P. J. H. Strategic assets and organizational rent. **Strategic Management Journal**, 1993. 33-46.

ANDERSON, T. W. **An Introduction to Multivariate Statistical Analysis**. Third Edition. ed. [S.l.]: John Wiley & Sons, 2003.

ARROYO, J. et al. Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments. **IEEE Access**, 2019.

AUDRETSCH, D. Entrepreneurship Research. **Management Decision**, 2012. 755-764.

BAKER, W. E.; SINKULA, J. M. Market Orientation, Learning Orientation and Product Innovation: Delving into the Organizaton's Black Box. **Journal of Market-Focused Management**, 2002. 5-23.

BENTO, F. R. D. S. R. Predicting Start-up Success with Machine Learning, 2017.

BER, H.; YAFEH, Y. Can venture capital funds pick winners? Evidence from pre-IPO survival rates and post-IPO performance. **Israel Economic Review**, 2007. 23-46.

BONAVENTURA, M. et al. Predicting success in the worldwide start-up network. **Scientific Reports**, 2020.

BROWN, R.; ROCHA, A. Entrepreneurial uncertainty during the Covid-19 crisis: Mapping the temporal dynamics of entrepreneurial finance. **Journal of Business Venturing Insights**, 2020.

BUSSBANG, J.; MONTUORI, C.; BRAH, W. How to Attract Startups and Tech Companies to a City Without Relying on Tax Breaks. **Harvard Business Review**, May 2019.

CAMPBELL, T. L.; FRYE, M. B. Venture Capitalist Monitoring: Evidence from Governance Structures. **The Quarterly Review of Economics and Finance**. 265-282.

CANTAMESSA, M.; MONTAGNA, F. **Management of Innovation and Product Development**. [S.l.]: Springer, 2016.

CATTELL, R. B. **Personality and Mood by Questionnaire**. London: Jossey-Bass, 1973.

CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, 2002. 321-357.

CIABURRO, G. **Regression Analysis with R**. [S.l.]: Packt, 2018.

COTEI, C.; FARHAT, J. The M&A exit outcomes of new, young firms. **Small Business Economics**, 2017.

CUMMING, D. et al. International entrepreneurship: managerial and policy implications. **Strategic Entrepreneurship Journal**, 2009. 283-296.

CUMMING, D. J.; FISCHER, E. Publicly funded business advisory services and entrepreneurial outcomes. **Research Policy**, Canada, 2012. 467-481.

DALPIAZ, D. **Applied Statistics with R**. [S.l.]: [s.n.], 2016.

DAVIS, S. J.; HALTIWANGER, J. Labor market fluidity and economic performance, September 2014.

DECKER, R. et al. **The Role of Entrepreneurship in US Job Creation and Economic Dynamism**. Journal of Economic Perspectives. [S.l.], p. 3-24. 2014.

DEMARIS, A.; SELMAN, S. H. **Converting Data into Evidence**. [S.l.]: Springer, 2013.

DENNING, P. J.; DUNHAM, R. **The Innovator's Way:** Essential Practices for Successful Innovation. Cambridge: The MIT Press, 2010.

DESSEIN, W. Information and Control in Ventures and Alliances. **The Journal of Finance**, 2005.

FELD, B.; MENDELSON, J. **Venture Deals**. [S.l.]: Josh Wiley & Sons, 2011.

FOX, J. **Applied Regression Analysis and generalized Linear Models**. Third Edition. ed. [S.l.]: SAGE, 2016.

GASTAUD, C.; CARNIEL, T.; DALLE, J.-M. The varying importance of extrinsic factors in the success of startup fundraising: competition at early-stage and networkd at growth-stage, Paris, 2019.

GINEVICIUS, R.; CIRBA, S. Determining Market Concentration. **Journal of Business Economics and Management**, 2007. 3-10.

GOMPERS, P. A.; LERNER, J. The Determinantes of Corporate Venture Capital Success: Organizational Structure, Incentives, and Complementarities. **Concentrated Corporate Ownership**, 2000. 17-54.

GOMPERS, P. A.; LERNER, J. **The Venture Capital Cycle**. Cambridge: MA: MIT press, 2004.

GRAEBNER, M. E.; EISENHARDT, K. M. The Seller's Side of the Story: Acquisition as Courtship and Governance as Syndicate in Entrepreneurial Firms. **Administrative Science Quarterly**, 2004. 366-403.

HALL, B. H.; LERNER, J. The financing of R&D and innovation. In: ______ **Handbooks in Economics**. [S.l.]: Elsevier, v. 01, 2010. p. 609-639.

HAN, J. K.; KIM, N.; SRIVASTAVA, R. K. Market Orientation and Organizational Performance: Is Innovation the Missing Link? **Journal of Marketing**, October 1998. 30-45.

HELLMANN, T.; PURI, M. The Interaction between Product Market and Financing Strategy: The Role of Venture Capital. **Review of Financial Studies**, 2000.

HILBE, J. M. **Logistic Regression Models**. [S.l.]: Taylor & Francis Group, 2009.

HILBE, J. M. **Practical Guide to Logistic Regression**. [S.l.]: Taylor & Francis Group, 2015.

HIRUKAWA, M.; UEDA, M. Venture capital and innovation: Which is first? **Pacific Economic Review**, 2011. 421-465.

HOENIG, D.; HENKEL, J. Quality signals? The role of patents, alliances, and team experience in venture capital financing. **Research Policy**, 2015. 1049-1064.

HONG, S.; SERFES, K.; THIELE, V. Competition in the venture capital market and the success of startup companies: Theory and evidence. **Journal of Economics & Management Strategy**, 2019.

HOPP, C.; STEPHAN, U. The influence of socio-cltural environments on the performance of nascent entrepreneurs: Community culture, motivation, self-efficacy and start-up success. **Entrepreneurship and Regional Development**, 2012. 917-945.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. Second Edition. ed. [S.l.]: JOHN WILEY & SONS, INC., 2000.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. Sixth Edition. ed. [S.l.]: [s.n.], 2007.

KIM, J. Y. R.; PARK, H. D. Two Faces of Early Corporate Venture Capital Funding: Promoting Innovation and Inhibiting IPOs. **Strategy Science**, 2017. 161-175.

KRUSKAL, W. H.; WALLIS, W. A. Use of Ranks in One-Criterion Variance Analysis. In: ______ **Journal of the American Statistical Association**. 260. ed. [S.l.]: [s.n.], v. 47, 1952. p. 583-621.

LANZA, A.; PASSARELLI, M. Technology Change and Dynamic Entrepreneurial Capabilities. **Journal of Small Business Management**, 2014. 427-450.

LITAN, R. E.; RIVLIN, A. M. **Beyond the dot.coms:** The economic promise of the internet. [S.l.]: Brookings Institution Press, 2001.

LOPES, J. et al. Bosiness Dynamism and Innovation Capacity, an Entrepreneurship Worldwide Perspective. **Journal of Open Innovation: Technology, Market, and Complexity**, 2021.

LUONG, H. V. et al. Compressive Online Robust Principal Component Analysis via n-1 Minimization. **IEEE TRANSACTIONS ON IMAGE PROCESSING**, 2018. 4314-4329.

MCKINSEY & COMPANY. Global VC view: Funding startups in the next normal. **Technology, Media & Telecommunications**, May 2021. 10.

MEGGINSON, W. L. et al. Financial distress risk in initial public offerings: How much do venture capitalists matter? **Journal of Corporate Finance**, 2016. 10-30.

METRICK, A.; YASUDA, A. **Venture Capital & the Finance of Innovation**. 2nd. ed. [S.l.]: John Wiley & Sons, 2011.

MILLIKEN, F. J. Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. **Academy of Management Review**, 1987. 133-143.

NAMBISAN, S.; WRIGHT, M.; FELDMAN, M. The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. **Research Policy**, 2019.

OECD. **Declining Business Dynamism: Structural and Policy Determinants**. [S.l.]. 2020.

PAHNKE, E. C.; KATILA, R.; EISENHARDT, K. M. Who Takes You to the Dance? How Partners' Institutional Logics Influence Innovation in Young Firms. **Administrative Science Quarterly**, 2015.

PAVANI, C. **O Capital de Risco no Brasil:** Conceito, Evolução, Perspectivas. [S.l.]: [s.n.], 2004.

PHELPS, E. S. **Mass Flourishing:** How Grassroots Innovation Created Jobs, Challenge, and Change. Princeton: Princeton University Press, 2013.

RAO, A. **A History of Silicon Valley:** The Greatest Creation of Wealth in the History of the Planet. 2nd. ed. [S.l.]: [s.n.], 2013.

RENCHER, A. C. **Methods of Multivariate Analysis**. Second Edition. ed. [S.l.]: John Wiley & Sons, 2002.

RIES, E. **A Startup Enxuta**. [S.l.]: Sextante, 2019.

RÖHM, P. et al. A world of difference? The impact of corporate venture capitalists' investment motivation on startup valuation, 2017.

SANZ-VELASCO, S. A. Opportunity development as a learning process for entrepreneurs. **International Journal of Entrepreneurial Behaviour and Research**, 2006. 251-271.

SAVIGNAC, F. Impact of financial constraints on innovation: What can be learned from a direct measure? **Economics of Innovation and New Technology**, Paris, 2008. 553-569.

SHANE, S.; STUART, T. Organizational Endowments and the Performance of University Start-ups. **Management Science**, January 2002. 154-170.

SONG, M. et al. Success Factors in New Ventures: A Meta-analysis. **The journal of product innovation management**, 2008. 7-27.

STARTUP GENOME. **The Global Startup Ecosystem Report GSER 2021**. [S.l.]. 2021.

STARTUPBLINK. **Global Startup Ecosystem Index**. [S.l.], p. 333. 2021.

STOICA, O.; ROMAN, A.; RUSU, V. D. The Nexus between Entrepreneurship and Economic Growth: A Comparative Analysis on Groups of Countries. **Sustainability**, 2020.

SZETO, E. Innovation capacity: working towards a mechanism for improving innovation within an inter-organizational network. **The TQM Magazine**, v. 12, p. 149-158, 2000.

VANBERG, V. **Innovation, cultural evolution and economic growth**. University of Michigan Press. Ann Arbor, p. 105-121. 1992.

VIRGLEROVA, Z.; ADDEO, F.; ZAPLETALIKOVA, E. Business Dynamism in the World Economy. **Problems and Perspectives in Management**, 2020. 160-169.

WANG, L.; WANG, S. Economic freedom and cross-border venture capital performance. **Journal of Empirical Finance**, China, 2011. 26-50.

WINSTON SMITH, S.; SHAH, S. K. Do innovative users generate more useful insights? An analysis of corporate venture capital investments in the medical devide industry. **Strategic Entrepreneurship Journal**, 2013. 151-167.

ZACH, M. H. If managing knowledge is the solution, then what's the problem. In: MALHOTRA, Y. **Knowledge Management and Business Model Innovation**. [S.l.]: Idea Group Publishing, 2001. p. 16-36.

ZBIKOWSKI, K.; ANTOSIUK, P. A machine learning, bias-free approach for predicting business success using Crunchbase data, 2020.

ZHAO, H. et al. Fault Diagnosis Method Based on Principal Component Analysis and Broad Learning System, 2019.

ZIDER, B. How venture capital works. **Harvard Business Review**, 1998. 131-139.