Degree project

Study of the environment of start-ups in the technological software industry

Report
27/06/2022

Autor: Josep Geli Gallart
Director: Jasmina Berbegal Mirabent
Ponent:
Convocatòria: June 2022

ETSEIB
Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona

ETSEIB

Summary

The objective of this bachelor thesis is to investigate the current status of the software start-up industry in Spain, and to dive deeper into the characteristics of the environment that shape their performance. More precisely, we are interested in determining which factors (namely demographic, economic, political and legal) might hinder their success.

To do this so we first embark on a literature review, in order to figure out the elements of the environment that interact with start-up companies. At this point we rely on the traditional PESTEL framework to characterize the environment. Also, we use the triple helix model, which serves as well to illustrate how start-ups are neither alone nor act in isolation, rather, they are part of a bigger ecosystem and consequently, interact with incumbents, the public administration and with the institutions responsible to create new knowledge, i.e., the universities. Based on these two analyses we determine and characterise a set of factors that might affect start-ups operating in the software industry.

In order to investigate the relevance of each of these factors, an empirical analysis is proposed. Specifically, we gather data from the case of Spain. Using multiple regression analysis we suggest two complementary approaches. First, we consider the region as the unit of analysis. That is, we aggregate the observations at the company level by territories (autonomous communities) and investigate the effect of the specific characteristics of the region where the start-ups are located. Second, we adopt an alternative approach and consider the start-up as the unit of analysis. The results of both approaches are reported and discussed. At this point it should be noted that to better approximate what it is meant by "performance" we use different proxies.

In a final stage, the results were confronted with an expert in the field. By means of an interview we shared with him the findings from the empirical analysis, being able to link our results with the theory of the clusters of innovation.

The main conclusion is that the most important traits of a region in order to host and boost the creation of successful start-ups are easiness of access to finance, a high concentration of talent, market size and being highly innovative.

INDEX

1. Glossary

The following abbreviations have been used throughout the bachelor thesis. Below we provide their meaning. Also, these terms will be introduced in the main text the first time they appear.

SaaS: Software as a Service

ERP: Enterprise resource planning

PESTLE: Political, Economic, Social, Technological, Legal, Environmental

CCAA: “Comunidades autónomas” of Spain

GDP: Gross domestic product

R&D: Resource and Development

INE: “Instituto nacional de estadística”

PIB: “Producto interior bruto”

H0: Hypothesis 0

H1: Hypothesis 1

VIF: Variance inflation factor

ERASMUS: European Region Action Scheme for the Mobility of University Students

2. Preface

2.1. Idea of the project

The idea of this master thesis came to me during my ERASMUS exchange in the Netherlands, where I took several courses related to entrepreneurship, marketing and economics of innovation. It was at that time when I decided that I wanted to get to know at a deeper level the start-up industry.

More recently, getting an internship at Holded, a software start-up based in Barcelona, helped me guide the project in right direction.

2.2. Motivation

There are three main drivers behind the choice of this topic. First, at some point, I would like to put forward an idea and start a company from scratch. Unfortunately, it has not been until recently that I have started getting familiar with the entrepreneurial ecosystem, its actors and the challenges it embraces. Having the chance to do research in the field by means of conducting this thesis it is an excellent opportunity for knowing more about the process and the main difficulties entrepreneurs have to go through in their beginnings. Seeing real examples will put me one step closer towards my ambition to create my own company in the near future.

Second, another motivator is that at first, I wanted to help the company where I was doing my internship. My original idea was to provide them a study reviewing the environmental factors that might influence the start-ups in the software industry, so that they can use this information to improve their business. Nevertheless, after several deliberations the scope of the project was slightly modified and although the final outcome might be of their interest, Holded is not the focus of this thesis.

Lastly, another reason that played a role in the choice of the topic is that over the industrial engineering degree there is not a course on entrepreneurship or how to build a business. Therefore, I was interested in knowing more, particularly the financial part (although how start-ups get funding is not the central part of this thesis), when it comes to being able to attract investors, how many funding rounds does a company need, etc.

2.3. Previous requirements

Before starting structuring and drafting the project, a lot of desk and field research was done. Using a snow ball technique, several papers were reviewed which in turn, open up the way to identify new wants, which were later downloaded and carefully read. Also, grey literature was critical, in order to capture the practitioners' perspective with real life examples of companies, etc. All these documents helped me to get a bigger picture of the start-up industry and decide the direction the project should take.

3. Introduction

3.1. Research objectives

The thesis's main goal is to study the environment of the software's start-up industry (more specifically, SaaS and ERP) and find out which factors (and their operationalisation through indicators) are of utmost importance for the creation and survival of this typology of new technology based firms. Moreover, identify the weight of each one as well as figuring out if it is the combination of all of them or just a few that enhances the creation of software firms.

Another goal the paper attempts to achieve is to establish a methodology that is able to identify almost all the possible indicators that take place in the industry and classify them according to their importance based on a mathematical study.

In order to achieve the previous objectives, another goal needed to be set, which is to find reliable and relatively new data where conclusions could be drawn of.

Lastly, another main goal of the project is to compare the results obtained with the opinion and research done by an expert in the industry.

3.2. Scope of the thesis

The thesis's scope is limited to the extent of the software industry. It does not pretend to embrace the technological sector as a hole. Moreover, the study is based in a given country, i.e. Spain, and uses the territorial division of autonomous communities.

Unfortunately, the database does not comprise information for the entire population of interest. Therefore, the results reported here should be interpreted with caution. It should also be noted that, for the purpose of this work, although we have followed a rigorous methodology, some key information was not available; consequently, we had to rely on a few, yet relevant, indicators that capture the different aspects of the environment. Likewise, performance of start-ups has been measured using alternative measures. One should take the results with a grain of salt, as alternative measures of performance might lead to different results.

Context

3.3. Start-ups in the software industry

Over the last decades, software has penetrated our lives by allowing individuals and businesses to develop daily tasks as well as relying on operating systems. As of today, nearly all businesses rely on some sort of software to maximize the efficiency of their operations and to organize all the data they need to process daily.

The software industry comprises all those companies that develop some kind of software. The development of the software can be done in many ways such as coding, modifying and supporting databases and webpages. [1]

The spectrum of products and services offered by this industry is very broad, yet, the most common ones, which are precisely those that are more widely used, are application software, systems software, maintenance software, custom software, entertainment software, Software as a service (SaaS), Enterprise resource software (ERP), and security software. [1]

Out of all these products and services, the ones that stand out the most are the enterprise resource planning (ERP) and the software as a service (SaaS). The companies that offer these products are the ones that will be analysed in this study. The later one, SaaS, is on an increase trend because the software industry is evolving, moving from traditional "on-premises software" to SaaS. Currently, SaaS covers a third of all the software market. The rationale behind this growth is that SaaS implies lower costs and has an easier implementation compared with more traditional software's. [2]

According to IBISworld [2], the software industry revenue over the next five years is predicted to increase at an average compound annual rate of 2.2%. Nowadays, the Spanish software industry has approximately 15.166 companies which sell software, employ over 96.757 workers and has a market size of 13 billion euros. For the purpose of this work, we will focus on the case of Spain.

3.4. PESTLE analysis

In order to better understand how the software industry (SaaS and ERP) in Spain behaves, it is necessary to obtain a more comprehensive picture. At this point, a PESTLE analysis is pertinent. This analysis consists in examining the political, economic, social, technological, legal and environmental factors that might affect a given industry.[3]

In the paragraphs that follow, we review each of these factors individually. First, we briefly describe their relevance, and afterwards we particularize it for the case under analysis: start-ups operating in the software industry.

Political factors have an important weight in a company's long-term profitability in a specific market. Each region may have different tax rates, wages legislations, intellectual property protection, legal policies, political stability etc. All of which may affect in a way the success of a software company. [4]

Economic factors such as government intervention in the free market, workforce education, labour costs, economic growth rate, unemployment rate or inflation and interest rate also have an impact in the firm which are worth studying. [5]

The regions culture, beliefs and mannerisms will not only impact the culture of the company, but also the working environment and work ethics. Hence, the education level, entrepreneurial spirit, class structure have to be classified for each region to see what is the impact in this area. [4]

Technology is evolving very rapidly over the last years, creating new competitors almost everywhere with different sort of technology that serves the same purpose in cheaper or more efficient way than in the past. Said evolution can be looked by focusing on the rate of diffusion, impact on the value chain structure or the patents application. [4]

Legal factors for software companies are of outmost importance because they can be used to protect data, patents as well as the situation in regards with the employment law. [3]

Lastly, the environmental factors are not as relevant for the software industry as they may be for other markets. Nevertheless, laws regulating pollution, recycling and producing eco-friendly products are important to take into account if the case was to produce a physical product.[23]

3.5. Interactions of the software industry with other agents

The software industry does not operate in isolation, rather it is part of a system in which different actors coexist and interact with each other. One model that can be used to easily illustrate and interpret these relationships is the Triple Helix model. According to this model, a system is integrated by three main agents, namely, the university, the industry and the public administration (or government). Over time, the agents experience two different movements: they evolve around their own axis (they become mature in the ecosystem) and develop interactions with the other agents. The sum of the two movements is supposed to improve their performance as well as to help the ecosystem become more competitive. [6]

From the view of the industry (which is the focus of this thesis) the interaction with the university is meant to create bidirectional channels to allow knowledge flows. Universities provide companies with the latest findings and cutting-edge discoveries; also, they contribute by means of the teaching mission. Through teaching activities, they train the individuals that will later enter the labour market, bringing to the table new ways of thinking and new strategies. [29]

Nevertheless, university-industry interactions pose some big challenges, being among the most widely documented the conflict of interest. For instance, if the university focuses too much on the companies of a sector in order to benefit the industry by only teaching students what they need to know to work in that industry, instead of teaching what should be taught then the university is preparing students to enter the market and not focusing on enhancing their knowledge as much as possible.[29]

The relationship between the university and the government is meant to on the one hand produce well educated people to improve and maintain society and on the other hand to fund the education of the individuals by funding universities. The higher the education students get, the better performance these students will later have on the marketplace. [30]

Likewise, the industry also works closely with the public administration. For instance, both spheres interact through policy and regulation. Government can prevent market failures to occur, make policies to prevent monopolistic situations, provide economic support through subsidies and grants, encourage the intellectual property protection, etc.[6]

The triple Helix model is best applied under a set of conditions for developed countries with democracies and where the market has an important value for the country. Moreover, it is not clear neither proven that these interactions between the agents promote economic growth. [6]

4. Theoretical background

In order to study the software start-up industry in Spain and determine the conditions outside of the company that are necessary for the creation and performance of such companies, it is of utmost importance to establish and define a set of indicators that englobe the firm's environment. Moreover, a clear and precise definition of these indicators is important to then find them for each CCAA and avoid inconsistencies in the research of the data.

The analysis of these factors will help to figure out why in some places there is a higher concentration of start-up software companies than elsewhere and why are start-ups performing better in specific locations.

Taking into consideration the two models explained before, Triple Helix and PESTLE, a set of theoretical factors have been chosen to be studied to shed light in the most amount of area regarding the environment of a software company.

There is another model that helps to explain the success of certain industries which is the Porter's five forces model. This model is based on the study of five factors that are competition in the industry, potential of new entrants, power of suppliers, power of customers and threat of substitute products. But since this model is used to study the industry itself and does not allow to explain why some companies inside that industry perform better than others based on their location, it will not be used in this research. [7]

Therefore, considering the before mentioned models, a set of components have been chosen that might play a role in explaining the success of a software company and its performance. These are economic components, technological components, political/legal components, demographic components and cultural factors. It is worth mentioning that inside said components there are many traits that conform them and that this research will select the ones with the most weight and that data is accessible. [8]

Before explaining each of the components, there is a difference between the theoretical indicators, the ones just mentioned, and the indicators used to analyse the data which will later be explained and that have to be calculated manually using data extracted from reliable sources.

4.1. Economic components

As derived from the Triple Helix model, the geographical location plays a paramount role, as the agents with those a company will interact, to some extent, will shape its growth, hindering or fostering it. In this regard, the level of economic development of the territory is worth of examination. In the specific context of start-ups in the software industry, key factors related to the economic environment that may play a role are:

- Inflation
- Subsidies and government help
- Tax exemption
- Gross domestic product (GDP)
- Income per capita
- Currency

The inflation a country has is a good indicator of economic growth. Inflation can be defined as a "sustained rise in overall price levels". A moderate rate of inflation is positive because it signals an economic boost, but a very high rate is not desired because it leads to increasingly high prices for the consumers. When the economy grows, businesses and customers spend more money on goods and services, which leads to an increase in demand and all together leads to producers raising prices. Hence, the inflation increases. When economic growth slows down it means that demand decreases in comparison to the supply of goods and a period of deflation occurs. [9]

Inflation is an aspect to consider specially when making investments. Since start-ups require, from the very beginning, external funding, whether national or international, it is more appealing for the investor point of view if the inflation rate is moderate. Two of the main arguments are the following. First, a very low inflation rate would indicate that there is almost none economic growth. Second, an extremely high inflation rate would be very risky, for instance, if a return on investment is planned at 5% annually and the inflation rate is at 6%, this means that there is a negative return of 1% when adjusted to inflation. From the above, it can be concluded that inflation is relevant. Nevertheless, given that this research only considers one country (Spain) and our analysis is at regional level (within Spain), Inflation does not seem to be so critical. [9]

Another subcomponent that does actually vary from CCAA to CCAA is the subsidies given and the government help companies from that region get. Those regions that are able to reward companies financially for their ideas and business plans are the ones that attract more some type of companies. Moreover, if a location gives away subsidies, these allow the start-ups and hence the industry to become more efficient and eventually reduce the price of goods and services.

Nevertheless, there is a big risk in conceiving a lot of subsidies to start-ups. Such financial help could make firms lazy and less innovative to achieve their objectives because they know they have a financial pillow given by the government. When the government stops giving subsidies to these companies, they might fail because they had been living of the money received and not from the one earned through actual profits.

The one kind of government help that is seeked by many new firms is tax exemption, being able from the first years not to pay taxes and prioritise the companies' growth since the money going to taxes is spent in improving and growing.

Moving on to another indicator that allows to measure the economic health of a country is the gross domestic product (GDP) since it measures the market value of goods and services produced within the borders of a country in a period of time. It is also looked closely by investors because it shows the possibility of growth in that country. If the GDP of a country is increasing steadily it means it has a fast-growing economy with opportunity to invest in. [10]

The income per capita of a region is a measure of average living standards and gives a hinge to producers if their product can be afforded by potential customers or if there is a market for them. For instance, a software company selling a financial software aimed at small companies to monitor in detail all of their accounting would not be very successful somewhere where the income per capita is very low if they are thinking of charging high prices. [11]

Lastly, the region's currency has to be analysed in comparison to its valuation to other currencies. If the valuation of a currency has ups and downs regularly it might be very risky from an investor point of view to invest there. Moreover, if the valuation increases in comparison to other ones, then that country will have it difficult to export goods since outsiders are not willing to pay that much if prices remain the same. And the other way around happens when the valuation decreases because the country gets poorer and cannot afford to import products as much as before. [12]

From an entrepreneur point of view there has to be a very stable currency to minimize risks and therefore reduce as much as possible the negative outcomes that could happen that are out of his control.

4.2. Technological components

The technological component refers not only to the advances in an industry in a specific location, but also the pace at which these advances took place. Advances in technology can refer to improving machines, making some techniques more efficient, developing new tools and skills to execute a function or improving an organizational method. [13]

Each region has its own pace in developing technology and implementing it in their sectors. The pace and the effort companies put into said component will be of study in this paper.

Although new born companies do not usually have a department of R&D as some of the bigger companies do, they do have an impressive level of creativity especially when it comes to the product or service they are selling and how can they sell it as well as protect it. Thus, reap the benefits of their idea.

A good indication of how technology is developing in a region is for instance looking at the number of patents that companies are applying to, because it means that they are innovating and developing new products or techniques. Although sometimes when companies patent an idea, they sometime do so with the idea of avoiding competition in a sector but that company has not in its plans to use that idea. Nevertheless, this case is still a sign of technology development as well as of innovation. [14]

If the patents were the only indication of advance it could have some hidden traps due to the fact that some companies may apply for a patent but then not execute into existence the device or technique. Hence, it is important to look at the new products developed sold into the market or different techniques of selling a service. In this way it can be seen as well the success start-ups have in eventually getting into the market the ideas and not just having and patenting them. [14]

When considering software start-ups, a consequence of excellent technology development is usually the easiness in attracting partners or investors. If the technology the company is developing is new into the market and is out of the box, then investors will eventually follow.

For all of the reasons mentioned above to study this component is crucial because there will be some CCAAs where there is almost non technology advancement and where the pace is very slow, whereas in other areas that pace is exponential. Taking into consideration this component will be of enormous help in figuring out what situation is the best for helping companies grow.

4.3. Political and legal components

When it comes to the legal laws and politics that a region has, it can impact directly in the decision of an entrepreneur in starting a business there. The less difficult it is to build up the company faster, the better for the entrepreneur. Moreover, if there is some kind of protection and motivation economically speaking for the company then the more attraction that region will probably have for start-ups. [15]

Legal policies regarding taxes are important because they may be a source of attraction for talent. If a region imposes high taxes for income, then talent and people working in a sector may decide to move elsewhere. For instance, the state of California in the US has recently increased their taxes on wealth and one of the consequences was that as is affected the wealthier people considerably, it was a good part of them who decided to move elsewhere with lower taxes. Therefore, it produces a runaway of talent in some cases. [16]

Policies that involve the protection of the employee and benefit them is a way to attract talent that wants to work in a healthy environment, knowing that there is a strong regulation protecting them. [17]

Another kind or regulation that is of utmost importance specially for companies in the software industry is the protection of the property rights. Being able to secure the idea that one has during a long period of time is essential if the entrepreneur wants to get financial benefits from that idea. Besides, this kind of protection boosts innovation because there is a reward for the person that develops an idea and puts it into the market. If no such policies were in place then there would be no reason for someone the put so much time and effort in an idea if then someone else will "steal" the benefits. [18]

By enabling protection to the inventor but at the same time making public the information of the patent there is constant evolution of improvement where the idea cannot be stolen or implemented by anyone else, but it can be used to improve upon it.

When studying different CCAA, it is important to know in some degree the government intervention in the economy and sectors of that region. The objective of government intervention is to change decisions made by some organizations or individuals and take some control of the situation on the marketplace. The government can intervene in several ways such as with specific taxes or imposing maximum and minimum prices. Some of the advantages are to prevent monopolies, correction of negative externalities or environmental protection. Whereas the disadvantages are lack of incentives, excess of limitation and limiting the free market, limiting the choice of products or services, etc. [19]

4.4. Demographic components

Besides all the previous components in the analysis that may seem to have a bigger impact, the demography of the region is also worth studying because it can impact in the profile of employee a company has as well as how well are these educated in that region.

First of all, when looking at the demography of a region the size of the population and the population per square meter is worth knowing for the entrepreneur to know whether there is a niche market he can aim at or not. Besides, for software companies that may sell a product or service at a high price it is interesting to know the income level of the population to see if the potential customers have the financial power to afford the product being sold. [20]

One of the most crucial aspects to look into of a region is how well educated the labour force is. If the region has few universities or there is little money spent in improving the education system as well as attracting students from elsewhere, then a start-up will have it difficult to find the wright fit for their jobs. [20]

A part from helping to determine the decision of whether to enter or not that region, demographics can also be used as a marketing tool to find the best way to reach customers, study their behaviour and determine the size of the potential market. It can basically be used to build a customer-based profile. [21]

4.5. Cultural components

Following the previous component, very similar information can be gathered through the cultural factors of a region. In this case it can be used as well to help decide if it is a smart business decision to enter the market in that region or to use it as an advantage and attract the talent from that location by adapting to their customs and values.

The cultural factors of a region can be divided into several areas. The first area involves the material culture, meaning the presence of wireless communication, excellent Internet, access to personal transport such as cars and buses. The second area involves the language spoken, because it could have an impact in the marketing techniques used, brand names or surveys made to customers. Thirdly, the most abundant religion can have an impact in many aspects of the business because most of the employees will follow the customs and norms of that religion and could affect the working hours, the relationship between employees and employer, the labour organization and hierarchy.

There is also a separate area that englobes the ethics and values that shape the community. In our case they are differences in culture between CCAA's that involve different ways of working and of social organization.

5. Data and method

5.1. Sample and data

Data has been gathered from several sources. Two different websites of prestigious recognition in the world of start-ups have been used, these are Crunchbase and INE ("Instituto nacional de estadística"). Each of them as explained below have been used for different metrics.

The analysis is done in two levels, by CCAA and by company. Moreover, all the study is based in the Spanish market.

5.2. Approaches

To quantify all the indicators, several Excels have been used to collect and keep together all the data. Each of the indicators have had a sheet of excel where the information downloaded from the websites was transformed by formulas into indicators per capita that are more representative of each region. It is the easiest and best way to keep the information because it is quicker to change and add if necessary. As well, it is needed to be in this way for then transporting all of the data into different software's where we will apply studies and mathematical models to analyse the data and extract solid conclusions.

In the website Crunchbase there is information of all companies in Spain regarding the location, number of employees, number of national and international investors, funding rounds, money raised and different other metrics. After all, Crunchbase is a website that in order to access all of the information you have to pay a monthly fee. So, filtering for software companies (SaaS and ERP) which are the most abundant ones in the industry and the ones that had all the data we required, the data was put manually into the Excel which has all of the information mentioned above. Because to download all of the data you have to pay a several amount.

A glimpse about how the excel looks can be seen in the image 1 found in the Annex section.

In the Excel with all the software companies, the most important columns are the location, the total funding amount, and the column of performance. It is worth mentioning that not all the companies had the data we required regarding the amount of money they raised as well as the funding rounds. Therefore, we had to take into consideration later on in the study of different mathematical models, the one's that did have the data. In the Annex (image 2) we show the number of companies for which we had solid data of out of all the software companies in that CCAA.

The other website we used to determine the majority of the indicators later explained is INE. Through the official government website, all kind of data could be downloaded in excel format that was much easier to work with compared to Crunchbase. For each indicator an excel sheet was created and a formula was applied to transform the raw data into the indicator we will use in the model to analyse the data. At the end, when all the indicators had been found, an Excel with each CCAA and all the indicators was made in order to introduce it to the software that will help to analyse it (see image 3 in the Annex section).

It is important to mention that the financial data extracted from different years from Crunchbase does not take into account the inflation rate of that year. Hence, for a more precise value, for each funding amount of every funding round, the inflation rate should be considered if the money values are studied together.

Each indicator will be explained further below, but the main idea was to have as many indicators as possible for each CCAA.

Two different excels were made, one with every software company in Spain (SaaS and ERP) with its respective money raised and another with each CCAA with several indicators. The goal is to study both excels through mathematical models that will help us to understand from the side of each CCAA and individually from each company.

5.3. Calculus of the indicators

Given the two approaches, at the region and company level, two different databases have been constructed. In the first one, the level of analysis is the the CCAA. In the second one, data are presented at the company level. In this latter case, we also included columns with the information regarding the CCAA in which the company is located.

5.3.1. Performance

As mentioned before, the collection of the data was divided into two excels. In the Excels where all the companies are found, an indicator was made in order to track the level of success of a company. By success we refer to the capability of that firm to keep growing and getting better rather than not making profits and going to bankruptcy.

So as to quantify such success an indicator named “performance” was created. According to the article [3] there are many ways to measure the success, but a very good one is to look at the percentage value increase during a period of time of a start-up between two funding rounds. So basically the performance will be given by the percentage increase in the money raised from one round to next round, because measuring the money raised is directly proportional to the value that company has. [8] If not no investor would put high amounts of money to put forward the idea of that company.

But there are a series of limitations in this approach that weaken a bit the analysis as a whole. Firstly, this approach can only be applied to those companies that have at least two founding rounds. If a company has only undergone one funding round then no measure can be made about its success or performance. Secondly, the ideal situation would be the measure the value in growth during a fixed period of time for all the start-ups, but every company goes through funding rounds at different stages during their growth, some quicker than other. So no fixed period of time can be used homogeneously for all the companies.

The ideal and theoretical formula for said indicator would be “the difference between the post money valuation of round n and the pre-money valuation of round n+1, divided by the number of months between both rounds”. [8]

From the data we extracted, as mentioned before, some companies do not have the information regarding the financial amounts. Therefore, the performance indicator was made only for those companies from which we have data of financially. Going back to the second limitation of the approach that mentioned that no period of time can be picked, and since Crunchbase only has some dates of some funding rounds, the performance indicator was calculated as the division of the money raised in the last funding round with the one raised in the previous of the last funding round and then make the average for all the companies of that CCAA, so between the last two rounds of financing.

Although it is not as precise as it would have been if considering each time period of every company, it still has a good correlation and direct relation with the performance of the company.

Moreover, another indicator used was the “Seed round achievement”. This indicator is based on ones and zeros. A one being that the company has reached the seed round and a zero being that it has not. It will probably give an estimation of the success of each company.

5.3.2. Demographic indicators

For this aspect, many indicators were selected in order to define each region through very precise data that could have an impact at the end when we analyse each factor in the model. So many indicators were selected because at the end of the study the model will tell us which of them are relevant and have an actual impact, so including more that will not have an impact, causes no problem at all. Moreover, at this point, we cannot know which are the ones with greater impact; hence including them all is the best solution.

Firstly, the population of every CCAA was included, which was extracted directly from the INE. So no transformation was necessary in this case. It is important to take into account the density of the population and the scope of the market for potential customers. Then, the PIB *(“producto interior bruto”*) per capita for each CCAA was calculated dividing the PIB with the population of that region. This indicator indicates how much does it produce in all of the sectors of that economy. Although it is not taking into account the PIB only of the software sector, it will give a wide indication about the economy of that place.

Moving on, the next one is the unemployment rate, which was obtained directly from the INE. It gives a view about the labour situation in that region as well as the demand and offer for specific jobs. It has to be taken into account that the lower the unemployment rate the better. A pretty unconventional indicator is the number of birth rates per a thousand people, it has been obtained directly from the INE as well and it gives a view about the number of future employees the region might have as well as the number of potential customers. If a region has a very low birth rate then in the next twenty to fifty years the market might not have grown and there is few room to grow and expand as a company. This indicator has the limitation that from one year to another due to X reason there could be a dramatic change in birth rates.

Another indicator is the income level per capita. Such indicator is extracted from the INE directly and indicates the purchase power of the customers. We need to know if the people from that location can afford to buy a specific product or service, therefore the higher the income level per capita the better.

Apart from the mere statistical indicators there is also an area about the demographics of a region that has to be precisely measured, and that is to measure the level of educated labour force. Two indicators were made in this regard. The first one is the number of universities per capita, which is calculated dividing the number of universities in a CCAA by the population.

The second indicator is the one named educated labour force which indicates the number of graduate students from degrees related to start-ups of software per capita. So firstly, from INE, an excel indicating the number of students from every major degree for each CCAA was downloaded. Then, the degrees that were related with building a software company were selected. The degrees were audio visual techniques and media communication, economy, social and behavioural sciences, management and administration, business and administration, law, psychology, mathematics and statistics, computer science and engineering. A software start-up not only needs engineers, but also people related to marketing, finance, law, etc. Then I summed up all the students from these degrees for each CCAA and divided it by the population of that region. This indicator before applying any model seems pretty useful to identify talent and well-prepared employees.

5.3.3. Cultural indicators

It is very tricky to measure quantitatively the culture of a region because there is no data that quantifies customs and values of people. Nevertheless, an indicator related to both culture and technological advancements is the innovation indicator of each region. The data was extracted from the official webpage of the European commission [22]. The study uses directly the indicator from them, which is calculated by taking into account 21 different aspects. For instance, digital skills, innovation expenditures per person employed, PCT patent applications, design and trademark applications, sales of new-to-market and new-to-firm innovations or R&D (resource and development) expenditures business sector. The European commission calculates an indicator for every CCAA considering all of these traits, and it is the one used as well in this study.

It is also considered a technological indicator because inside the calculus of the innovation indicator, the formula takes into account several traits regarding patent applications, IT skills, etc.

5.3.4. Legal indicator

To study the legal advantages or disadvantages in a region, an indicator about the amount of taxes paid by the population is created. This indicator will give a glimpse about what are the CCAA's where less taxes are paid per capita. This might influence talent or entrepreneurs to make a decision about where to move to live, since everyone's desire is to pay less taxes.

In order to get this indicator, data from the website of hacienda about the amount paid by population in taxes in each CCAA was downloaded. [23] Then divided this amount by the number of people in that location. Moving on, the percentage of the total was calculated for each CCAA and that gave us the final indicator that will be used in the model to analyse precisely the data.

It is worth mentioning that the amount paid in taxes per region takes into account the taxes paid equally through each CCAA plus the taxes called "*propios*" which depend on every region. Although at first this indicator might seem to tell about where people pay more taxes, it also indicates in a sense where are people making more money. Since there is a direct relationship between earning more money and paying more taxes.

5.3.5. Access to finance indicator

A very key aspect for a company that has recently been created is the ability to attract financial help from investors. It may be the most important factor for the entrepreneur because without this help he or she may not put into existence their idea.

I considered it to be a factor that will have weight in the analysis because theoretically a company based in a region with easier access to finance than another elsewhere will have better changes to succeed. It is worth studying as well how are the international investors looking at the different CCAA's in Spain, and where do they invest more. Besides, regarding national investors, a look at the density of Venture Capitals, angel investors and accelerators throughout the country will give a glimpse about where are these organizations and individuals concentrated.

All in all, to look at said area of financial help, three different indicators have been created to shed light into the topic. The first one is the amount of money invested in software start-ups for each CCAA, so as to know where does the money flow to primarily. This indicator was created by taking the data from the excel obtained from Crunchbase where we had all of the companies put together with the financial data and dividing it by the number of companies we had data of. Filtering for each CCAA, the sum of the money of all the companies was put together.

Another indicator that allows us to know the presence of international investors in the software start-up industry is the percentage of international investors out of the sum of these and the national ones. Some research [24] shows that international investors only pick to invest in those companies that have a great business plan, an assessment risk has been made and overall have a great idea with an excellent management team. Hence, knowing where the international investors money is going will indicate where the most promising companies, which will may or may not have a correlation with some environmental factors that have helped these companies along their creation and journey.

In relation to this last indicator, another was made that looks at the average number of funding rounds in that CCAA. Which tries to explain if the start-up in that region is likely or capable to obtain more financial help as well as performing better. Linked to this, another almost identical indicator about if the company reached the “Seed round” was made. Where a 1 illustrates that the firm did reach the seed round and a 0 if it did not.

The most common organization that provides financial aid in form of investments to new companies are the venture capitals, which are found across the country and are constantly making investments for a living expecting that someday that investment will pay off and earn some profit from it. Thus, to study the distribution and density of these venture capitals through the country, will help us in order to study the facility or difficulty a company X from the region Y has to obtain such help in terms of money.

So as to cover the previous concept, an indicator was made based of the number of venture capitals per CCAA per capita. The higher the value of the indicator, the more presence of these organizations in that region.

5.4. Summary of the indicators

The table 1 summarizes all the indicators used in the mathematical study.

Table 1: Indicators

Indicator	Calculus	Data source
Unemployment rate	Given	INE
Income level of population	Given	INE
Birth rates per 1000 people	Given	INE
PIB per capita	Given	INE
VC's indicator	Number of Venture capitals / population of that CCAA	Crunchbase
Educated labour force	Number of graduate students from degrees related to software companies / population of that CCAA	INE
Universities	Number of universities / population of that CCAA	INE
Funding rounds	Number of funding rounds in that CCAA / total companies in that CCAA	Crunchbase
Money raised	Money raised in that CCAA / total companies in that CCAA	Crunchbase
% of international investors	Number of interntional investors * 100 / Sum of national and international investors	Crunchbase
Innovation indicator	Given	European commission
Tax incentive indicator	Average tax payment for an individual in a CCAA / population of that CCAA. Then calculate what percentage of the total regions represents that CCAA.	Ministerio de Hacienda
Performance	Percentage increase in the money raised from one round to next round	Crunchbase
Seed round achievement	1= If the company has reached the seed round. 0 = Else	Crunchbase

ETSEIB

6. Empirical analysis: regression models

6.1. Objectives

Once all the indicators have been calculated as previously explained, a multiple regression analysis is used in order to extract some conclusions out of the data collected and fulfil the objectives of the study. To carry out the analysis, we used the software “Minitab” in its 2019 version.

Given that for each dimension several indicators were collected and we have a limited number of observations we decided to use a step-wise regression. With this strategy it is possible to figure out which are the most relevant variables to be included in the model.

To better scrutinize what makes software start-ups successful, we run two series of approaches, with 34 models in total. First, taking as the unit of analysis the region (the autonomous communities), and second, using the individual company as the main focus of analysis. Also, different independent variables have been used, as the performance of software start-ups can be measured using alternative metrics. In the sections below we report the procedure followed.

6.2. Methodology

So as to find the “perfect” equation, or the one that resembles the most to the ideal one, four different models have been studied, which will later be explained. For each of these models, different approaches, variables and values of the response “Y” have been used, which led to different equations and in some cases different significant variables.

In order to analyse each model, a methodology was needed. In Minitab there are many ways to make successfully a regression study. First of all, a histogram of all the variables was made to affirm that they are normalized. For some cases such as for the variables of “population”, “average money raised in that CCAA” and “Total funding”, they were divided by a constant to make them proportional to the other indicator’s variables. Because if not, in the regression equation they had a very low coefficient which is difficult to work with, hence to make it more practical this change was made. For “average money raised in that CCAA”, “Total funding” and “population” each row was divided by 10^6 (1 million).

The next step was to normalize the response for each model, which was done by making the neperian logarithm for the column “Y”.

After making these adjustments, several models were proposed. Given the high number of independent variables, in first attempt we used the option “best subsets”, which takes into account the best variables to make the study according to their p-value, the CP Mallows (the lowest one, the better) and the R^2 value. Once the best subsets are identified, a manual regression model could be made where in each step, some variables are eliminated by looking at the p-value and the number of alpha chosen. This process is very time consuming and does not guarantee the best solution since having so many variables, it takes a lot of steps and human errors could occur.

An alternative but most robust approach was to select variables using the “stepwise” method. By using this approach, the software is able to find the optimal equation which only retains those variables that have a p-value below the number specified by the researcher. More precisely, we used the “forward selection”. In this case the method starts with an empty model and the user indicates all the variables that might be worth to be considered in the final equation. Then, the software adds the most significant term for each step. Minitab stops when all the variables not in the model have p-values that are greater than the specified alpha value. [25]

In the study of each model a set of hypotheses were defined in order to apply and make use of the p-value and find the limit between a significant variable and one that is not.

Hypothesis 0 – A variable is not significant if p-value >= 0,2

Hypothesis 1 – A variable is significant if p-value < 0,2

If a variable does not verify H0 then it will verify H1. In the study the Alpha value=0.2, and it has been considered this way because the data and variables calculated did not give a minimum number of indicators as significant that it was considered as necessary to reach some conclusions in the research. Therefore, a higher probability of mistakes or less certainty in the results was chosen over getting more indicators as significant.

Hence, an Alpha value of 0.2 means that the results gotten can be affirmed to be correct with an 80% of certainty, which is considered enough for this exploratory study. Although, for each model, some of the variables that are significant will have p-values way lower than 0.2, but it only serves as a barrier to eliminate those indicators above said value.

- A p-value < 0.01 is very significant, with 99% of certainty
- 0.011<p-value<0.05 is significant with 95% of certainty
- 0.051<p-value<0.1 is significant at 90%

For each variable, the p-value will be analysed carefully.

Having an Alpha value of 0.2 does not guarantee a certainty of 100%, therefore there is a certain chance that some mistakes are made. First, the type 1 mistake, which consists of accepting the H1 when actually the correct one is H0. Secondly, the type 2 mistakes consist of accepting the H0 when the H1 is actually correct. [26]

6.3. Regression models

In order to find the most significant variables/indicators that help for the creation and success of a software start-up, four different models were studied were in each one the response (Y) is calculated in a different way. The difficult part here, is to find a response that represents the success of the company, hence, many different calculations were made to represent as ideally as possible the concept of “success”.

Moreover, two different approaches were taken into consideration based on the two excels that were made when the data was collected. The first approach is looking only at each CCAA and the other one is studying the company in that CCAA individually. One model corresponds to the first approach and the other three models correspond to the second approach.

The goal of studying these four models is to find a repetition in several models of the variables that are significant. If a variable appears in more than one of the models then it is highly likely that in our different representations of “success”, said variable is significant. If it only appears once then we will have to question whether it should be considered significant or not. Besides, each model has its flaws and neither of them is ideal, one of them may be “stronger” than another model and vice versa, therefore in the analyses of the results one model may have more weight than another because it has less flaws.

6.3.1. Model 1

The first model is the one that focuses on each one of the CCAA’s and not in each company individually. So for each CCAA, the information of all the companies of the region are gathered and put together so as to compare them with other CCAA’s.

The value of the response (Y) needs to represent or relate as much as possible to the level of success of the companies in that CCAA. In this model, this response is calculated as the average performance of all the software companies in that region. As a remainder, the level of performance of a company was calculated by the percentage increase in the money raised from one round to the next. Therefore, the response being the average performance of all the companies, gives a representation of the success and growth of the firms in that CCAA by considering that the increase in money raised is directly proportional to the success of that company.

Out of the seventeen CCAA's that Spain has, only 8 CCAA's had enough data about the money raised of its companies and the total funding rounds to consider them in the regression model. These eight CCAA'S were Andalucía, Asturias, Cataluña, Comunitat Valenciana, Galicia, Madrid, País Vasco and La Rioja. The other nine CCAA that were not included, all of them had a few software companies but the financial details of each of them was not publicly available. Hence, no study could be made considering those regions. In every model, only the previous eight regions are considered and studied.

The variables that have been used in this model are "Population", "Unemployment rate", "Income level of population", "birth rates per 1000 people", "PIB per capita", "Educated labour force", "Universities indicator", "Funding rounds", "Money raised of all the companies of the region", "Percentage of international investors average", "Total software companies", "Innovation indicator" and "Tax incentive indicator". All of which have been previously explained.

The following table 2 illustrates all the original variables used, before any transformation was made, in relation to their components and with some traits. Only considering those companies that we have financial data of, which will be part of the model.

Table 2: Descriptive statistics of the variables

Dimension	Indicator	Mean	St. Dev.	Max	Min
Cultural indicators	Innovation indicator	86,95	13,6	103,6	67,5
Legal indicators	Tax incentive indicator	0,1778	0,0291	0,2228	0,1383
Access to finance indicators	Funding rounds [number]	2,920	0,936	4	1,5
	Money raised of all the companies of the region [€]	2.071.500	1.897.700	5.790.000	80.000
	International investors [%]	28,35	16,6	53,85	0
Demographic indicators	Population [inhabitants]	4.285.798	3.152.445	8.472.407	319.796
	Birth rates per 100 people	6,799	1,084	7,68	4,7
	Educated labour force [number]	2,382	0,708	3,941	1,829
	Unemployment rate	11,84	3,77	20,18	8,43
	Income level of population [€]	12.956	1936	15.813	9.990
	Universities indicator [number]	3,933	1,288	6,852	2,926
	Total software companies in the region [number]	57,6	70,4	200	1
	PIB per capita	24.696	5.089	32.048	17.747

Before applying the methodology of stepwise, all the variables as well as the response were looked at so as to see if they had to be normalized or not. For instance, the response "Y" had to be normalized by doing the neperian logarithm. To know whether a variable or response is normalized, the Histogram was made and checked if the shape was similar to that of a normalized variable. Also, “Population” and “Money raised” were divided by a constant of 10^6 in order to be more practical with the respective coefficients in the equation.

Another step to be made before doing the regression model is to study if some of the variables are correlated and are generating some kind of noise. To do so, the Pearson correlation matrix was performed, confronting all variables against. Values range from -1 to 1. A negative value means that the variables are negatively correlated and if it is close to -1, then it is strongly correlated. Whereas if the value is greater than zero, then the correlation is positive and if, close to 1, the correlation is stronger. Values close to 0 denote no correlation effect.

The Pearson matrix of model 1 is the following, where we can see that the majority of the variables have low Pearson values. But since so many variables are being taken into account to have a wider vision of the environment, it is almost impossible to have no correlation at all between indicators. Hence, there are some cases where some values are close to -1 and 1. For instance between “Money raised” and “Total software companies” with a value of 0.932, or between “Unemployment rate” and “Income level of the population” with a value of -0.835. These both values are the highest values in the model. Said previous analysis will be taken into account in determining if the model is valid or not to extract conclusions.

The pearson correlation representation shows a visual of the correlation of each variable with another, said graphic can be found in image 4 in the Annex section. The next table 3 illustrates the values for the interactions between every variable.

Table 3. Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Population	1,000												
2. Unemployment rate	0,561	1,000											
3. Income level of population	-0,248	-0,835	1,000										
4. Birth rates per 1000 people	0,644	0,323	0,110	1,000									
5. PIB per capita	-0,026	-0,729	0,932	0,332	1,000								
6. Educated labour force	0,384	-0,242	0,502	0,469	0,711	1,000							
7. Universities indicator	-0,383	-0,377	0,449	-0,232	0,307	0,132	1,000						
8. Funding rounds	0,018	0,033	-0,423	-0,504	-0,393	0,004	-0,29	1,000					
9. Money raised	0,650	-0,098	0,097	0,331	0,271	0,339	-0,333	0,285	1,000				
10. % international investors	0,619	0,370	-0,098	0,07	-0,051	0,183	-0,039	-0,103	0,166	1,000			
11. Total software companies	0,705	-0,144	0,333	0,517	0,49	0,521	-0,339	0,038	0,906	0,296	1,000		
12. Innovation indicator	0,149	-0,592	0,794	0,380	0,861	0,73	0,461	-0,208	0,463	-0,093	0,576	1,000	
13. Tax incentive indicator	0,457	0,517	-0,508	0,14	-0,244	0,369	-0,459	0,404	0,073	0,385	0,043	-0,316	1,000

Once all the pre-analysis is done, the stepwise method can be applied. For this case, three steps were taken, which can be seen in image 5 in the Annex section. The first one to include “Innovation indicator” in the model, with a p-value of 0.001. The second step to include “Funding rounds”, with a p-value of 0.037. The third and last step to include “Population”, with a p-value of 0.182. As a remainder, the alpha value was 0.2, therefore variables with a p-value > 0.2 have not been considered.

Other values have to be looked at to reassure that the model is a valid one. The first one is the R^2 which the closer to 100% the better. This value is of 96,21%, which strengthens the model. Secondly the deviation (“S”) has to be studied, because the higher the value, the higher the likeliness of mistake there is. In this case S= 0.166.

Lastly the FIV (family importance value), which is a value that indicates how affected the variance of a coeficient caused by correlation between variables in the model [31]. A FIV value close to 1 indicates no correlation, 1<FIV<5 indicates a bit of correlation and a FIV value greater than 5 indicates a lot of correlation. In our study, the three indicators selected in the model have a FIV value of almost one, hence we can cosider there is no correlation between them. Image 6 in the Annex section shows all the FIV values, coeficients and p-values for each variable.

The most important part of the methodology is to figure out the equation that resembles the indicators with the response “Y”. Said equation is the following.

$$\ln(y) = -1{,}725 - 0{,}0326 \text{ population} - 0{,}2178 \text{ Funding rounds} + 0{,}04151 \text{ innovation indicator}$$

As it can be seen, “population” and “funding rounds” have a negative coefficient, which means that they are negatively impacting on the outcome. This result can be interpreted saying that the lower the number of funding rounds, the better the performance. The effect of population seems to dilute as the p-value is higher than 0.1, therefore, this effect is found to be not as significant. On the other hand, the “innovation indicator” is positive and significant, meaning that the greater its value the greater the dependent variable will be.

To sum up the model 1, the next table illustrates the main aspects.

Table 4: Main aspects of model 1

Significant variables	Conclusions	Flaws
- Population - Funding rounds - Innovation indicator	Less funding rounds is linked to higher performance and that this makes sense as it means that with less rounds they are able to raise higher amounts of money. In the case of innovation indicator, the effect is positive and significant, meaning that regions that are more advanced in this dimension are more likely to attract investors, and therefore, companies located there have greater chances to succeed.	A flaw is that having a response value “Y” as average performance of companies in a specific CCAA is not an ideal definition of success. Moreover, some CCAA’s have a lower number of companies than others, hence there is no total certainty that that number represents the whole region due to the fact that some companies do not have their financial information public.

6.3.2. Model 2

For the next three models including model 2, the approach made is taking into account every single software company (SaaS and ERP) individually instead of taking every CCAA as in model 1.

In this case, the response “Y” has been calculated as the division between performance and funding rounds. Dividing the performance by the number of funding rounds allows the distinction of the following situation. In the hypothetical situation where two companies have the same performance but one company has less funding rounds than the other, then it has been thought that the one with the less amount of funding rounds has been more successful, because it has achieved the same rate of performance requiring less financial rounds.

In the excel where all the data was collected from Crunchbase, there are 102 companies with the sufficient data that is needed.

For models 2,3 and 4, the same variables have been used but the response is different in each case. Those variables are “Population”, “Unemployment rate”, “Income level of population”, “birth rates per 1000 people”, “PIB per capita”, “Educated labour force”, “Universities indicator”, “Funding rounds average of CCAA”, “Money raised of all the companies of the region”, “Percentage of international investors average”, “Innovation indicator”, “Tax incentive indicator”, “Acquisition”, “Total fundings in that CCAA”, “Seed round achievement”, “VC indicator” and “Total software companies in that CCAA”.

The following table 5 illustrates all the original variables used, before any transformation was made, in relation to their components and with some traits. Only considering those companies that we have financial data of, which will be part of the model.

Table 5: Descriptive statistics of the variables

Dimension	Indicator	Mean	St. Dev.	Max	Min
Cultural indicators	Innovation indicator	95,745	9,065	103,6	67,5
Legal indicators	Tax incentive indicator	0,1778	0,0326	0,2228	0,12036
Access to finance indicators	Funding rounds for every company [number]	3,657	2,056	15	2
	Money raised of all the companies of the region [€]	4.332.304	1.743.600	5.790.000	319.000
	International investors [%]	33,176	8,42	53,846	0
	Acquisition [1;0]	0,0294	0,1698	1	0
	Seed round achievement [1;0]	0,3627	0,4832	1	0
	VC indicator [number]	10,554	4,609	16,441	1,534
	National investors [number]	2,735	2,044	11	0
Demographic indicators	Population [inhabitants]	6.819.897	1.801.075	8.472.407	661.537
	Birth rates per 100 people	7,3908	0,6494	7,68	4,7
	Educated labour force [number]	2,7048	0,7319	3,941	1,829
	Unemployment rate	10,908	2,353	20,18	8,43
	Income level of population [€]	13.751	1.289	15.813	9.990
	Universities indicator [number]	3,2969	0,7167	7,8029	2,926
	Total software companies in the region of that company[number]	147,61	67,97	200	2
	PIB per capita	27.498	3.927	32.048	17.747
	Total fundings in that CCAA [€]	5.637.091	19.259.743	187.300.00	50000

The response "Y" has been normalized using the neperian logarithm, as it can be seen in the following histograms (Figure 1).

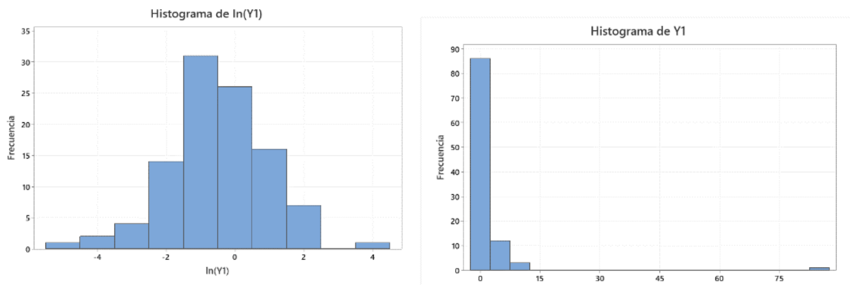

Figure 1: Histogram of response "Y"

Since models 2,3 and 4 have the same variables, just one Pearson correlation matrix is needed to study whether there is a strong correlation between indicators. The matrix is the following.

Table 6: Pearson correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Aquisition	1,000															
2. Total funding	0,042	1,000														
3. Seed round achievement	0,110	0,216	1,000													
4. Population	0,059	0,085	0,051	1,000												
5. Un-employment rate	-0,057	-0,057	-0,157	0,076	1,000											
6. Income level of population per capita	0,075	0,062	0,101	0,262	-0,861	1,000										
7. Birth rates per 1000 people	0,06	0,061	0,001	0,841	-0,004	0,438	1,000									
8. PIB per capita	0,077	0,041	0,068	0,281	-0,739	0,925	0,525	1,000								
9. VC indicator	0,081	0,042	0,069	0,423	-0,61	0,835	0,581	0,954	1,000							
10. Educated labour force indicator	0,043	-0,029	-0,038	0,046	-0,262	0,483	0,398	0,748	0,788	1,000						
11. Universities indicator	-0,051	-0,079	-0,112	-0,774	0,042	-0,132	-0,355	-0,114	-0,322	0,144	1,000					
12. Funding rounds	-0,026	-0,044	0,053	-0,549	-0,048	-0,286	-0,566	-0,18	-0,102	0,18	0,189	1,000				
13. Money raised	0,061	0,121	0,182	0,673	-0,492	0,506	0,464	0,334	0,336	-0,256	-0,657	-0,353	1,000			
14. % International investors	0,03	-0,004	-0,073	0,501	0,223	0,078	0,313	0,19	0,346	0,216	-0,553	-0,464	-0,031	1,000		
15. Total software companies in that CCAA of that company	0,076	0,117	0,156	0,771	-0,525	0,669	0,625	0,532	0,558	-0,03	-0,69	-0,45	0,95	0,169	1,000	
16. Tax incentive indicator	-0,004	-0,081	-0,11	-0,176	0,26	-0,158	-0,006	0,192	0,314	0,744	0,146	0,392	-0,664	0,37	-0,521	1,000

It is worth mentioning that there are some variables that have a high correlation with values close to 1 and -1. These interactions are caused by “Average money raised in that CCAA” with “Total software companies in that CCAA”, a value of 0,95. Moreover “PIB per capita” and “VC indicator” also present high correlation with a value of 0,954. Nevertheless, most variables present low values in comparison with other ones. It will later by studied if such correlation is also present in the significant variables.

When applying the stepwise methodology, only one step is needed to include the variable “Innovation indicator”, with a p-value of 0.102.

The value of R^2 for this model is pretty low, equivalent to 2,66 %. Whereas the deviation value “S” is 1.44.

The FIV value is 1 in this case since there can not be interaction with other variables.

There have also been detected a couple of observations that had a remainder, “Residuo” in Spanish, that was abnormal in comparison with all the other observations. Thus, they have not been considered.

The final equation with the significant variables is the following. We can see that “innovation indicator” has a direct relation with the response Y.

ln(Y1) = -3,04 + 0,0262 Innovation indicator

To sum up model 2, the next table summarizes the main aspects.

Table 7: Main aspects model 2

Significant variables	Conclusions	Flaws
- Innovation indicator	“Innovation indicator” can be assured to be significant with an 89,8% certainty and a positive correlation with the response.	One of the main flaws is that the R^2 is extremely low, which means that what is trying to be predicted (Y), is very difficult to do so with an equation of said characteristics. This model will not be considered.

6.3.3. Model 3

The approach taken is the same as model 2, which is basically focusing on every software company (SaaS and ERP) individually. In this case, the response “Y” has been calculated as the performance of each company divided by its total funding. By doing so, it is being considered as well all the fundings the company has been given. So for instance theoretically if two companies with the same performance ratio, one has obtained more funding than the other one, the one with less funding but same performance has been more successful because it has achieved the same level of performance with less financial resources.

The study has been made with 102 companies from 8 different CCAA’s, the same as in model 2. Therefore, the same variables have been used so the Pearson correlation matrix is the one mentioned in the previous model.

The response “Y” has been normalized using the neperian logarithm, as it can be seen below (Figure 2).

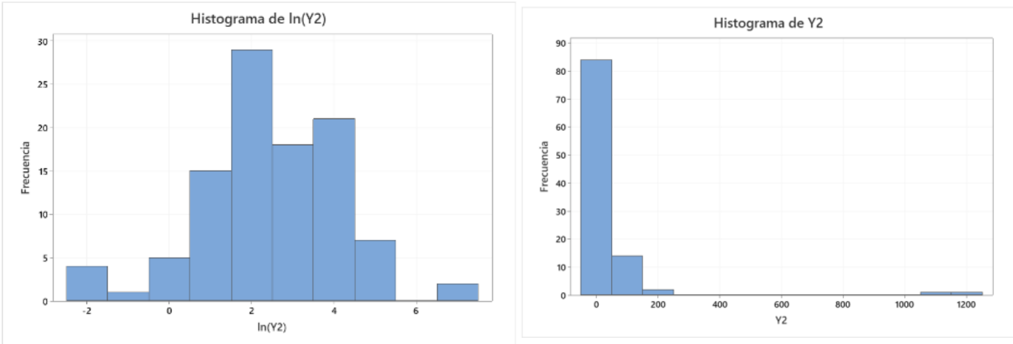

Figure 2: Histogram of the response

Applying the stepwise methodology, the first step is to include “Seed round achievement” with a p-value of 0. The second step is to include “International investors” with a p-value of 0.031. Then “national investors” is added with a p-value of 0.043. In step number four, “Universities indicator” is included with a p-value of 0,121. And lastly, “Innovation indicator” with a p-value of 0,127 is also added into the model. Image 7 in the Annex section shows the process of the stepwise methodology for model 3.

The value of R^2 is equal to 28,49%, the deviation is equal to 1,48 and all the FIV values are close to 1, meaning that there is almost no correlation between the variables, as it can be seen in image 8 in the Annex section.

There have also been detected a series of observations that have a “remainder” pretty big in comparison with the other observations as well as some market as “Less freqüent” observations. The observations with a big remainder have not been considered in the model.

Finally, the regression equation is the following.

ln(Y2) = -0,15 - 0,1591 International investors - 0,1790 National investors
- 1,077 Seed round achievement + 0,367 Universities indicator
+ 0,0258 Innovation indicator

It can be seen that five variables have been considered as significant. “International investors”, “National investors” and “Seed round achievement” have a negative coeficient, therefore they do not relate positively with the response.

Whereas “Universities indicator” and “Innovation indicator” have a positive relation with the response.

To sum up model 3, the next table summarizes the main aspects.

Table 8: Main aspects model 3

Significant variables	Conclusions	Flaws
- International investors - National investors - Seed round achievement - Universities indicator - Innovation indicator	The biggest p-value of all the significant variables is of 0,127. So it can be assured with a 87,3% certainty that these variables are significant in this model.	The value of R^2 is low which makes the response "Y" difficult to predict through an equation of this kind. The value of R^2 is low but it is consistent with the typical value obtained in studies in the field of social sciences. Therefore, despite being low, the model is robust enough. The response which symbolizes success, does not take into account the number of funding rounds needed for each company, so it is not an absolute representation of success.

6.3.4. Model 4

The approach taken is the same as model 2, which is basically focusing on every software company (SaaS and ERP) individually. In this case, the response "Y" is calculated as total funding divided by founding rounds of the company. It is the only model where the performance indicator is not considered. This alternative way of representing the "success" of a firm is based on the idea that the greater the fundings and the lower the number of funding rounds, the better the company has performed over time.

The number of companies is 102 and the variables taken into account is the same as in models 2 and 3. Therefore, there is no need in studying again the Pearson correlation matrix to see if there is strong interaction between variables.

Again, the response "Y" has been normalized using the neperian logarithm. As it has been illustrated below in Figure 3.

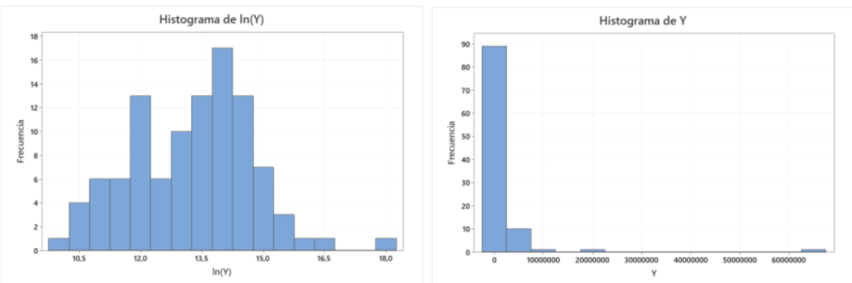

Figure 3: Histogram of the response

When applying the stepwise methodology, the first step is to include the variable that suposedly has the lowest p-value, which is “Seed round achievement” with a p-value of 0. Then, in the second step the variable “International investors” is added with a p-value of 0. Thirdly, “National investors” is added with a p-value of 0.028. In the last step, with a p-value of 0.037, “Birth rates per 1000 people” is also considered to be significant in the model. All the steps can be seen in image 9 in the Annex section.

The value of R^2 is 44,16% with a deviation of 1.125. The FIV values are all around 1, hence it can be affirmed that there is almost no interaction between variables.

Once all variables are put together in the same model, the respective p-values are the ones shown in the image 10 in the Annex section.

Those observations that have a high remainder value and that Minitab selects, have been cut off the model.

Lastly, the regression equation is the following.

$$\ln(Y) = 9{,}46 + 0{,}2135 \text{ International investors} + 0{,}1248 \text{ National investors} + 1{,}220 \text{ Seed round achievement} + 0{,}367 \text{ Birth rates per 1000 people}$$

As it can be seen, all the variables have a positive coeficient, meaning that they are positively and direclty related with the response “Y”. So the higher the value of these indicators, the higher the response will be.

The next table 9 summarizes the main aspects of model 4.

Table 9: Main aspects of model 4

Significant variables	Conclusions	Flaws
- International investors - National investors - Seed round achievement - Birth rates per 1000 people	Out of all the variables in the final model, the biggest p-value is 0.037, which means that it can be affirmed with a 97,3% of certainty that those values are significant. The sign of the coeficient will later be evaluated.	The main flaw of this model is that by not considering the performance ratio in the response “Y”, it is not being taken into account the growth between the last two funding rounds. So no study of the progression of the performance of the company is being made, only looking at the bigger picture.

7. Results and comparison of the models

Out of the four models that have been run in the study, the second one has not been taken into consideration because the value of R2 is below 25%. The other models have a value above 25% which is an acceptable threshold for studies in this area, where the response "Y" is difficult to represent.

Concerning the remaining models (1, 3 and 4), the variables that are statistically significant are consistent in all of them, strengthening the robustness of the results. Note that the dependent variable (performance) has been operationalised differently in each model. Obtaining results that hold true for the three models further validates the relevance of explanatory variables that are statistically significant regardless the measure of the outcome employed.

Diving deeper into the results, we found that the presence of both international and national investors as well as seed round achievement are positively linked with a higher performance in model 4, but have a negative effect in model 3. Therefore, according to the study no conclusion can be drawn of whether the relation is positive or negative due to the fact that two of the models contradict each other. Nevertheless, what can be affirmed is that the three variables can be said to be significant and have weight in the success and creation of software companies.

For model 3 where the response is performance dived by funding rounds, the effect is negative, which means that a higher value of these variables, a lower value of the response. Whereas in model 4 is the other way around. The common trait is that for both models the three variables have the same sign, but the difference is that for each model the definition of success is different which could alter the sign of the variables if a condition or situation considered in model 3 is not considered in model 4.

Most models have in common the presence of the "Innovation indicator" appears in both model 1 and 3, having a positive relation with the response. Hence, it can be affirmed that if the region where the start-up is located is characterized for being highly innovative, the likelihood of the company performing better is higher.

Then, several indicators show up in the different models separately without repetition, these are "Population", "Funding rounds", "Universities indicator" and "Birth rates per 1000 people". All of them have a positive correlation except "Funding rounds" and "Population". This result goes against the previous thoughts that the more funding rounds the better, instead, the lower funding rounds a company has the better. This result can be attributable by assuming that a successful company does not need a lot of funding rounds in a short period of time to be able to perform, it does not need constant external financial support. Whereas the positive relation indicators confirm the firsts hypothesis that the higher the level of talent/universities the better, as well as of birth rates. Thus the higher the birth rates, the more demand there will be. In the case of “Population” it kind of contradicts the birth rate indicator, but having the “population” a negative indicator it may mean that the bigger the market the more difficult it is to succeed because there will probably already be companies competing for the same market share.

The following table summarizes the main results (taking into account the three models). It only includes those variables that at least in one of the models was found to be statistically significant.

Table 10: Variable relationship with response

Indicator	Relation with response “Y”
Population	Negative
Funding rounds	Negative
Innovation indicator	Positive
International investors	Not defined
National investors	Not defined
Seed round achievement	Not defined
Universities indicator	Positive
Birth rate per 1000 people	Positive

8. Qualitative research

In order to confront the results obtained in the empirical analysis, in a second stage analysis a qualitative study was designed. The ultimate purpose was to have the opportunity to hold a conversation with an expert in the field, share our findings with this person and exchange points of view in order to come up with a final list of recommendations for start-ups in the software industry.

Due to the tight schedule, we were only able to interview one expert. In this case, we selected Josep M. Piqué Huerta, who has a strong profile as practitioner but also has a background in research, holding a PhD in Innovation Ecosystems from La Salle URL. Currently he is the president of La Salle Technova Barcelona, the research park at La Salle. Among the most renowned positions it is worth mentioning that currently, he is the president of the Triple Helix Institute, the president of XPCAT (the network of Catalan Science Parks), the vice-president of APTE (Spanish association of science parks) and the past president of IASP (international association of science parks). His strong CV makes him ideal for this research.

An online interview was scheduled on June 22, 2022. In the paragraphs that follow we include the questions and a summary of his responses.

What external factors are the most important ones to consider when it comes to decide where to locate for creating a software company?

Josep mentioned that for him there are five aspects to look at. These aspects are:

- The level of “talent” that region has in relation to the activities done in the company.
- The level of infrastructure in the location so as to perform the tasks correctly and efficiently.
- That the level of demand is high enough to have a profit margin.
- The possibility of having individuals or entities financing the company.
- A high level of networking in the environment.

What weight would you give to each of those factors?

All these factors have a similar weight and importance. The next question gets into more detail about this issue.

Do you think that a place should have a couple of excellent factors or, on the contrary, is the combination of all of them what enhances the creation and improves the performance of a company?

According to Josep, it is the combination of all the factors what creates a better ecosystem for the company. Moreover, he emphasises the importance of the quality of life and personal life of the employees. This is critical to make them perform as better as possible. Also, it is a strategy for attracting and retaining talent.

Do you think that the above factors are the same over the lifecycle of a company?

He states that some factors are more important at some earlier stages than others. In the creation stage, the most important factors refer to the talent and the technology that company is selling and is able to develop. On the contrary, during the development stage, the key aspects are the easiness to get funding, the demand and the infrastructure.

What relations do you consider to be the most important ones in the Triple Helix model between the university-industry-government? What are the actions that benefit the most the three agents involved?

As a quick summary of the complex relationship between these three agents, Josep mentioned that the universities are key for the creation of talent, the creation for new technologies and for the research they conduct. Whereas for instance, the government is important because it acts as a purchase mechanism, it buys some of the products the companies make and helps them (with policies) to put them in the marketplace. Lastly, the industry is important because a part from the aspects mentioned in the theoretical part of the project, the industry “buys” those companies in order to get the most out of them.

Is there any other model able to explain why some companies are more successful that others due to regional differences?

At this point, Josep mentioned the theory of the “Clusters of Innovation”. This theory bases its premises in the named “hot spots” where new technologies are created at an astonishing rate thanks to pools of capital, new approaches of doing businesses, creation of new industries and pools of capital and knowledge. The key elements (enablers) are illustrated in Figure 4.

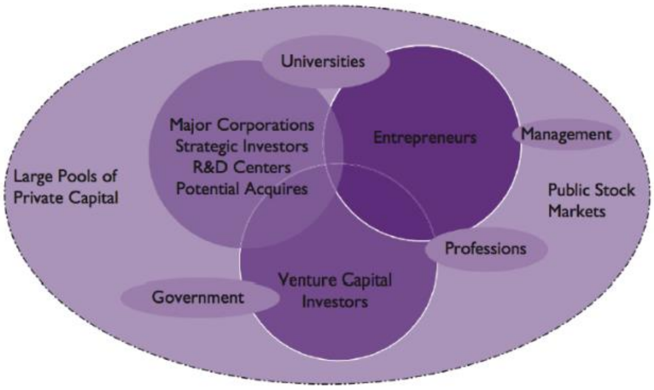

Figure 4: The Innovation Engine of Clusters of Innovation. Source[32]

9. Discussion and concluding remarks

9.1. What have we learnt?

The study has identified those indicators that are more significant out of all the ones that were calculated of the data collected, for the creation and success of the software start-up companies.

Although for the quantitative results the conclusions are based on the models of Triple Helix and PESTLE analysis and for the qualitative results obtained in the interview with Josep Piqué he focuses on 5 factors and mentions the theory of Cluster of Innovation that has not been used in this project, a comparison can be made.

The quantitative results show a similar conclusion to the one mentioned by professor Piqué, where some significant variables relate to some important factors the professor mentions. These have been summarized in the following table 11 so as to illustrate easier the comparison.

Table 11: Relation between quantitative and qualitative research

Quantitative indicator	Qualitative factor
Funding rounds International investors National investors Seed round achievement ⇨	Finance
Population Birth rates per 1000 people ⇨	Demand/market
Universities indicator ⇨	Talent
Innovation indicator ⇨	Level of innovation in the region

Therefore, although different models have been used, the conclusions are similar and it can be said that the most important aspects in a region are the finance, the demand or market, the talent it can be found there and the level of innovation the region has (for example if there is a high concentration of entrepreneurs).

9.2. Avenues for future research

More into depth research should be made to figure out the type of relation the variables that have not been defined have with the response, by applying different types of models or approaches where the difference were identified and verified.

In order to make the study more convincing, a more precise definition of success and performance of a company should be defined taking into account several traits and extracting more up to date information.

If the conclusions would want to be extrapolated to the technological start-ups as a whole, then a big sample of data regarding companies of all types of sectors related with technology should be gathered and then analysed.

This thesis is expected to be useful for companies in the software industry. Particularly, we shed new light on the elements of the territory that should be taken into account when selecting a location for establishing a start-up. Also this study can be useful for policy-makers as it signals which territories are more fertile for start-ups, and which territories need additional incentives and help in order to attract talent, investors and clients. Finally, the academic community can also benefit from this study as it offers an insight about alternative typical indicators that could be of use for further research about this or related topics.

10. Budget

The budget of this thesis has been almost none. On the one hand all the data downloaded from INE and Crunchbase, was done for free although there was the possibility of paying a fee in order to get the data directly without having to write it down manually. On the other hand, the software used to apply the regression methodology, Minitab version of 2019, is completely free if the license of the UPC is used, which was the case.

The existence of a slight budget is due to the cost of the energy used to charge the computer as well as the lighting in the room, the cost of some pieces of paper and pens, but other than that no extra money was spent. Thanks to the possibility of having meetings online, no cost in fuel for transport was spent in order to meet my tutor, Jasmina, neither to make the interview to Josep Miguel Piqué.

11. Environmental impact

The environmental impact of the project is very low, because all the work has been done remotely from a computer. Therefore, the impact is related to the material used for the construction of the computer, the electricity needed to be able to work with it as well as the software's used to do online meetings and analyse the data.

First of all the use of the computer is not strictly for this project. Hence, it will be used for further research and for personal use, so the useful life of the computer will be optimized as much as possible, and some components will be recycled.

Regarding the energy in shape of electricity, the computer has been needed to charge many times throughout the time of the project, approximately 30 time of 2 hours each, representing a total of 60 hours of battery charge. It is known that the excessive use of energy has a negative impact for the environment because on the one hand it leads to a lack of those energies that are not renewable. On the other hand, it pollutes the atmosphere to a certain extent. [32] A solution for the future would be to make use of a more efficient and up to date computer that did not consume as much as the one used, so then less electrical energy it will need to perform the same functions.

Nevertheless, there has not been use of a transport for meeting with Jasmina Berbegal, the tutor of the project, neither with Josep Piqué to make him an interview. Such activities were done through the software "Zoom" and "Microsoft teams". Thus, this generates a positive impact by reducing as much as possible the pollution that comes from transportation.

Lastly, all the data gathered has been analysed with "Minitab" version 2019, so no other software's were used nor physical paper for the treatment of the information.

Acknowledgements

First of all, I want to thank my tutor Jasmina Berbegal Mirabent for her great advice and guidance throughout the development of the thesis, her consistency and implication whenever I needed her help and opinion.

Secondly, I would like to thank Josep Miguel Piqué Huerta, president of the Triple Helix Institute and director of the Technova, the science park at La Salle, for his time and consideration to accept answering some questions and shedding some light into the project.

Lastly, I want to express my gratitude to the start-up software company where I did my internship, Holded, where the inspiration about this TFG came to me. Also, their willingness to help and their knowledge whenever I had questions about the early days of their creation and the difficulties they had at the beginning.

Bibliography

Bibliographic references

[1] Statista, *Software Industry Insights,* Published 2020, [https://www.statista.com/markets/418/topic/484/software/#overview, 15 May 2022]

[2] IBISWorld, *Software Development in Spain,* Published 19 January 2021, [https://www.ibisworld.com/spain/industry/software-development/3595/, 9 May 2022]

[3] CoveyCS, *How to Do PESTLE Analysis For Software Projects,* Published 3 April 2019, [https://medium.com/@coveycs/how-to-do-pestle-analysis-for-software-projects-c2841aaec684, 16 May 2022]

[4] Fern Fort University, *Tableau Software, Inc. PESTEL & Environment Analysis*, [http://fernfortuniversity.com/term-papers/pestel/nyse4/3479-tableau-software--inc-.php, 15 May 2022]

[5] Casesteam, *Global Software Industry In Case Study Help,* [https://casesteam.com/chicago-booth/global-software-industry-in/pestel-analysis.php, 16 May 2022]

[6] Stanford University Triple Helix Research Group, *The Triple Helix Concept,* Published 11 July 2011, downloaded on 17 May 2022

[7] Gordon Scott, Patrice Williams, *Porter's 5 Forces*, [https://www.investopedia.com/terms/p/porter.asp, 7 March of 2022]

[8] Siddharth Vedula, Markus Fitza (2019) Regional Recipes: *A configurational Analysis of the Regional Entrepreneurial Ecosystem for U.S Venture Capital-Backed Startups*. Strategy science, published on the 26/02/2019 , downloaded on 15/03/2022

[9] Pimco, *Understanding inflation,* [https://europe.pimco.com/en-eu/resources/education/understanding-inflation, 7 March 2022]

[10] Thomas Brock, Pete Rathburn, *Per Capita GDP, Published 30 January 2022,* [https://www.investopedia.com/terms/p/per-capita-gdp.asp, 15 March 2022]

[11] Andy Schmitz, Saylor academy, *The economic importance of currency markets*, Published on 2012, [https://saylordotorg.github.io/text_money-and-banking-v2.0/s21-01-the-economic-importance-of-cur.html#:~:text=Key%20Takeaways&text=It%20is%20important%20because%20the,the%20people%20residing%20in%20it, 20 March 2022]

[12] Suman Pandit, *Types and Component of Technological Environment, Published 16 January 2020,* [https://businessmantalk.com/technological-environment/, 2 April 2022]

[13] Adam Hayes, Chip Stapleton, Diane Costagliola, *Patent Troll,* Published 9 april 2021, [https://www.investopedia.com/terms/p/patent-troll.asp#:~:text=A%20patent%20troll%20is%20a,court%20system%20to%20earn%20money, 2 April 2022]

[14] Samiksha S, *5 Major Components of Business Environment*, [https://www.yourarticlelibrary.com/business-environment/5-major-components-of-business-environment-business-studies/8638, 2 April 2022]

[15] Robert W. Wood, *California Proposes Tax increases again with Wealth Tax,* Published 7 March 2022, [https://www.forbes.com/sites/robertwood/2022/03/07/california-proposes-tax-increases-again-with-wealth-tax/?sh=656152002719, 11 April 2022]

[16] Hiring Monster Webpage, *How to Comply With Hiring Laws and Reduce Legal Risk,* [https://hiring.monster.com/resources/recruiting-strategies/talent-acquisition/legal-hiring-process/, 11 April 2022]

[17] Armen A. Alchian, *Property Rights,* [https://www.econlib.org/library/Enc/PropertyRights.html#:~:text=The%20fundamental%20purpose%20of%20property,with%20competition%20by%20peaceful%20means, 14 April 2022]

[18] Study Smarter, *Government Intervention In Markets,* [https://www.studysmarter.de/en/explanations/economics/macroeconomics/government-intervention-in-markets/#:~:text=The%20types%20of%20government%20intervention,demerit%20goods%2C%20and%20environmental%20protection, 14 April 2022]

[19] Dr.M.K.Singh, *Demographic profile of India*, [http://www.rncollegehajipur.in/rn/uploads/products/Demographic%20profile%20in%20business%20environment.pdf, 28 April 2022]

[20] Adam Hayes, Charles Potters, Katharine Beer, *Demographics,* Published November 7 2021, [https://www.investopedia.com/terms/d/demographics.asp#:~:text=Segmenting%20a%20population%20by%20using,that%20company's%20most%20influential%20consumers, 28 April 2022]

[21] European Commission, *Regional Innovation Scoreboard 2021 – Databases,* Published 29 June 2021, [https://ec.europa.eu/docsroom/documents/46031, 4 May 2022]

[22] Ministerio de Hacienda y Función Pública, *Tributación autonómica 2021,*

[https://www.hacienda.gob.es/es-ES/Areas%20Tematicas/Financiacion%20Autonomica/Paginas/Tributacion-Autonomica-2021.aspx, 4 May 2022]

[23] Ben Mcclure, Suzanne Kvilhaug, *How Venture Capitalists Make Investment Choices,* [https://www.investopedia.com/articles/financial-theory/11/how-venture-capitalists-make-investment-choices.asp, 8 May 2022]

[24] Minitab 18 support, *Perform stepwise regression for Fit Regression Model,* [https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/how-to/fit-regression-model/perform-the-analysis/perform-stepwise-regression/, 23 May 2022]

[25] María Vicenta Cabalgante Perera, *Pruebas de Hipótesis,* Published on 2003, [http://recursostic.educacion.es/descartes/web/materiales_didacticos/Muestreo_Inferencia_Estadistica/pruebas_hipotesis.html#:~:text=H0%20se%20llama%20hip%C3%B3tesis%20nula,signos%20distinto%2C%20mayor%20y%20menor, 23 May 2022]

[26] VIVIENDAsaludable, *Consecuencias del consumo de energía,* [https://www.viviendasaludable.es/sostenibilidad-medio-ambiente/ahorro-energetico/consecuencias-del-consumo-de-energia, 24 May 2022

[27] Josep M. Pique, Jasmina Berbegal- Mirabent, Henry Etzkowitz, *Triple Helix and the evolution of ecosystems of innovation: the case of Silicon Valley,* published on 2018, downloaded on 15/03/2022.

[28] Jerome S.Engel, Jasmina Berbegal-Mirabent, Josep M. Pique, *The renaissance of the city as a cluster of innovation*, published on the 08/10/2018, downloaded on 15/03/2022.

[29] Yuzhuo Cai, Marcelo Amaral, *The Triple Helix Model and the Future of Innovation: A reflection on the Triple Helix Research Agenda*, Published 25 November 2021, [https://brill.com/view/journals/thj/8/2/article-p217_1.xml?language=en, 17 May 2022]

[30] Loet Leydesdorff, Henry Etzkowitz, *The Triple Helix as a Model for Innovation Studies,* [https://www.leydesdorff.net/th2/spp.htm, 17 May 2022]

[31] Minitab 18 support, *Tabla de Coeficientes para Analizar diseño de mezcla,* [https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/doe/how-to/mixtures/analyze-mixture-design/interpret-the-results/all-statistics-and-graphs/coefficients-table/#vif, 23 May 2022]

[32] Monica Botey, Josep Miquel Piqué, Francesc Miralles, *The Evolution of Silicon Valley's Innovation Ecosystem: From 2006 to 2016,* Downloaded on 16 June 2022.

Annex

Company name	CCAA	Total fundi	Acquisition	Funding rour	Pre Seed rc	Convertible note (number of investors; money	Angel roun	Venture rou	Seed rounc	Series A	Series B	Series C	Series D	Performanc	Number of Investc	National investo	International Investc	Seed round achievement
Holded	Cataluña	2,66E+07	Yes	4			2,75E+05		1,60E+06	6,00E+06	1,50E+07			3,75E+00	7	5	2	1
Amenitiz	Cataluña	7,30E+06	No	3	6,20E+05	2,00E+05			6,50E+06					3,25E+01	12	3	9	1
Nextail	Madrid	1,20E+07	No	3				4,00E+05	1,60E+06	1,00E+07				6,25E+00	4	2	2	1
Lanbot	Cataluña	1,02E+07	No	2					2,20E+06	8,00E+06				3,64E+00	6	6	0	1
Haddock	Cataluña	2,10E+06	No	3	1,00E+05			1,00E+06	1,00E+06					1,00E+00	5	3	2	1
Whisbi	Cataluña	1,74E+07	No	7	1,00E+06	1,00E+06	4,40E+06	1,00E+06	1,00E+06	4,00E+06	5,00E+06			1,25E+00	5	1	4	1
iCommunity Labs	Madrid	1,80E+06	No	9	2,00E+04	8,50E+04	1,20E+05	2,30E+05	7,50E+04	3,00E+05	1,00E+06			3,33E+00	11	11	0	1
Fracttal	Madrid	7,40E+06	No	4					1,40E+05	2,00E+06	5,30E+06			2,65E+00	4	2	2	1
Smart Protection	Madrid	1,93E+07	No	6	5,00E+04			8,80E+05	7,60E+05	5,20E+06	1,00E+07			1,92E+00	9	6	3	1
Lodgify	Cataluña	6,60E+06	No	4					1,40E+06	5,00E+06				3,57E+00	9	3	6	1
TROOP	Madrid	8,90E+06	No	6	2,30E+05	7,00E+04	5,00E+05	5,00E+04		8,00E+06				1,00E-01	15	5	10	0
Twenix	Madrid	3,80E+06	No	5	1,00E+05	3,35E+05	1,00E+06	3,35E+05	2,00E+06					5,97E+00	6	4	2	1
ForceManager	Cataluña	1,65E+07	No	8	4,80E+05	6,60E+05	5,50E+05	2,60E+06	1,20E+07					4,62E+00	10	6	4	1
Pulpo	Madrid	7,80E+06	No	5	8,45E+05	1,00E+06	1,50E+05	5,00E+06						3,33E+01	5	3	2	0
Bloobirds	Cataluña	3,60E+06	No	2					3,30E+05	3,00E+06				9,09E+00	5	4	1	1
Geoblink	Madrid	7,30E+06	No	4	2,15E+05	1,53E+05	1,10E+06	5,00E+06						4,55E+00	5	5	0	0
Caravelo	Cataluña	2,70E+06	No	2					7,00E+05	2,00E+06				2,86E+00	3	2	1	1
Sales Layer	Comunidac	5,40E+06	No	4				1,23E+05	1,20E+06	3,50E+06				2,92E+00	7	4	3	1
Rentals United	Cataluña	6,50E+06	No	3				5,75E+05	1,50E+06	4,30E+06				2,61E+00	2	0	2	1
Citibeats	Cataluña	3,50E+06	No	2				1,50E+06	2,00E+06					1,33E+00	9	6	3	1
Blueliv	Cataluña	8,40E+06	No	3			2,50E+06	1,90E+06	4,00E+06					2,11E+00	6	6	0	1
Pridatect	Cataluña	3,70E+06	No	3				1,00E+06	2,20E+06	5,00E+05				2,20E+00	5	5	0	1
Neo	Cataluña	5,00E+06	No	1				5,00E+06							2	1	1	0
Councilbox	Galicia	5,90E+06	No	2				9,00E+05	5,00E+06					5,56E+00	4	4	0	1
VONZU	Cataluña	1,70E+06	No	3			4,00E+04	1,72E+05	1,50E+06					8,72E+00	5	1	4	1
Validated ID	Cataluña	2,40E+06	No	3			3,50E+05	2,00E+06	3,00E+06					1,50E+00	4	2	2	1
Billin	Madrid	3,80E+06	No	3			1,00E+06	1,50E+06	1,30E+06					8,67E-01	4	4	0	1
Decclarando	Comunidac	2,70E+06	No	2			5,25E+05	2,20E+06						4,19E+00	4	4	0	0

Image 1: Companies with their financial data

CCAA	Number of co	Total software companies in that CCAA
Andalucía	7	36
Aragón	0	11
Asturias	3	6
Baleares	0	4
Canarias	0	7
Cantabria	0	4
Castilla y Leór	0	4
Castilla- La M	0	2
Cataluña	76	200
Comunitat Va	9	42
Extremadura	1	3
Galicia	5	21
Madrid	35	131
Murcía	0	8
Navarra	1	2
País Vasco	2	24
La Rioja	0	0
	139	505

Image 2: Number of companies with financial data

CCAA	Population	Unemployment ra	Income level of popul	Birth rates per 1000 peopl	PIB per capita (euros	VC's indicato	Educated labour force indic	Universities	Funding rounc	Money raisec	International investors out of all the investor	Number of companies we have data of money	Total software companies in that C	Innovation indicator	Tax incentive indicat
Andalucía	8472407	20,18	9990	7,65	17747	1,534392765	1,949227028	2,975374	16	8,03E+06	0,538461538	7	36	67,5	0,208487949
Aragón	1326261	9,02	13097	6,79	26512	6,031995211	2,215959891	4,5269865			1	0	11	80,9	0,168959234
Asturias	1011792	10,01	12786	4,7	21149	3,953381723	1,858242383	3,7673439	12	1,53E+06	0,33333333	3	6	73,7	0,173730751
Baleares	1173008	14,9	12658	7,72	22048	1,70501821	0,790254495	2,7187655				0	4	67,4	0,171399967
Canarias	2172944	18,94	9935	5,86	17448	1,380615423	1,512722021	4,2759102			0	0	7	48,8	0,191107582
Cantabria	584507	11,52	12748	5,81	22096	1,710843497	2,04363839	6,7951401				0	4	73,5	0,154110083
Castilla y L	2383139	10,33	12697	5,7	23167	1,258843903	3,946223931	2,405012				0	4	76,9	0,1829049
Castilla- La	2049562	13,31	10485	7,13	19369	1,463727372	0,873179253	6,735267				0	2	64,4	0,176441388
Cataluña	7763362	10,16	14170	7,58	27812	10,82005451	2,354199808	2,9259587	2,13E+02	4,40E+08	0,328185328	76	200	98,9	0,148620729
Comunitat V	5058138	14,39	11332	7,07	20792	3,163219351	2,580500012	4,3957414	36	2,20E+07	0,114285714	9	42	91,3	0,181788216
Extremadu	1059501	18,94	9147	6,94	18301	0,943840544	1,278609571	4,5469757	1	1,00E+04		1	3	61,1	0,172590381
Galicia	2695645	11,01	11469	5,64	21903	2,225812375	1,828867881	4,0018991	18	1,38E+07	0,238095238	5	21	78,9	0,189699067
Madrid	6751251	10,12	14580	7,68	32048	16,44139731	3,941317914	3,4085064	1,11E+02	1,13E+08	0,382352941	35	131	101	0,222817129
Murcía	1518486	12,92	9850	9,06	19838	1,317101376	2,010474374	2,7269914				0	8	76,3	0,191981002
Navarra	661537	9,92	15094	7,67	29314	4,53489374	2,946363083	7,8028683	2	3,19E+05	0	1	2	98,1	0,120357664
País Vasco	2213993	8,43	15813	6,74	30401	5,420071337	2,625874549	6,852491	3	1,20E+06	0,333333333	2	24	103,6	0,138272123
La Rioja	319796	10,45	13504	7,33	25714	3,126993458	1,921618046	3,1347766	0	0	0	0	0	80,7	0,158587994

Image 3: Each CCAA with indicators

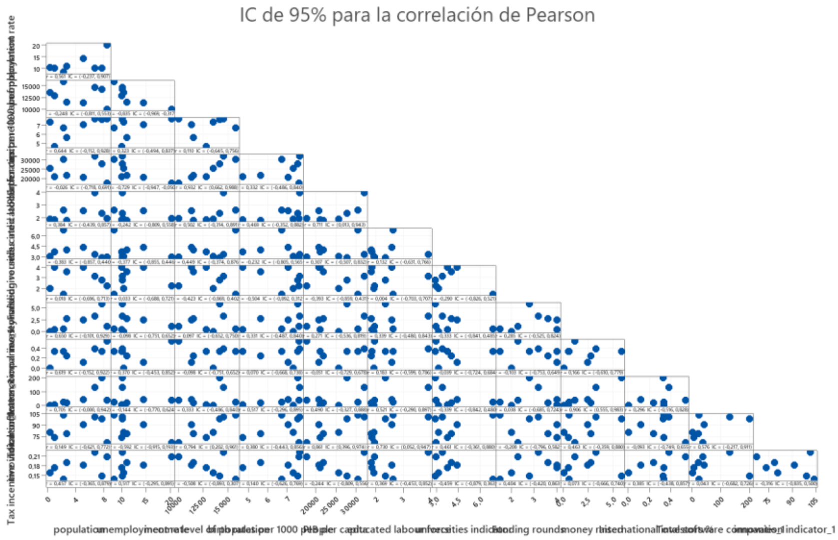

Image 4: Pearson correlation visual representation

	------Paso 1-----		------Paso 2-----		------Paso 3-----	
	Coef	P	Coef	P	Coef	P
Constante	-2,674		-1,744		-1,725	
innovation indicator_1	0,04351	0,001	0,04031	0,001	0,04151	0,001
Funding rounds			-0,2234	0,037	-0,2178	0,034
population					-0,0326	0,182

Image 5: Stepwise methodology model 1

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	-1,725	0,497	-3,47	0,026	
population	-0,0326	0,0202	-1,61	0,182	1,03
Funding rounds	-0,2178	0,0688	-3,17	0,034	1,05
innovation indicator_1	0,04151	0,00479	8,67	0,001	1,07

Image 6: FIV values model 1

	-----Paso 1----		-----Paso 2-----		-----Paso 3-----	
	Coef	P	Coef	P	Coef	P
Constante	2,981		3,137		3,484	
Seed round achievement	-1,495	0,000	-1,306	0,000	-1,112	0,001
International investors			-0,1659	0,031	-0,1541	0,042
National investors					-0,1583	0,043
Universities indicator						
Innovation indicator						
S		1,55777		1,52892		1,50479
R-cuad.		17,85%		21,65%		24,87%
R-cuad.(ajustado)		17,02%		20,07%		22,57%
AICc		384,11		381,44		379,37
BIC		391,74		391,53		391,87

	-----Paso 4-----		-----Paso 5-----	
	Coef	P	Coef	P
Constante	2,391		-0,15	
Seed round achievement	-1,058	0,002	-1,077	0,002
International investors	-0,1460	0,052	-0,1591	0,035
National investors	-0,1652	0,034	-0,1790	0,022
Universities indicator	0,328	0,121	0,367	0,083
Innovation indicator			0,0258	0,127

Image 7: Stepwise methodology for model 3

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	-0,15	1,81	-0,08	0,935	
International investors	-0,1591	0,0743	-2,14	0,035	1,10
National investors	-0,1790	0,0768	-2,33	0,022	1,13
Seed round achievement	-1,077	0,332	-3,24	0,002	1,18
Universities indicator	0,367	0,210	1,75	0,083	1,04
Innovation indicator	0,0258	0,0168	1,54	0,127	1,06

Image 8: FIV values model 3

	-----Paso 1----		-----Paso 2----		-----Paso 3----	
	Coef	P	Coef	P	Coef	P
Constante	12,657		12,437		12,149	
Seed round achievement	1,627	0,000	1,360	0,000	1,199	0,000
International investors			0,2344	0,000	0,2246	0,000
National investors					0,1314	0,028
Birth rates per 1000 people						
S		1,25496		1,16795		1,14508
R-cuad.		28,39%		38,60%		41,58%
R-cuad.(ajustado)		27,68%		37,36%		39,79%
AICc		340,02		326,50		323,64
BIC		347,65		336,59		336,14

	-----Paso 4----	
	Coef	P
Constante	9,46	
Seed round achievement	1,220	0,000
International investors	0,2135	0,000
National investors	0,1248	0,033
Birth rates per 1000 people	0,367	0,037
S		1,12527
R-cuad.		44,16%
R-cuad.(ajustado)		41,85%
AICc		321,30
BIC		336,16

Image 9: Stepwise methodology for model 4

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	9,46	1,29	7,36	0,000	
International investors	0,2135	0,0561	3,81	0,000	1,09
National investors	0,1248	0,0579	2,16	0,033	1,12
Seed round achievement	1,220	0,251	4,87	0,000	1,17
Birth rates per 1000 people	0,367	0,173	2,12	0,037	1,01

Image 10: FIV values model 4