

Master Thesis

Viabile Framework 1.0, evaluating viability of startups by using a data based framework

Jaime Parodi Bardón

Supervised by Dr. Miguel Godinho de Matos

15th October, 2017

Content

1. Abstract
2. Statement of the research
3. Relevance of the research
4. Literature review
5. Data sources and methods used to collect data
6. Data treatment and analysis
7. Presentation of conclusions
8. References
9. Appendices

Abstract

Most individuals have at least one business idea that they believe might be great. At the same time governments, academics and the media are also encouraging these individuals to come forth and become entrepreneurs. There is little doubt today that the entrepreneurial sector is critical for the future development and sustainability of the economy (representing 33% of global GDP and 45% of total employment). However, not all the ideas from entrepreneurs are viable. Many times, entrepreneurs do not have enough knowledge and/or skills to evaluate and assess their own business ideas during the early stage. On the other side, investors are often times swamped with hundreds of business plans to evaluate, having to allocate a large percentage of internal, but limited resources to this process often times causing them to miss interesting opportunities. How can be determined if a business venture is viable in advance? If it will be able to survive and reach success?

Viable Framework 1.0 aims to provide a data based framework as a tool to evaluate the viability of early stage startups. This framework can help investors (and corporations) and entrepreneurs by creating a common language between both sides of the development process of new fast growing potential ventures. The methodology to create the framework has been structured in four steps. The first step is the literature review with the goal of analysing previous research in similar fields, defining the concepts of viability, success and survival, and collecting a comprehensive list of factors that might determine success. The second step is to shortlist the initial list of success determinants with the criteria of number of appearances in the articles, academic works and books studied in the literature review. The third step is to ask investors about their own definition of viability, success and survival, and about the importance of these collected and shortlisted determinants of success in order to compare the theoretical and the practical points of view. And the fourth step is to analyse Crunchbase, as a defacto standard database on startups in order to detect some insightful information, contrast with the factors collected in the literature review then measured by investors, and try to find predictive models using data mining and machine learning.

After executing the four steps, we concluded that **the theory and practice meet on the need of strong team capable to lead a brilliant execution**. There is a general alignment, among all the sources analysed about the need of a complete, strong and determined team with the ability to design, plan and execute an scalable business model. The contradictions detected between literature-investor inputs and data inputs reveal the need of building better datasets in order to predict viability by looking at information from different investment stages. Other contradictions, for which we did not find a clear explanation, would require further research and that is why we

Viable Framework 1.0, evaluating viability of startups by using a data based framework



consider Viable Framework 1.0 as the initial version of a more comprehensive framework in the future.

In the meantime, the application and evolution of the Viable Framework 1.0 can be followed at the entrepreneurial project <http://viablereport.com/>.

Statement of the research


The objective of this Master Thesis is to create a framework to evaluate the viability of startups automatically by using data. The framework is called *Viabie Framework 1.0*. The framework is composed by a list of dimensions, areas and factors provided of a weight, corresponding to the relevance of the factor within the total set of factors, together with machine learning predictive models. The purpose of the framework is to help investors (and corporations) and entrepreneurs to assess the viability of the ventures early in the lifetime of their businesses.

The framework is the outcome derived from a research work where we have identified the main determinants of viability (literature review), then we have shortlisted them and we have asked investors which of those would be the most important on their decision making processes (investors survey) and, finally, we have executed different machine learning exercises trying to get the main viability predictors from the patterns founded in other sources of information.

In the literature review section we will introduce the startup development and financing process, what stages exist, who participates on them and how the mechanism works. Then we will define the concepts of viability, survival and success and we will link them to the investment process. After that we will compare emotional approaches and analytical approaches to investment, and then we will finish by studying those data based approaches, understanding why access to certain information is critical on the decision making process before an investment made by a business angel or a venture capital investor. Finally, and supporting the theory development, we will come up with a preliminar but extensive list of dimensions, areas and factors that will establish the first brick in the construction of the framework. This list forms the base of the potential dependent variables that might indicate and provide the needed signals to determine viability, survival and success of startups.

In the data sources and methods used to collect data section we will select sources of information capable of operationalizing the dependent variables and independent variables that we selected from the literature review, and we will present the structured databases, and their curated datasets, together with the survey data from the investors survey used for our analysis. The investors survey aims to get the definition of viability from the investor perspective and detect the most relevant dimensions and factors on their decision making process. The shortlisting of the dimensions, areas and factors to be asked to the investors provides the second brick in the construction of the framework.

In the data treatment and analysis section we will examine the investors survey, first to sort out the definition of viability and then study the importance of the dimension and factors in the



investment process. This last step will set the third brick in the construction of the framework. In addition, we will analyse predictive models based on different datasets and configurations to detect the main signals of viability from the data. By implementing different predictive models of machine learning we will be able to assess if it is possible to predict startup viability from data and the predictive power of such models. This data mining process composes the fourth and final brick in the construction of the framework.

In the end, we will present our conclusions in the form of a table with the main dimensions, areas and factors together with their final relevance, obtained as the sum of the partial relevances determined in the four steps of the framework building. In addition, we will study the value of information, meaning how different datasets and specific variables contribute towards the predictive accuracy of the machine learning models that we will implement, and discuss the different results obtained from the research.

Relevance of the research

Why this research is important? This work aims to analyse the weaknesses and flaws of the startup development and financing process, directing the focus on the investment decision making. The outcome of the research is a data based framework, Viable Framework 1.0, whose purpose is to increase the efficiency of the aforementioned processes at the same time that the risk is reduced. The relevance of the study is given by the theoretical knowledge provided by previous research that is contrasted by what investors really think and what data reveals.

Having a framework to evaluate the viability of startups automatically by using data would lead to increase the efficiency of the investment process, therefore, it would help investors (and corporations) to analyse more and better opportunities. At the same time, having a tool to make informed decisions would lead to decrease the uncertainty, therefore, it would help entrepreneurs to create more successful businesses. Helping investors (and corporations) and entrepreneurs by providing them with efficient tools can have an important impact on the global economy. The entrepreneurial sector is critical for the future development and sustainability of the economy. According to the report *Small and Medium Enterprises (SMEs) Finance* ([1] The World Bank; 2015), entrepreneurship activities and SMEs represent 33% of global GDP and 45% of total employment. Therefore, having ways to measure and predict viability of startups, as one of the main areas of entrepreneurship and SMEs, can be of great help and impact on the shaping of the future economies.

The entrepreneurial economy is booming. In 2014: 4.7 million startups were born in the U.S.; India created 1.91 million successful new businesses; Germany is considered the economic bedrock of the European Union but, per capita, Nigeria has just as many startups; surprisingly the country with the most startups per capita in the world is not US, China or Israel but Uganda ([2] Georgiev, G.; 2015). From those millions of startups that are created every year around the globe, not all of them will be backed by venture capital. And from these ones, 38.1% will remain private, 22.5% will exit on an IPO, 23.8% will finish on a M&A and 15.6% will end up on bankruptcy ([3] Gompers, P.E.; 1995).

On the other side, the failure rate of early stage startups is very high. Some references show qualitative measures: "Some make it, most don't" ([4] Tobak, S.). Others provide catchy numbers and shocking headlines: "Conventional wisdom says 90% of startups fail" ([5] Griffith, E.; 2017) and "90% Of Startups Fail: Here's What You Need To Know About The 10%" ([6] Patel, N.; 2015). And others offer a range of rates: "70% to 90% of ventures that fail" ([7] Furr, N. and Ahlstrom, P.). However, there are numerous data collection problems about which startups are actually created and which of them fail. These ranges of numbers do not allow us to have a clear picture of the

industry and reveal a need for more systematic data collection. Cambridge Associates, a global investment firm based in Boston, concluded that the failure rate has not risen above 60% since 2001 ([5] Griffith, E.; 2017). Data from the US Census Bureau shows that 50% of ventures created in 1992 survived for at least 4 years and 30% of the ventures survived for at least 10 years ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011).

The ecosystem is booming and the opportunities are growing, however, the lack of more structured systems of information on startups make hard to predict success and decide where to invest. There is a need of data based mechanisms to increase the efficiency and mitigate the risk in the process. Investors see around 5,000 companies per year. However, 50% of the investments do come from introductions through the investor's network ([9] Martínez, R.; 2015). In addition, around 5% of the investments they do, were done to companies that have sent them a cold email as first contact. Helping investors in the startup screening process could increase the efficiency drastically and mitigate the risk. By mitigating the risk, investors would feel more comfortable and they would be willing to open their channels to receive investment proposals. According to the NVCA and PwC MoneyTree, during 2014, the total venture capital investment was above \$48 billion, 40 "mega-deals" (investments exceeding \$100 million) were closed, where 2 of them were above \$1 billion, and 71% of the companies that went public were not profitable yet ([10] Roth, M.; 2015). Professional investors review thousands of opportunities before investments. Having access to high quality deal flow is key to building a strong portfolio as well as have access to critical information. On average, it takes angel investors between 20 and 50 hours of due diligence to make an investment decision ([11] 1000 Angels (A Onvest Company); 2016). There are thousands of startups trying to raise money, and it is hard to know which ones will make you money. The whole point of venture investing is to identify startups that have high-growth potential early on and invest before they actually take off. By doing this, the investor assumes more risk, but he/she will be rewarded with significant returns if he/she choose wisely. This level of uncertainty on the investor side makes investors try to find ways to reduce such uncertainty by automated information analysis.

A way to reduce the uncertainty is to establish a common language between investors and startups. Having a common language is very important to determine valuation of early stage companies ([12] Villalobos, L.; 2007). The perspectives on returns on investment for entrepreneurs and investors are different. It is critical to find the path that leads to a common ground between both. The main problem relies on the fact that most entrepreneurs and investors have oblique points of view and that they do not even speak the same investment language. Understanding this divergence could reduce contentiousness and facilitate the construction of an effective negotiation territory between both sides, investors and entrepreneurs. This lack of agreement is often contentious, however, the valuation usually follows certain standard parameters: early stage companies are usually valued at between \$1 million and \$3 million,

investors expect them to grow to between \$50 million and \$100 million, along a lifecycle of between five and eight years ([13] Lipper, G.; 2007). The main concern of investors is about the capacity of the management team to grow the company rapidly. Investors have the goal of helping entrepreneurs to achieve these growth objectives. However, there is a rule of 85/15 ROI where investors typically receive the 85% of their portfolio returns from the 15% of their portfolio companies ([14] Villalobos, L. and Payne, W. H.; 2007).

Additionally, we need to talk about an important emerging stakeholder in the ecosystem: corporations. Some of them act as traditional investors, looking for investment opportunities in their related industry verticals (corporate venture capitals), and others act as innovation partners, looking for synergies with startups where they both can share the benefits of the collaboration. There is a growing trend of corporations looking at startups as a source of innovation, in a more inexpensive, efficient and “certain” way than creating, managing and operating their own R&D departments. U.S. companies invested \$7.5 billion in corporate venture programs in 2015, the highest level this century, regarding the PwC’s 2015 MoneyTree report ([15] Parnell, B.A.; 2017). Other sources provide the following examples: General Motors invested \$1 billion in Cruise Automotive, an autonomous vehicle startup with no product in the market yet; Unilever spent \$1 billion on Dollar Shave Club, a razor startup that contributed just \$200 million in revenue to its global €53.3 billion bottom line; and Wal-Mart invested \$3.3 billion in Jet.com. In 2016, huge multinationals such as Johnson & Johnson and GlaxoSmithKline participated in a venture fund valued at \$200 million where each of them contributed with \$50 million to support biotech startups ([16] Alsever, J.; 2017). Campbell, the soup company, invested \$125 million in a venture fund to work with startups in the food industry. Corporations, like investors, need of tools that help them to discover, evaluate and capture the right startups in efficient ways, meaning on time and at affordable cost. Firms get involved in the creation of prototypes and proofs of concept through the investment in startups which might be later acquired by these big companies through a M&A process. Detecting those appearing opportunities is critical for corporations. “If the first time an established company is made aware of a startup is by receiving a deal book from an investment banker, it’s already too late. Personal knowledge is the first place to start. Most times, established companies are woefully unaware of startups. These companies are too small and fly under their radar” ([17] Yoon, E. and Hughes, S.; 2016). However, there is still a need of creating a common language (as we saw between startups and investors). There are some frictions between startups and corporations and they are reluctant to work together ([18] Goldstein, M.; 2017). In addition, startups shouldn’t feel like they are “selling out” by teaming up with bigger companies, and corporations shouldn’t feel wary of startup culture not jiving with company ethos ([19] Griffith, E.; 2016). Tools such as our framework could lead to reduce those frictions by having access to objective information at the negotiation time. Dave Zilberman, Managing Director at Comcast Ventures, says “The rise in corporate venture capital comes from the realization that

Viable Framework 1.0, evaluating viability of startups by using a data based framework



corporations must look beyond their four walls to access innovation, understand market changes and compete".

Literature review

We aim to create a framework as a tool to help investors (and corporations) to make more agile decisions on the startups they want to invest in. On the other side, we want to help entrepreneurs to know their ventures better and approach investors with a common language. The goal of the literature review is to understand the startup development and financing process, define concepts such as viability, success and survival, differentiate between analytical approaches and emotional approaches on investment decision making and study previous research on similar frameworks. In addition, the outcome from the analysis of previous studies will be a list of factors that could determine viability as the first step of Viable Framework 1.0 building.

Startup development and financing process

An ideal venture needs to go through different stages of development along its life cycle. Each stage has certain development goals, requires different types of financing and presents different levels of available information. The available information is crucial on the decision making process. This information is presented in the form of metrics that will determine the ratio between rewards and risk. And therefore, those different stages will attract different investor profiles. Our framework will evaluate startups in different stages. We need to identify these stages, the information that could be available in each of them, the participants in the process and the way they make their decisions in order to design *Viable Framework 1.0*.

The book *Entrepreneurial Finance* ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011) introduces the following development stages: **Opportunity**, where the entrepreneur identifies and assesses an opportunity, the initial organization is set up, and seed financing is obtained; **Research and development**, where the research & development starts, the market is tested, and R&D financing is obtained; **Start-up**, where initial revenues started to come, the production and sales processes are refined, and start-up financing is needed; **Early-growth**, where the venture works towards breakeven revenue, expands the team and facilities accordingly, and early-growth financing is needed; **Rapid-growth**, where the venture works through proven viability, the team and facilities continue expanding, the track record for harvesting is built, and rapid-growth financing is needed; and **Exit**, where continuing financing is obtained through IPO, M&A or Buyout, and where early investors harvest. In all the aforementioned stages the business plan is created and updated as an alive document that pictures the current situation of the company at any moment.

Although it is clear that startups need to pass through different development and financing stages, the names and labels associated to them are not standard. The study *Determinants among the Internet Startup Life Cycle* ([20] Menkveld, D.J.; 2012) offer four different perspectives

on the startup life cycle stages. According to Kazanjian (1988), the four development stages are: Conception and Development, Commercialization, Growth and Stability. According to Kim and Ha (1999), the four development stages are: Startup, Early Growth, High Growth and Mature. According to Blank (2007), the four development stages are: Customer Discovery, Customer Validation, Customer Creation and Company Building. And according to Berman et al. (2011), the four development stages are: Discovery, Validation, Efficiency and Scale.

Such map of different stages could make us think that we are talking about different processes. However, we can see that is a question of nomenclature. In order to facilitate the understanding, and using the knowledge gathered during the research, we have mapped all the different development and financing stages identified in the literature to our own structure:

Literature Review stages	Viabie Framework 1.0 stages
Opportunity, Research and development, Conception and Development, Customer Discovery, and Discovery	Pre-seed
Start-up, Commercialization, Startup, Customer Validation, and Validation	Seed
Early-growth, Growth, Early Growth, and Customer Creation	Series A
Rapid-growth, High Growth, Customer Building, and Efficiency	Series B
Exit, Stability, Mature, Company Building, and Scale	Series X

In the startup development process, the role of the entrepreneur, as the founder of the startup, and the investor, as the funder of the startup, are key figures in the successful ending on the venture. They both play a game of allocating resources and buying options along the different stages. As we saw before, different stages attract different type of funding and participants in the process. *Entrepreneurial Finance* ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011) shows us the participants involved on the providing economical resources aspect, meaning who is more likely to back the venture in the different stages (according this book definition of stages). We can see that for the R&D stage, the most likely backers are the Entrepreneur, Friends & Family, Business Angels and Corporate Strategic Partner. For the Start-up stage, the most likely backers are Business Angels, Corporate Strategic Partner, Venture Capital and other sources of financing that are out of the scope of our study. For the Early-growth and Rapid-growth stages, the most likely

backers are Corporate Strategic Partner, Venture Capital and other sources of financing that are out of the scope of our study. For the Exit stage, IPO, Acquisition, Buyout and others ways of lending such as public debt are the most common sources of funding.

Again, we did the exercise of mapping the most common ways of financing to our own development stages structure. The result is the following:

Viability Framework 1.0 stages	Most common sources of financing
Pre-seed	Entrepreneur, Friends & Family, Business Angels and Corporation
Seed	Business Angels, Venture Capital and Corporation
Series A	Venture Capital and Corporation
Series B	Venture Capital and Corporation
Series X	IPO, M&A and Buyout

The participants of the financing process described above need to make their decisions to invest. Many business angels work alone, others are affiliated in networks, that allows them joining forces to investigate, discuss and co-invest in opportunities. On the other hand, Venture capital limited partners, who are the capital providers or investors of the venture capital fund, rely on the general partner, who is the decision maker of the venture capital fund, to decide by using sophisticated methodologies ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011). It means that investment decision making is different from one investor to the other. This process is basically composed by screening plans, conducting due diligence on prospective investments and negotiating deals ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011). Our framework aims to help in the efforts of screening plans, partially in the conduction of the due diligence and, thanks to the common language built, it would facilitate the deal negotiations.

Definition of Viability, Success and Survival

Viability Framework 1.0 aims to evaluate and predict the viability of startups by using data. It is very important to define first the meaning of Viability and two related concepts: Survival and Success.

Looking at the dictionary ([21] Cambridge Dictionary), we can find the following straightforward definitions:

1. Viability: Able to work as intended or able to succeed. Able to continue to exist as or develop into a living being.
2. Success: The achieving of the results wanted or hoped for. Something that achieves positive results.
3. Survival: The fact of a person, organization, etc. continuing to live or exist. Something that has continued to exist from a previous time.

We have analysed and compared different opinions on the topic. Some sources shows viability from the perspective of success. The book *The Illusions Of Entrepreneurship* ([22] Shane, S.A.; 2008) defines success as the capacity of reaching the performance objectives, that for any company means to meet the expected sales and profits. In addition, a continuous growth of sales, profits, capital and assets, employees and net value is an indicator of success. It also associates success with something beyond business economics: contribution to society. On the other side, it considers success, from a less optimistic point of view, as the fact of closing a business voluntarily where the entrepreneurs and creditors don't lose their money.

From the perspective of survival, Eduardo Punset compares startup survival process with the natural selection of the evolution theory created by Darwin ([23] Punset, E.; 2004). According to Darwin, natural selection is the engine of the species propagation. Evolution of living beings and companies are based on 2 strategies: cost reduction (efficiency) and the most adequate niche search (specialization and differentiation). Finding the niche means to meet a place within an ecosystem that enables solving the feeding problem (energy needed to exist) and competing with other species to avoid being killed (survive). These actions have to be executed in a very efficient way in order to avoid being extinct. The threat of extinction can only be avoided by differentiation. Viability depends on the balance between both strategies: cost reduction and niche search.

Other references, such as *Entrepreneurial Finance* ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011), offers definition of viability as a mix between survival and success. Viable companies are able to resist and they are durable along the time, reducing the risks and uncertainty and attract external investment. This capacity of surviving would enable to continue operating the venture, meeting the objectives of growth (sales revenue, net income, cash flow available to investors, etc.) until the venture manages to successfully exit by going public (IPO), private selling of the venture to another firm (M&A), selling of the business to employees or other members of the team, or by a management buyout (MBO). These authors also mention that success depends on building a competitive advantage that is sustainable and success is facilitated by a brilliant execution with a lack of failure. Again, as the previous references, it leaves room for less optimistic definitions of viability such as the lack of bankruptcy. *Business for Punks* ([24] Watt, J.; 2016) introduces a punk and straightforward relation between success and survival: "Survival is always the first step to

success". The book presents the fundamental ingredient in the formula of viability, then meaning survival, and then meaning success: cash, "The lifeblood of your business is cash. if you can't manage a cash flow, then you can't run a business".

Relating those concepts with the dangers behind choosing the proper development and investment strategy, Eze Vidra, Managing Partner at *Google Ventures*, reveals that the entrepreneurs are focused on trying to create unicorns (startups which are valued at over \$1 billion ([25] Fortune; 2016)) because it is what investors are looking for ([26] Pimentel, A.; 2016). This behaviour, in his opinion, is creating a dangerous bubble. Nowadays it is cheaper than ever to create a startup. However, it is more difficult to make them sustainable. He recommends to focus on the survival factors, the ones that make a startup survive to a nuclear war, like a cockroach (a startup comprised of resourceful founders who survive no matter what ([27] Draper, A.; 2015)), instead of trying to create unicorns.

The literature review gave us different visions on the definitions of viability, success and survival. However, they are still framed on a theoretical perspective. We wanted to know what the investors think of viability, success and survival. They are the ones that will make their decisions based on their abilities to read the future behaviour of companies. So we collected all these definitions (*Appendix 1. Definition of Viability*) in order to ask investors with a survey.

According to the book *Entrepreneurial Finance* ([8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011), the viability will depend on two times:

1. Short-run: Entrepreneurs and investors will use the best available methods and information to decide which ventures to pursue and how to structure the deals
2. Long-run: Entrepreneurs and investors will develop organizational structures that are agile and can react quickly to changing environments (transformative, disruptive and transitory fluctuations) and opportunities

We can say that the viability of a startup will be determined by its capacity to survive and be successful, meaning sustainable in the first instance and profitable in the long run. The target variable of our framework will be Viable. The value can be "Yes", when we detect present indicators of survival for future success, and "No", when we detect present indicators of survival for future failure. Therefore, in our framework we could consider a Viable company as the one that poses the present indicators about the concept development ready for chasing new opportunities, a good market fitting together with the ability to compete and the metrics of financial growth, managed by a team capable to perform a brilliant and sustainable execution. These ingredients of the formula would lead to a successful exit (by IPO, M&A or buyout) in the future. We have the theoretical definition of Viability, that will be contrasted with the investor's definition through the investors survey. In addition, the potential determinants of viability will be

studied and collected along the following sections of the literature review. Then, again, we will measure the importance of these indicators from the investor's point of view (with the investors survey) and finally we will contrast them with the data mining exercise.

Analytical approach vs. emotional approach on the investment decision making process

Viable Framework 1.0 is an analytic framework that uses data to facilitate the investment decisions. Before analysing different data based frameworks as the foundations of ours, we would like to study briefly the other method of investment and making decisions, based on intuitions and feelings (emotional approach), and analyse the differences with our data based method (analytical approach). By comparing both approaches and see how they could interact together, we could understand better the strengths and weaknesses on the investment decision making process that are purely based on data.

As we have seen, managing objective information is key at making decisions. The objective information is relevant for investment decisions and having access to data about the team behind a startup, or some execution and financial metrics facilitates the selection of the right options. But this information is not only important at micro level, in the relation between the entrepreneur and investor but also at macro level, where the environment and context plays an important role. Knowing where an entrepreneur or an investor is playing is fundamental. In some cases, decisions are made by common wise beliefs. The book *The Illusions Of Entrepreneurship* ([22] Shane, S.A.; 2008) describes some notes and topics that are commonly assumed and they are not true (demonstrated by analysing data). For example, "The United States is not one of the most entrepreneurial countries in the world", or even more surprising facts that could make the worldwide entrepreneurial scene shake on its ground such as "Psychological factors account for very little of the difference between entrepreneurs and other people, much less than demographic factors like age, race, and gender", or "Most entrepreneurs do not select the most profitable industries but instead pick industries with the highest firm failure rates", or "Contrary to most people's mental image, the typical start-up is a very ordinary, not-very-innovative, home-based business that starts and stays tiny". Having access to data could remove those common wise beliefs in order to make an informed decision based on reality.

Data is very important but investing is not only an analytic science. It has much of art and human side, and emotions are working under the surface in ways that we can't escape ([28] Statman, M.; 2013). According to the book *What they don't teach you at Harvard Business School* ([29] McCormack, M.H.; 1984), the decision making process is more based on intuitions than on information. It points out that having so much data to analyse is dangerous because it can undermine the power of intuitive decisions ([29] McCormack, M.H.; 1984). Emotions are very

important in every decision making process. "Everything starts with an emotion" is said by Eduardo Punset on his book *El viaje a la felicidad: las nuevas claves científicas* ([30] Punset, E.; 2005). Again, Eze Vidra, in the interview *Em vez de unicórnios, transformem-se em baratas. Sobrevivam à guerra nuclear* ([26] Pimentel, A.; 2016), pointed out the importance of the human empathy at the time of investing.

Both approaches (analytical and emotional) are needed on the investment process, as a decision making process. The most recent and revolutionary discovery on this topic was made by Dylan Evans, from the School of Computer Science, Engineering and Mathematics of the University of the West England in Bristol, who demonstrated that decisions, all decisions, are emotional ([30] Punset, E.; 2005). Then, what is the roadmap of every decision making process? At the beginning there is an emotion (Fast). Then, a rational computational process analyses the options with all their available information (Slow). With so much information logic is not able of determine on its own. Finally, emotions re-appear to save the situation. Without emotions we can not make decisions. The presence of emotions is bipolar: they are at the beginning and the end of every human projects. "The final outcome of any complex problem still relies on a blend of human intuition and artificial intelligence" ([31] Reddy, S.; 2017).

Although Viable Framework 1.0 is on the analytical side, we now understand the importance of certain emotions at the time of making a decision, and how they could support, complement or even abort an informed decision.

Analysis of different data based frameworks to assess startups viability

Ira Sager, from Businessweek magazine, invited Bob Goodson, Quid AI's CEO, to take part in a challenge in 2009. The challenge consisted in asking a computer to pick the 50 startups that would rock the world in the future. The results, published in 2017, were astonishing. Evernote, Spotify, Etsy, Zynga, Palantir, Cloudera, OPOWER, and the list goes on. The list not only contained the names of some of the most widely-known companies nowadays but it contained other cases of great performers in terms of growth. This fact could end with the traditional venture capital's belief about the impossibility of using machines to facilitate the investment decision making, and even the prediction of startup success ([31] Reddy, S.; 2017).

Numerous factors affect how investors assess startups. The most typical way of asses a startup is the valuation. The importance of a strong team, the opportunity size or the potential of the company to scale are the main drivers for the investors valuation ([32] Payne, W. H.; 2007). There are proto tools, in the form of worksheets, helping entrepreneurs to understand what investors look for and investors can compare different opportunities. The Valuation Worksheet ([33] Payne, W. H.; 2007) is one of the basis of our *Viable Framework 1.0*. Until this point, the focus is on valuing pre-money, that means to determine the value of a startup before that external investors

come to the equity structure and, therefore, it is not based on a financial valuation. However, the method most used by venture capitalists is to determine the post-money valuation: “Post-money valuation: The valuation of the company immediately after a round of investment is closed. Pre-money valuation: The valuation of the company just before closing a new round of investment, including the value of the idea, the intellectual property, the assembled management team, and the opportunity. Post-money valuation = Money + Pre-money valuation” ([34] Payne, W. H.; 2007). Valuation of startups brings other complexities and incorrect assumptions to the table, such as the valuation divergence, which states that the shares value increases proportionally with the company valuation increase, and it is made by both entrepreneurs and investors. Similarly, an increase in the company’s market valuation will not imply a proportional increase in the returns on investments ([35] Villalobos, L.; 2007). Valuing early-stage startups is an art and very complex process, even if we have tools available, such as the Valuation Estimator ([36] Payne, W. H.; 2007). This tool, created by Akira Hirai, managing director of the Phoenix office of Cayenne Consulting, can value startups by answering only twenty-five questions.

The approach of using an analytical framework to assess early-stage startups was also brought to the table by Walter Palma, Director at Caixa Capital, during his conference at *Lisbon Investment Summit 2016* ([37] Palma, W.; 2016) where he described a framework called 4M’s which uses the following factors to evaluate startups from the VC (Venture Capital) perspective: Large market, taking a small pie still makes a big business; Disruption, disrupting the rules & incumbents of this market; Capital-efficient model, limited capital needs to grow fast; Unique strengths, making credible that you will scale & win.

Almost every venture capital firm has its own framework and it has not been the first time to research about the standardization of this evaluation process. There were other academic projects attempting this mission. However, any of these previous studies take into account the three different angles considered in this one: literature review, investor's perspective and data insights. *Determinants among the Internet Startup Life Cycle* ([20] Menkveld, D.J.; 2012) tries to find the determinant factors in the success prediction from the theoretical perspective based on previous academic research. For that, its framework offers the perspective of the following stages, called *Internet Startup Life Cycle*:


1. Discovery can be best described as finding the customer. Besides: create a founding team, conduct customer interviews, find value proposition, build a minimal viable product, join an incubator, invite first advisors for the board, and finance it by themselves
2. Validation can be described as validating the business model. Furthermore: refine core features, initiate user growth, implement metrics and analytics, get seed funding, hire first employee, and find your product market fit

3. Efficiency can be best described as optimizing product and processes. Besides: refine the value proposition, improve the user experience, optimize conversion funnels, achieve viral growth, and find repeatable sales process and scalable user acquisition channels
4. Scale can be best described as conquering the market. Besides: get a large round, start massive customer acquisition, improve back-end scalability, hire first executive, implement processes, and establish departments

Our goal is to identify the determinants of viability (or success or survival). The million (or billion) dollar question is: What makes a venture capital investment successful? The article *4 Factors That Predict Startup Success, and One That Doesn't* ([38] Marion, T. J.; 2016) describes the analysis of a dataset with 300 startups and 600 entrepreneurs. The factors analysed are related to the founders themselves such as age, gender, education, location, and prior work and startup experience. The study discovered correlations between four of those factors and success: gender, age, education, and prior work and startup experience. However, the analyses didn't find any correlation with the location factor. The article also mentions the change in the investment behaviours of the BAs (Business Angels) and VCs (Venture Capitalists). They used to invest exclusively through their own networks but now they are opening their minds to other channels and sources. This fact creates new opportunities for tools and systems based on information services, such as the framework that we are building, which can act as filters and funnels facilitating the decision making.

Statements like "It is not only about ideas but it has more to do with execution" are frequently heard. However, it is not only about execution but about to choose the right time as well. The correct timing needs to be complemented by other relevant factors, such as the persistence and the determination of the founders ([29] McCormack, M.H.; 1984).

There is a change on the VCs and BAs mindset about their approach to investment. This change is partially thanks to the myriad available tools and access to information that exist nowadays. On one side, the amount of data to be analysed is much bigger. On the other side, and since the information belongs to private markets, it could be opaque in comparison to the public markets. On the other side, the information provided by the entrepreneur is too positive and optimistic that makes very difficult the objective analysis of the real opportunity behind the project ([39] Shalman, W.A.; 1996). "Business plans, each proposing to "revolutionize" an industry, each "conservatively" projecting at least \$50 million in revenues within five years based on a modest market share of under 10%, and each containing a projection of likely investor returns of over 100% per annum.". At the same time these tools, help investors to be aware of the signals that could help to identify a successful opportunity during the screening, evaluation and further due diligence process ([40] 1000 Angels (A Onvest Company); 2016).



By analysing these articles, academic works and books, we have produced the first brick in the construction of the *Viable Framework 1.0*. The full list contains 194 factors that have been identified during the literature review and they have been categorised into 25 areas and 6 dimensions (Concept, People, Context, Execution, Financial and Legal). Each factor has been provided with the “Relevance in the Literature Review”, a qualitative value depending on the number of appearances of this factor in the articles, academic works and books checked in our literature review and the distance to the average of appearances: 2.65 (and rounded to 3). This first piece of the *Viable Framework 1.0* can be found in the *Appendix 2. Viable Framework - Step 1*. We have identified some insightful common grounds among the different methodologies. However, this is still a field very theoretical and it is very far away from the main users of these frameworks: investors (and corporations). With the purpose of building a standard and more comprehensive framework, we decided to create a survey, that will be analysed in the further sections, to compare the collected theoretical knowledge with the mindset of the investors. In addition, we will complement the study by contrasting with data, analysing one of the main sources of information about startups: Crunchbase.

Data sources and methods used to collect data

The data to be analysed will come from two different sources:

- Investors survey
- Crunchbase (structured database)

Investors survey

In order to collect the information regarding the investor perspective on the topic, we created a survey. The name of the survey is: “Viability from the Investor perspective” and it contains 2 sections:

- General Information
- Viability Factors

General Information section will provide the information about the investor profile, investment preferences such as the stage of investment, geographical area and industry vertical. This section includes one of the most relevant questions and part of the core of our study: the definition of viability. Each of the asked investors will select its definition of viability by completing the statement starting by “Your definition of viability has to do with...” and selecting among the values offered (and they will be the same discovered in our literature review and summarised in the section Definition of Viability, Survival and Success).

On the other side, Viability Factors section will go deeper on the analytical factors that could help investors to make their decisions. These factors were shortlisted from the general list of Dimensions, Areas and Factors composing the first piece of the *Viable Framework 1.0 (Appendix 2. Viable Framework - Step 1)*. The shortlisted version will be composed by 60 factors, 10 factors corresponding to each dimension. We selected the 10 factors with the highest “Relevance in the Literature Review” (Number of appearances in the Literature Review) of every dimension. The final shortlisted version can be found on the annex *Appendix 3. Viable Framework - Step 2*. This shortlist exhibits the second step in the building of our *Viable Framework 1.0*.

The details about the investors survey can be checked on the *Appendix 4. Investors survey - Viability from the Investor perspective*.

Crunchbase

Crunchbase is our selected structured database to operationalize the dependent variable (Viable) and potential independent variables. We chose Crunchbase due to its relevance in the startup investment scene. The details about this database are on the *Appendix 5. Crunchbase - Details*.

We wanted to analyse two datasets based on different investment stages, Seed and Series B, to get different insights and conclusions. As we saw previously, these two different stages should present different available information that could facilitate the decision making. Crunchbase presents information according to the last update of the data regardless of any particular stage. In order to avoid the data leakage, a problem of machine learning that appears when external information from the training dataset is used to create the model and produces optimistic or invalid predictive models ([41] Brownlee J.; 2016), we performed the featured selection and then data preparation within the cross validation folds technique. Cross validation allows the data rescaling process to have information on the distribution of data in the training dataset at the time of scaling factors calculations ([41] Brownlee J.; 2016). Based on that, we built the two datasets by selecting the possible attributes that could be available for a seed stage company and for a series B stage company. For example, we could not use variables that seem very interesting for our analysis, such as `employee_count`, because the database does not provide neither the date from when this information was collected nor the stage it can be associated with. If we included information like this in the model, we might have an optimistic and invalid prediction. This fact reveals the importance of collecting accurate and updated information in order to produce predictive models with better quality and precision in the future.

For the operations of retrieving the data (from a XLS format database), preparing the data, modeling the data and visualizing the data, we used the tool *RapidMiner Studio 7.6.000 Free Edition for Students*.

Dataset 1 - Seed6040

We have identified companies that passed through a seed stage and they had the corresponding available information. Our target variable is Viable (with the values equal to “Yes” or equal to “No”). We defined a Viable company (Viable = “Yes”) as a company privately acquired by another firm and / or presenting a sustainable competitive advantage that allows to perform well in the market and meet the objectives of continuous growth: sales, profits, capital & assets, employees and net value. However, in Crunchbase we could not identify these criteria (not enough data available). We will assume that a Viable or Successful company is a company that has been privately acquired by another firm (“acquired”) or has been offered to the public (“ipo”). On the other side, a Non Viable or Failed company (Viable = “No”) is a company that has ended of

business (“closed”). Some metrics of performance in machine learning algorithms are sensitive to class proportions and we want to take this fact into consideration. As we saw in the section *Relevance of the research*, it is very hard to come up with a concrete and concise failure rate. For this exercise, we will choose 60% as startup failure based on the previous statement: “Cambridge Associates, a global investment firm based in Boston, concluded that the failure rate has not risen above 60% since 2001”.

With these assumptions in mind, our dataset 1 has the following form:

- 2581 examples (60% corresponding to failed companies and 40% corresponding to successful companies):
 - 1549 with status = “closed”
 - 516 with status = “acquired”
 - 516 with status = “ipo”
- Statistics of the variables:

Name	Meaning	Values
new_Viable (label)	The target or dependent variable. It was newly generated by mapping the values from the existing status variable: Viable is equal to “Yes” when status is equal to “acquired” or “ipo”, Viable is equal to “No” when status is equal to “closed”	Least: Yes (1032) Most: No (1549) Values: No (1549), Yes (1032)
country_code	The code of the country where the startup is based in	Least: CAN (110) Most: USA (2471) Values: USA (2471), CAN (110)
state_code	The code of the state where the startup is based in	Least: WV (1) Most: CA (932) Values: CA (932), NY (262), ...[55 more]
region	The region where the startup is based in	Least: Winston-Salem (1) Most: SF Bay Area (653) Values: SF Bay Area (653), New York City (230), ...[182 more]
city	The city where the startup is based in	Least: Yuba City (1) Most: San Francisco (247) Values: San Francisco (247), New York (197),

		...[626 more]
founded_on	The date when the company was founded	Least: 2015-06-25 (1) Most: 2007-01-01 (120) Values: 2007-01-01 (120), 2010-01-01 (116), ...[542 more]

- Other considerations:
 - We filtered all the examples with any of their attributes missing (empty). We kept the examples with all their attributes filled

Dataset 2 - Series6040

We have identified companies that passed through a Series B stage (`funding_rounds = 2`) and they had the corresponding available information. Our target variable is *Viabie* (with the values equal to “Yes” or equal to “No”). We defined a *Viabie* company (*Viabie* = “Yes”) as a company privately acquired by another firm and / or presenting a sustainable competitive advantage that allows to perform well in the market and meet the objectives of continuous growth: sales, profits, capital & assets, employees and net value. However, in Crunchbase we could not identify these criteria (not enough data available). We will assume that a *Viabie* or *Successful* company is a company that has been privately acquired by another firm (“acquired”) or has been offered to the public (“ipo”). On the other side, a *Non Viabie* or *Failed* company (*Viabie* = “No”) is a company that has ended of business (“closed”). Some metrics of performance in machine learning algorithms are sensitive to class proportions and we want to take this fact into consideration. As we saw in the section *Relevance of the research*, it is very hard to come up with a concrete and concise failure rate. For this exercise, we will choose 60% as startup failure based on the previous statement: “Cambridge Associates, a global investment firm based in Boston, concluded that the failure rate has not risen above 60% since 2001”.

With these assumptions in mind, our dataset 1 has the following form:

- 750 examples (60% corresponding to failed companies and 40% corresponding to successful companies):
 - 450 with status = “closed”
 - 150 with status = “acquired”
 - 150 with status = “ipo”
- Statistics of the variables:

Name	Meaning	Values
------	---------	--------

new_Viable (label)	The target or dependent variable. A newly generated variable by mapping the values from the existing status variable: Viable is equal to “Yes” when status is equal to “acquired” or “ipo”, Viable is equal to “No” when status is equal to “closed”	Least: Yes (300) Most: No (450) Values: No (450), Yes (300)
country_code	The code of the country where the startup is based in	Least: CAN (34) Most: USA (716) Values: USA (716), CAN (34)
state_code	The code of the state where the startup is based in	Least: VT (1) Most: CA (325) Values: CA (325), NY (66), ...[41 more]
region	The region where the startup is based in	Least: WI - Other (1) Most: SF Bay Area (245) Values: SF Bay Area (245), New York City (59), ...[92 more]
city	The city where the startup is based in	Least: Yorba Linda (1) Most: San Francisco (86) Values: San Francisco (86), New York (54), ...[265 more]
founded_on	The date when the company was founded	Least: 2015-06-25 (1) Most: 1999-01-01 (39) Values: 1999-01-01 (39), 2005-01-01 (35), ...[186 more]
funding_total_usd	The total funding that the startup has captured up to Series B, in US dollars	Min: 4500 Max: 2100000000 Average: 28329145.322
new_funding_period	A newly generated variable by calculating the difference between the existing variables: last_funding_on and first_funding_on	Min: -0.074 Max: 39.027 Average: 1.763

- Other considerations:
 - We filtered all the examples with any of their attributes missing (empty). We kept the examples with all their attributes filled
 - We took into account startups in Series B, meaning that the attribute `funding_rounds` was equal to “2”

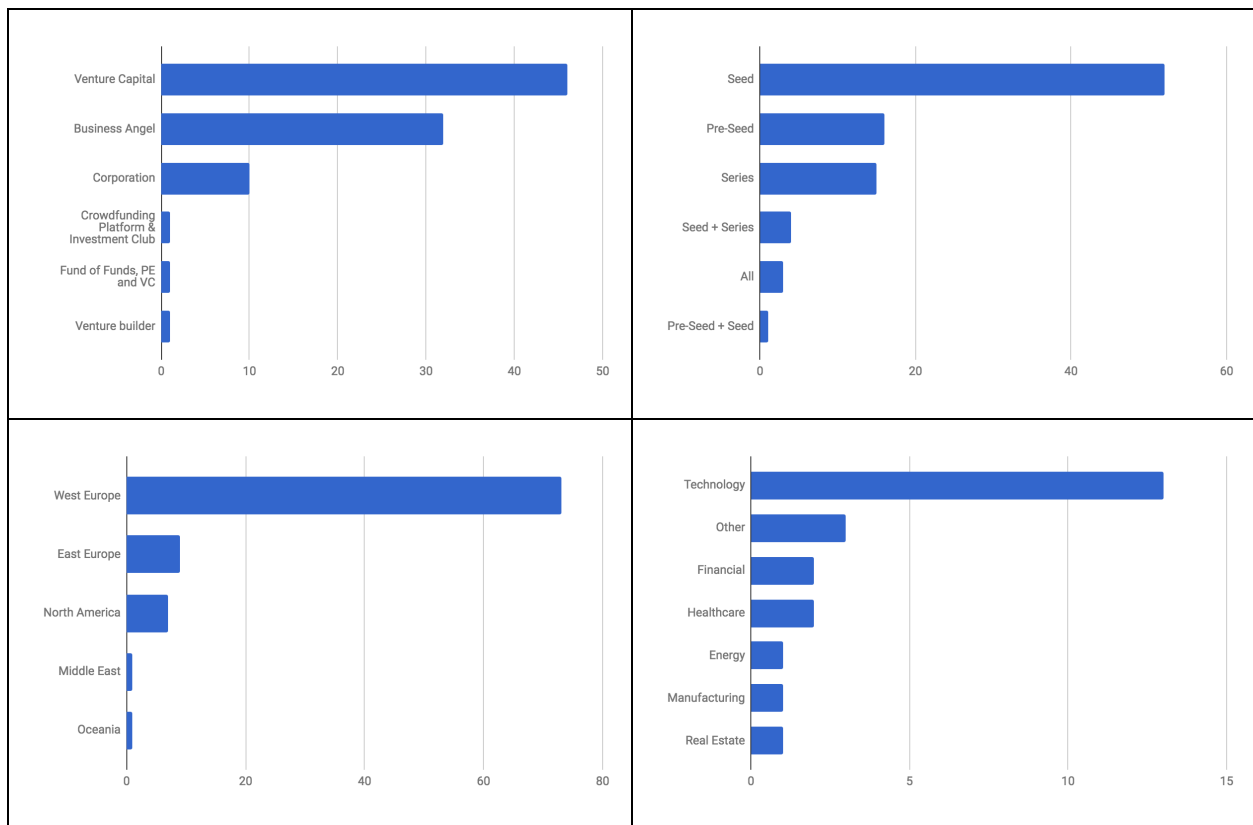
Data treatment and analysis

Investors survey

Demographics of the survey

We are presenting the demographics of our respondents because we want to identify and highlight the demographic group that our conclusion can generalize to. The survey has been sent to more than 1,300 investors and we got 91 answers, meaning a response rate of 7%. We collected the e-mails of investors from different sources: government and public lists with accredited investors contacts, websites of venture capital firms, business angels and investors associations, and our own private contacts. E-mail response rates go between 25% and 30% without follow up, and they are decreasing in the last years ([42] Fincham J.E.; 2008).

The structure of the answers, in terms of demographics, is very diverse. The following graphics show the statistics about the different “personas”:



According to the Investor profile, 50.5% of the answers are from Venture Capital investors, 35.2% belong to Business Angels, 11% correspond to Corporations and the rest is divided into other profiles: Different type of funds, Venture builders or Crowdfunding platforms.

The most common Stage of investment has been Seed with 57.1% of the share, Pre-Seed holds a 17.6% of the share and Series holds a 16.5% and, again, the rest is for other categories that in most of the cases mix the previous mentioned stages: Seed, Series and Pre-Seed.

In a similar way, in terms of Geographical area, the answers are more abundant in West Europe, represented by a 80.2%, in the meantime that East Europe and North America represent 9.9% and 7.7% respectively. The remaining portion aggregates the rest of the designated regions: Middle East, Asia, South America and Oceania.

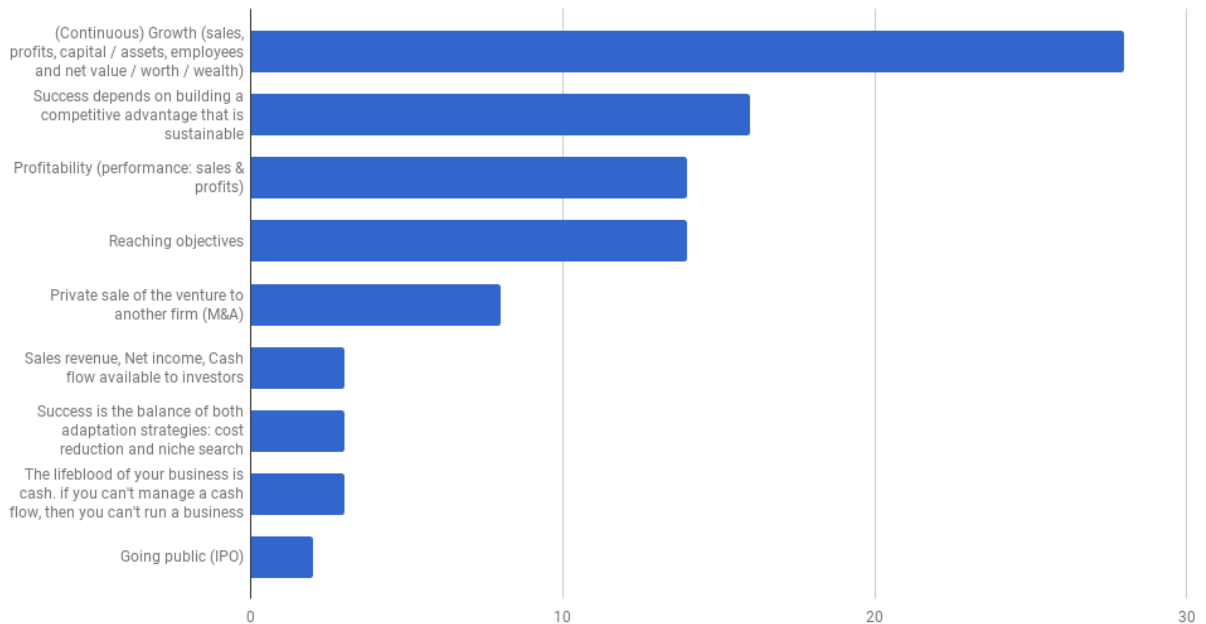
Industrial vertical was a question added later and it doesn't contain the full set of answers. We managed to get 23 answers for this new question. Most of the answers attend the vertical of Technology with a 47.8% of them, followed by Healthcare and Financial verticals with 8.7% each, and then other categories such as Real Estate, Online, Manufacturing and Energy with ratios of 4.3%.

Therefore, our average investor is a Venture Capital focus on Seed stage startups from West Europe looking for opportunities in the industry of Technology.

The whole detail of information can be checked on the *Appendix 6. Investors survey - Viability from the Investor perspective - Answers*.

Definition of Viability

In the section Definition of Viability, Success and Survival, we defined those concepts based on previous research, articles, academic works and books. But still we wanted to know the perspective from the investors (and corporations), the potential users of our framework. We asked investors "Your definition of viability has to do with..." and they answered:



The range of answers are the definitions identified in our literature review and that we catalogued and summarised in the *Appendix 1. Definition of Viability*. The highlights of the graphic above are:

- (Continuous) Growth (sales, profits, capital / assets, employees and net value / worth / wealth), backed by 30.8% of the investors asked
- Success depends on building a competitive advantage that is sustainable, backed by 17.6% of the investors asked
- Profitability (performance: sales & profits), backed by 15.4% of the investors asked
- Reaching objectives, backed by 15.4% of the investors asked
- Private sale of the venture to another firm (M&A), backed by 8.8% of the investors asked

After analysing previous research and contrasting with investors, our definition of Viability is about being privately acquired by another firm and / or presenting a sustainable competitive advantage that allows to perform well in the market and meet the objectives of continuous growth: sales, profits, capital & assets, employees and net value. Surprisingly, although IPO is a (public) acquisition and an objective successful exit, it is not in mind of most of the investors asked in this survey. The reason could be that IPO is an exit that is still distant in time and not in the short term of what investors are looking for, so we will assume that and we will consider it as part of the definition. Therefore, the target variable of our framework, Viabie, will be “Yes” when we find a startup that matches the aforementioned criteria. And it will be “No” in any other case.

The whole detail of information can be checked on the *Appendix 6. Investors survey - Viability from the Investor perspective - Answers.*

Viability factors

We asked investors about the potential viability factors as well, meaning the factors that may drive a startup to be viable. The possible answers (factors) of the survey are the ones shortlisted from the whole set of the *Viability Framework 1.0*. They can be found on *Appendix 3. Viability Framework - Step 2.*

First, we asked investors about the level of importance of the dimensions (the groups of factors organised by category) and we computed the average of the scores, assuming an equally spaced scale going from “very high” valued at 4 to “very low” valued at 0, and then we have ranked the dimensions by score with its mapped (we mapped above the .5 as the next level) qualitative values:

Dimension	Average	Mapping
People	3.87	Very High
Execution	3.53	Very High
Context (Market, Competition...)	2.99	High
Concept (Idea, Product...)	2.92	High
Financial	2.22	Medium
Legal	1.89	Medium
Other	1.62	Medium

We will set the *Relevance in the Survey (Dimensions)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viability Framework - Step 3.*

Then we analysed all the dimensions in order to identify the most important factors in each of them. We asked investors about the level of importance of the CONCEPT factors and we assumed and followed a similar process to get the qualitative values:

CONCEPT factor	Average	Mapping
Product/Market fit	3.65	Very High
Problem/Solution fit	3.53	Very High
Opportunity	3.22	High
Innovation	3.18	High
Vision	2.95	High

Mission	2.34	Medium
Product vs. Service	2.32	Medium
Location	1.88	Medium
Trendiness	1.84	Medium
Social impact	1.74	Medium

We will set the *Relevance in the Survey (Factors)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viable Framework - Step 3.*

We asked investors about the level of importance of the PEOPLE factors and we assumed and followed a similar process to get the qualitative values:

PEOPLE factor	Average	Mapping
Commitment	3.84	Very High
Motivation	3.80	Very High
Adaptability	3.52	Very High
Completeness of the team	3.22	High
Industry experience	3.08	High
Entrepreneurial experience	3.01	High
Professional experience	2.88	High
Reputation	2.82	High
Network	2.70	High
Academic level	2.01	Medium

We will set the *Relevance in the Survey (Factors)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viable Framework - Step 3.*

We asked investors about the level of importance of the CONTEXT factors and we assumed and followed a similar process to get the qualitative values:

CONTEXT factor	Average	Mapping
Product/Service differentiation	3.40	High
Go-to-market strategy	3.30	High
Market growth	3.03	High
Market size	3.01	High

Barriers to entry	2.88	High
Competitors' strength	2.82	High
Industry	2.62	High
Competitors' structure	2.46	Medium
Partners' structure	2.29	Medium
Economical environment	2.24	Medium

We will set the *Relevance in the Survey (Factors)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viable Framework - Step 3*.

We asked investors about the level of importance of the EXECUTION factors and we assumed and followed a similar process to get the qualitative values:

EXECUTION factor	Average	Mapping
Scalability	3.57	Very High
Business model	3.08	High
Sales & Distribution channels	2.99	High
Momentum	2.76	High
Operations plan	2.66	High
Number of users	2.56	High
Number of clients	2.55	High
Marketing plan	2.54	High
Financial plan	2.52	High
Development stage	2.49	Medium

We will set the *Relevance in the Survey (Factors)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viable Framework - Step 3*.

We asked investors about the level of importance of the FINANCIAL factors and we assumed and followed a similar process to get the qualitative values:

FINANCIAL factor	Average	Mapping
Burn rate	3.09	High
Credible external investors interested in investing now	2.97	High

Previous external investors interested in investing now	2.85	High
Cash flows	2.82	High
Amount required	2.79	High
Revenues in 5 years	2.77	High
Previous external investors (Cap table)	2.76	High
Revenues in the next 12 months	2.49	Medium
Capitalization	2.32	Medium
Current revenues	2.09	Medium

We will set the *Relevance in the Survey (Factors)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viable Framework - Step 3.*

We asked investors about the level of importance of the LEGAL factors and we assumed and followed a similar process to get the qualitative values:

LEGAL factor	Average	Mapping
Terms & Conditions of the deal	3.07	High
Fairness of the deal	3.05	High
Robustness of the deal	2.87	High
Lack of perverse incentives	2.81	High
Simplicity of the deal	2.70	High
IP status	2.63	High
Incentive effects of the allocation risk vs. reward	2.62	High
Intellectual property firm	2.38	Medium
Lack of foreclose valuable options	2.33	Medium
Incorporation form	1.95	Medium

We will set the *Relevance in the Survey (Factors)* with the same values of the Mapping column above. The final results can be checked on the *Appendix 3. Viable Framework - Step 3.*

We did the analysis about the investor profiles and their answers based on a descriptive and aggregate terms. We would like to further explore this topic and segment different investors

profiles and conduct an analysis based on clustering. This point is beyond our initial objective and we will leave it for the future.

Comparing the results of the relevance of factors according the literature review and the investors survey, we see that some outcomes are aligned, however, others are contradictory. For example, at the concept level, we can see that Product/Market fit is relevant and very relevant in both sources, the same happen with Location that shares a medium level importance in the literature and the survey. On the other side, Opportunities presents a very low relevance in the literature review in the meantime it presents high importance in the investors survey. At the people level, we can see almost a full alignment if we make obvious the light difference between high and very high levels of importance in some cases. Something very similar happens with the context, execution and financial dimensions. Where we see the highest degree of contrast is in the legal dimension. Besides the Incorporation Form factor that presents a medium level of relevance in both sources, and IP Status which is similar with high values, the rest of factors reveals a low / very low level of importance in the literature review in the meanwhile they present high / very high levels in the investors survey. It could indicate a “new” concern on investors for the legal process of building the entrepreneur-investor agreement, that still has not been studied in deep from the academic perspective.

The full details of the answers can be found on *Appendix 6. Investors survey - Viability from the Investor perspective - Answers*.

The answers of the investors, about the importance of the dimensions and factors, represent the third brick in the construction of the *Viable Framework 1.0*. The details can be found in the *Appendix 7. Viable Framework - Step 3*.

Crunchbase

Finally we turn to the objective of predicting startup future behaviour only from data on business ventures that is available for investors. The chosen source is Crunchbase that offers an extensive database on startups. Our target variable is Viable and it can take the value 'Yes', a viable company and 'No', a non viable company. We are going to try to predict what factors determine the value of Viable equal to 'No' ('No' will be the positive class).

The problem that we are trying to solve is a classification problem. Classification models predict the class (usually the classes are mutually exclusive) which an individual of a population belongs to ([43] Provost F. and Fawcett T.; 2013). This is a supervised problem because we have a target variable and we know the values of it in the training dataset. Our problem is a classification (rather than a regression) one because the value of the target attribute is categorical (yes or no) rather than numerical ([43] Provost F. and Fawcett T.; 2013). The classification tree or decision tree is a type of model that solves and predicts classification problems by the segmentation of classes generated by induction. The tree is made up of nodes, interior nodes and terminal nodes, and branches emanating from the interior nodes, and it creates a segmentation of the data where every data point will correspond to one and only one path in the tree, and thereby to one and only one leaf. In other words, each leaf corresponds to a segment, and the attributes and values along the path give the characteristics of the segment ([43] Provost F. and Fawcett T.; 2013). We decided to use a decision tree for its effective way of modeling and representing data for the kind of problem that we are trying to solve.

The decision trees have some specific features and parameters associated. The definitions of the control parameters and performance metrics can be find on *Appendix 8. Crunchbase - Definitions*.

Dataset 1 - Seed6040

In this exercise with seed stage startups we have used a Decision Tree predictive model. We tuned the parameters of our model using 10-fold cross validation to compute out of sample measures of performance and assess the quality of our model.

We have configured the following control parameters:

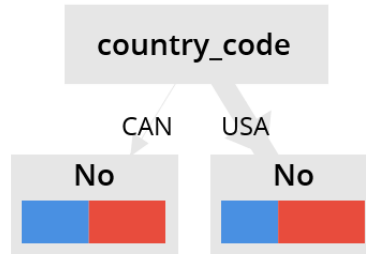
- Criterion: gain_ratio
- Maximal depth: 20
- Apply pruning: yes
- Confidence: 0.25
- Apply prepruning: yes

- Minimal gain: different values
- Minimal leaf size: 2

We have done a sensitivity analysis playing with different values of the parameter Minimal gain and different outcomes of the AUC (ROC):

Minimum gain	AUC	Minimum gain	AUC
0	0.506 +/- 0.010	0.01	0.500 +/- 0.000
0.0001	0.506 +/- 0.010	0.05	0.500 +/- 0.000
0.0005	0.506 +/- 0.010	0.1	0.500 +/- 0.000
0.001	0.503 +/- 0.008	0.5	0.500 +/- 0.000
0.005	0.500 +/- 0.000	1	0.500 +/- 0.000

0.0005 is the value of Minimum gain which maximises the AUC. Therefore, it is the one to be used to control our Decision Tree:



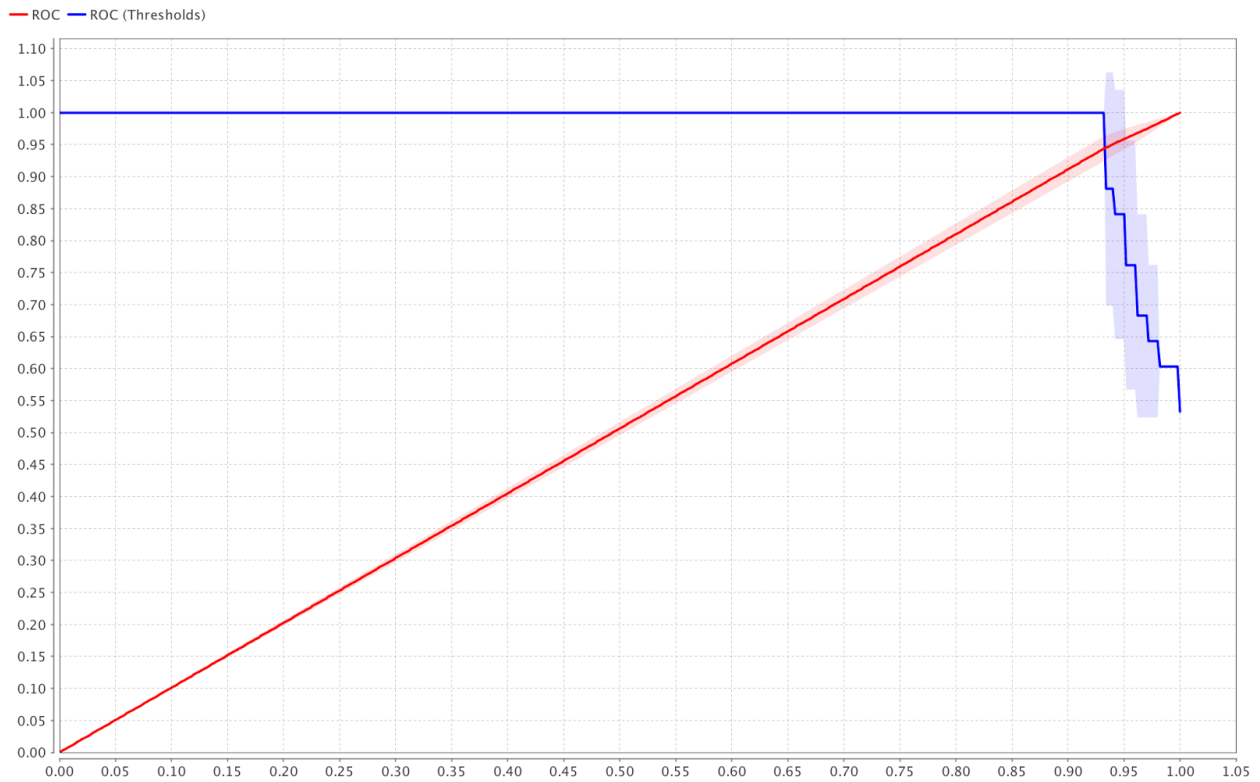
The performance metrics of the model are the following:

- Accuracy: 60.02% +/- 0.13%
- Precision: 60.02% +/- 0.13% (positive class: No)
- Recall: 100.00% +/- 0.00% (positive class: No)
- AUC (optimistic): 0.965 +/- 0.010 (positive class: No)
- AUC: 0.506 +/- 0.010 (positive class: No)
- AUC (pessimistic): 0.047 +/- 0.016 (positive class: No)

With the following confusion matrix:

	true Yes	true No	class precision
pred. Yes	0	0	0.00%
pred. No	1032	1549	60.02%
class recall	0.00%	100.00%	

And with the ROC diagram:



By looking at the Decision Tree we can see that the main (and only) predictor is “country_code” which corresponds with Location in our framework. We assigned a relevance of “Very High” to it, since we consider the data mining prediction with a high impact on our study. Predictors from data constitute the fourth and last piece in the construction of the Viable Framework 1.0. The details can be found in the *Appendix 9. Viable Framework - Step 4*.

However, by analysing the performance metrics and the ROC diagram, we can conclude that the predictive capacity of this model is very limited. We are not able to predict success and this is actually useless for investors. Looking at the tree and its metrics, we can predict that a company will fail 100% of the times and then, in the 60% of the cases, our predictive model will fail. Therefore this model does not work to predict viability of startups on seed stage. The reason might be due to that Crunchbase does not provide enough information to predict at seed stage. It was not easy to determine the metrics of viability as defined in this work. In order to build models

that would predict the viability, success and survival of seed stage startups, firms (investors themselves, corporations or other enterprise services companies) should start building datasets that have enough information on what investors say they need to know, according to our investors survey. Then, exercises similar to the one executed above could be repeated in the future with better results.

It is surprising that even a widely known dataset for research and startup tracking does not provide with the information that it is needed to predict at this stage. The biggest problems detected are data leakage and data quality. It is needed to ensure that data is available in the correct temporal frame in order to predict viability at a certain stage. In this case, for seed stage startups viability prediction, it is needed to have information available on seed stage and those main predictors would be the ones identified during the investors survey. As we saw, they are related to the proper team behind the project and its capacity to lead an effective execution on the right market opportunity.

Dataset 2 - Series6040

In this exercise with series B stage startups we have used a Decision Tree predictive model. We tuned the parameters of our model using 10-fold cross validation to compute out of sample measures of performance and assess the quality of our model.

We have configured the following control parameters:

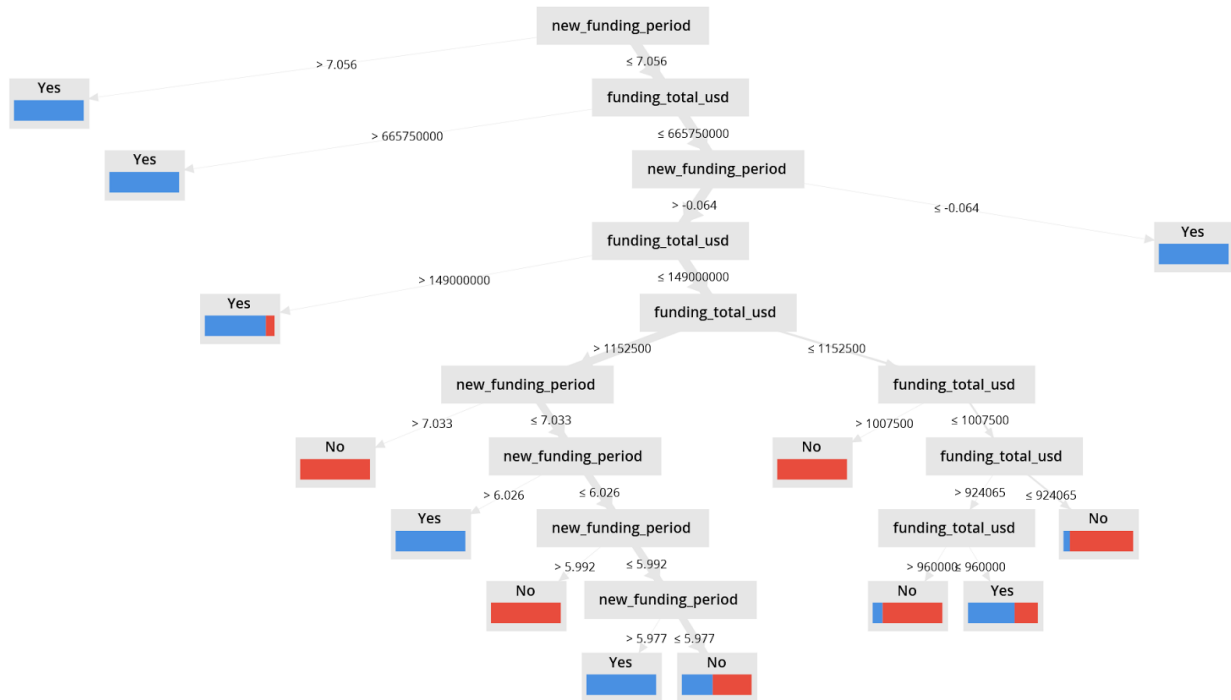
- Criterion: gain_ratio
- Maximal depth: 20
- Apply pruning: yes
- Confidence: 0.25
- Apply prepruning: yes
- Minimal gain: different values
- Minimal leaf size: 2

We have done a sensitivity analysis playing with different values of the parameter Minimal gain and different outcomes of the AUC (ROC):

Minimum gain	AUC	Minimum gain	AUC
0	0.623 +/- 0.040	0.01	0.623 +/- 0.040
0.0001	0.623 +/- 0.040	0.05	0.616 +/- 0.049
0.0005	0.623 +/- 0.040	0.1	0.539 +/- 0.034
0.001	0.623 +/- 0.040	0.5	0.500 +/- 0.000

0.005	0.623 +/- 0.040	1	0.500 +/- 0.000
-------	-----------------	---	-----------------

0.01 is the value of Minimum gain which maximises the AUC. Therefore, it is the one to be used to control our Decision Tree:



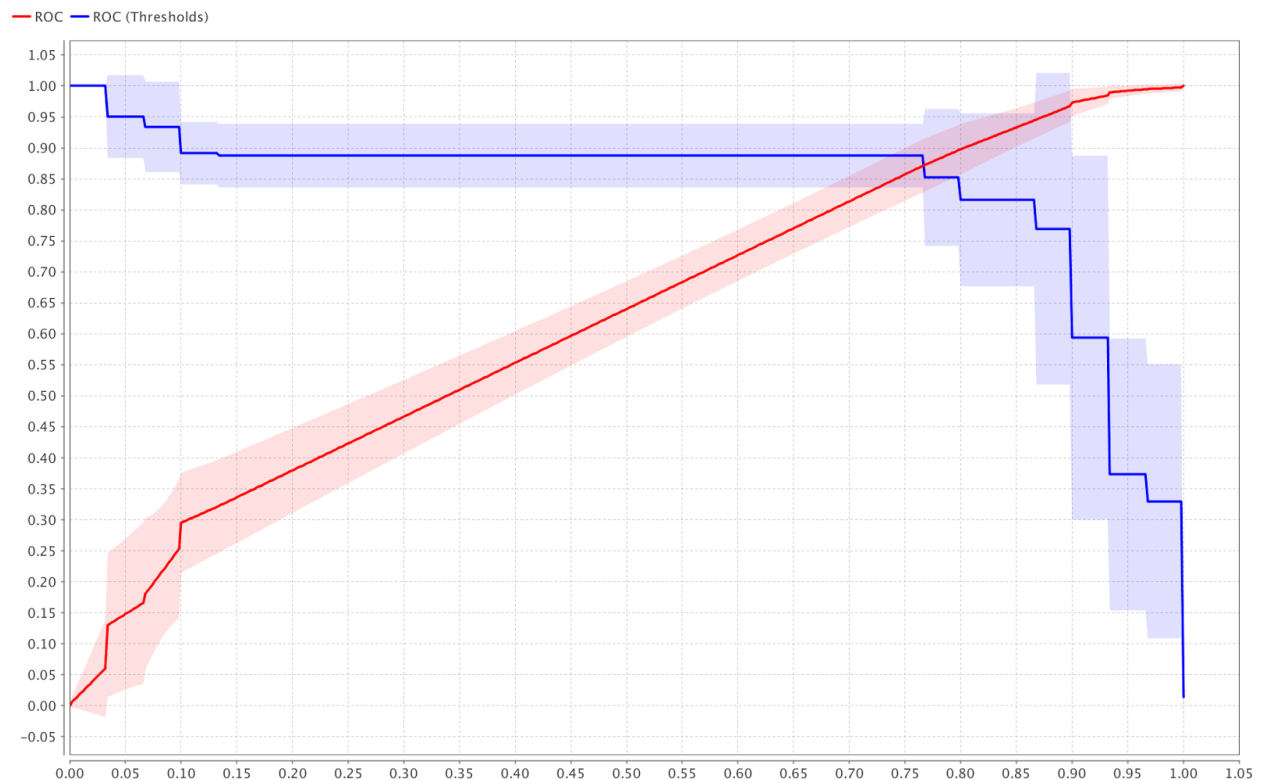
The performance metrics of the model are the following:

- Accuracy: 63.07% +/- 1.89%
- Precision: 62.27% +/- 1.33% (positive class: No)
- Recall: 97.78% +/- 1.99% (positive class: No)
- AUC (optimistic): 0.924 +/- 0.029 (positive class: No)
- AUC: 0.623 +/- 0.040 (positive class: No)
- AUC (pessimistic): 0.326 +/- 0.072 (positive class: No)

With the following confusion matrix:

	true Yes	true No	class precision
pred. Yes	33	10	76.74%
pred. No	267	440	62.23%
class recall	11.00%	97.78%	

And with the ROC diagram:



By looking at the Decision Tree, we can see that the following are the main predictors:

- `New_funding_period`, this factor corresponds with “Funding period” in our framework. It was never present, so we have included this factor in the framework and we assigned a relevance of “Very High” to it, since we consider the data mining prediction with a high impact on our study
- `Funding_total_usd`, this factor corresponds with “Capitalization” in our framework. We assigned a relevance of “Very High” to it, since we consider the data mining prediction with a high impact on our study

In this case, there was more available information about the startups on series B stage and we obtained better results on our predictive model. We can predict if a startup will be viable or not depending on different values for the factors indicated above. In addition, the performance metrics are slightly better than in the exercise with seed stage startups. These facts reinforce our message that we would be able to produce better predictive models if we are able to build better datasets and have access to more accurate information.

Viable Framework 1.0, evaluating viability of startups by using a data based framework



Predictors from data constitute the fourth and last piece in the construction of the Viable Framework 1.0. The details can be found in the *Appendix 9. Viable Framework - Step 4.*

Presentation of conclusions

The original goal of this Master Thesis was to create a model able to predict the success of early stage startups. With the development of the work we detected that the original goal was not feasible and the study evolved to the design of a data based framework able to assess the viability of early stage startups. This assessment is based on an analytical evaluation of present factors that could determine the chances of surviving and reaching success in the future. Therefore, our framework would facilitate the decision making process, for investors (but also corporations) when they have to decide if invest or not, and for entrepreneurs when they have to decide if continue pursuing the venture or not. The framework has been called Viable Framework 1.0.

Viable Framework 1.0 has been fed by different inputs: literature review, investors survey and data mining. In order to build the framework, we followed a methodology structured in four steps. The first step was devoted to theory development, the definition of viability and the identification of potential determinants of viability through an extensive and comprehensive literature review. We were able to collect 194 factors that could determine viability, success and survival according to our definition. We categorised them into 25 areas and 6 dimensions (Concept, People, Context, Execution, Financial and Legal). The second step was dedicated to shortlist the factors identified in the step one according to its relevance, a metric that was based on the number of appearances of a factor in the literature review. We came up with a list of 60 factors, the top ten most relevant ones for each dimension. The third step was devoted to the elaboration and conduction of a survey for investors with the objective of detecting the most important factors from the investor perspective, which means from the practical point of view. We compared the results obtained from both, theoretical and practical, approaches. We saw a general alignment between both except for the factors related to legal aspects. The fourth step was dedicated to the design and execution of a data mining and machine learning exercise. In the experiment we used two different datasets, taken from the Crunchbase database, with information related to startups on seed stage and startups on series B stage. First, we saw that the factors of viability previously identified in the literature review and the investors survey were not present in the Crunchbase database. In other words, Crunchbase does not offer the desired information for the investment stages that we were considering. Then, during the exercise with seed stage startups we were not able to produce any valid predictive model for investors. However, with the exercise with series B startups we managed to produce a predictive model. Although the performance metrics showed limited predictive capacity, accuracy and precision, the most important discovery was that with more and better information we would be able to produce better predictive models. This fact reinforces our premise for the need of building better datasets, based on more accurate

and updated information, that allow us to predict viability by looking at data from different investment stages. This is why we considered Viable Framework 1.0 as the initial version of a more comprehensive framework to be built in the future. The further versions of the Viable Framework and should analyse in deep these detected limitations and contradictions. This future research would have their foundations on building our own datasets, collecting the information that is really important to investors (according to the literature review and the investors survey). With these new datasets we would design new data mining exercises, similar to the ones executed in the present work, that would try to predict the startup success.

Along the study we concluded that **the theory and practice meet on the need of strong team capable to lead a brilliant execution**. We detected a general alignment among all the inputs analysed. We identified the need of a complete team with the proper background and experience, and presenting the right determination to pursue the venture. In addition, the ability to design, plan and execute an scalable business model driving to sales is key to produce a viable startup. And moreover, other factors such as location, product / market fit and capitalization complement the formula of success.

The final step of our Master Thesis is to present the Viable Framework 1.0, which is composed by a set of 20 factors, categorised into 8 areas and 4 dimensions:

Dimension	Area	Factor	Relevance in the Literature Review	Relevance in the Survey (Dimensions)	Relevance in the Survey (Factors)	Relevance in the Data	Total Relevance
Concept	General	Location	Medium	High	Medium	Very High	High
	Product	Product/Market fit	High	High	Very High	Very Low	High
People	Founders	Entrepreneurial experience	Very High	Very High	High	Very Low	High
		Professional experience	Very High	Very High	High	Very Low	High
		Industry experience	Very High	Very High	High	Very Low	High
		Network	High	Very High	High	Very Low	High
		Reputation	High	Very High	High	Very Low	High
		Motivation	High	Very High	Very High	Very Low	High
		Adaptability	High	Very High	Very High	Very Low	High
	Team	Completeness	High	Very High	High	Very Low	High
		Commitment	High	Very High	Very High	Very Low	High

Execution	Strategy	Business model	Very High	Very High	High	Very Low	High
		Scalability	High	Very High	Very High	Very Low	High
	Plan	Business plan	High	Very High	High	Very Low	High
		Marketing plan	High	Very High	High	Very Low	High
		Momentum	High	Very High	High	Very Low	High
	Sales	Channels	High	Very High	High	Very Low	High
		Clients	High	Very High	High	Very Low	High
		Users	High	Very High	High	Very Low	High
	Financial	Capital	Capitalization	High	Medium	Medium	Very High

In order to generate the previous table, we calculated the total relevance of the factors based on the average of their partial relevances obtained in the aforementioned steps. We computed the total relevance by assigning a numeric value to the different qualitative values (4 - Very High, 3 - High, 2 - Medium, 1 - Low and 0 - Very Low), we computed the arithmetic mean and then we assigned back the qualitative values to the results. Finally, we selected the factors with a Total Relevance equal or above 3 (High or Very High). The detailed calculations to come up with the final list of dimensions, areas and factors can be found on the *Appendix 10. Viable Framework - Total*.

TO BE CONTINUED... In the meantime, the application and evolution of the Viable Framework 1.0 can be followed at the entrepreneurial project <http://vialereport.com/>.

References

- [1] The World Bank; 2015. [*Small and Medium Enterprises \(SMEs\) Finance*](#). The World Bank. The World Bank, Washington, D.C., U.S.A.
- [2] Georgiev, G.; 2015. [*What the startup world looked like in 2015*](#). Medium.com. The Obvious Corporation, San Francisco, California, U.S.A.
- [3] Gompers, P.E.; 1995. [*Optimal Investment, Monitoring, and the Staging of Venture Capital*](#). The Journal Of Finance. Wiley-Blackwell for the American Finance Association, Hoboken, New Jersey, U.S.A.
- [4] Tobak, S.; 2014. [*9 Reasons Why Most Startups Fail*](#). Entrepreneur. Entrepreneur Media, Inc., Irvine, California, U.S.A.
- [5] Griffith, E.; 2017. [*Conventional Wisdom Says 90% of Startups Fail. Data Says Otherwise*](#). Fortune. Time Inc., New York City, New York, U.S.A.
- [6] Patel, N.; 2015. [*90% Of Startups Fail: Here's What You Need To Know About The 10%*](#). Forbes.com. Forbes Media, Jersey City, New Jersey, U.S.A.
- [7] Furr, N. and Ahlstrom, P.; 2011. [*Nail it then scale it: the entrepreneur's guide to creating and managing breakthrough innovation*](#). CIMMYT. NISI Institute, Lexington, Kentucky, U.S.A.
- [8] Smith, J.K., Smith R.L. and Bliss, R.T.; 2011. *Entrepreneurial Finance*. Stanford University Press. Stanford University Press, Stanford, California, U.S.A.
- [9] Martínez, R.; 2015. [*9 Tips to Cold-Call VCs Like a Pro*](#). Medium.com. The Obvious Corporation, San Francisco, California, U.S.A.
- [10] Roth, M.; 2015. [*Unicorns and the next Venture Capital Crash*](#). Bison.co. Bison, Boston, Massachusetts, U.S.A.
- [11] 1000 Angels (A Onvest Company); 2016. [*The million dollar question... which startups are going to make you money?*](#). 1000 Angels (A Onvest Company). 1000 Angels (A Onvest Company), New York, New York, U.S.A.
- [12] Villalobos, L.; 2007. [*Investment Valuations of Seed- and Early-Stage Ventures*](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.
- [13] Lipper, G.; 2007. [*Is Valuation a Key Issue in Funding Startups?*](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.
- [14] Villalobos, L. and Payne, W. H.; 2007. [*Startup Pre-money Valuation: The Keystone to Return on Investment*](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.

- [15] Parnell, B.A.; 2017. [*More Corporations Are Looking to Corporate Venture Capital to Access Innovative Startup Technologies*](#). RocketSpace. RocketSpace, Inc., San Francisco, California, U.S.A.
- [16] Alsever, J.; 2017. [*Attention Startups: Big Corporations Want to Work With You*](#). Inc.com. Mansueto Ventures, New York City, New York, U.S.A.
- [17] Yoon, E. and Hughes, S.; 2016. [*Big Companies Should Collaborate with Startups*](#). Harvard Business Review. Harvard Business Publishing, Brighton, Massachusetts, U.S.A.
- [18] Goldstein, M.; 2017. [*The Lost Partnership: How and Why Startups, Corporations Can Work Together*](#). Wired. Condé Nast Publications, San Francisco, California, U.S.A.
- [19] Griffith, E.; 2016. [*2016 Is the Year Startups and Corporates Became BFFs*](#). Fortune. Time Inc., New York City, New York, U.S.A.
- [20] Menkveld, D.J.; 2012. [*Determinants among the Internet Startup Life Cycle*](#). Faculty of Science. Department of Information and Computing Sciences. Utrecht University. Utrecht University, Utrecht, The Netherlands
- [21] [*Cambridge Dictionary*](#)
- [22] Shane, S.A.; 2008. *The Illusions Of Entrepreneurship*. Yale University Press. Yale University Press, New Haven, Connecticut, U.S.A.
- [23] Punset, E.; 2004. *Adaptarse a la Marea*. Espasa Libros. Grupo Planeta, Barcelona, Spain
- [24] Watt, J.; 2016. *Business for Punks*. Portfolio Penguin. Penguin Books Ltd., London, U.K.
- [25] Fortune; 2016. [*The Unicorn List*](#). Fortune. Time Inc., New York City, New York, U.S.A.
- [26] Pimentel, A.; 2016. [*Em vez de unicórnios, transformem-se em baratas. Sobrevivam à guerra nuclear*](#). Observador. Observador On Time, Lisboa, Portugal
- [27] Draper, A.; 2015. [*Be the Cockroach*](#). Medium.com. The Obvious Corporation, San Francisco, California, U.S.A.
- [28] Statman, M.; 2013. *Investor, Know Yourself*. Wall Street Journal. Dow Jones & Company, Inc., New York City, New York, U.S.A.
- [29] McCormack, M.H.; 1984. *What they don't teach you at Harvard Business School*. HarperCollins Publishers Ltd. HarperCollins Publishers Ltd, Glasgow, U.K.
- [30] Punset, E.; 2005. *El viaje a la felicidad: las nuevas claves científicas*. Destino. Grupo Planeta, Barcelona, Spain
- [31] Reddy, S.; 2017. [*A computer was asked to predict which start-ups would be successful. The results were astonishing*](#). World Economic Forum. World Economic Forum, Cologny, Switzerland
- [32] Payne, W. H.; 2007. [*Fundability and Valuation of Startups: An Angel's Perspective*](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.

- [33] Payne, W. H.; 2007. [Valuation Worksheet](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.
- [34] Payne, W. H.; 2007. [Valuation of Pre-revenue Companies: The Venture Capital Method](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.
- [35] Villalobos, L.; 2007. [Valuation Divergence](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.
- [36] Payne, W. H.; 2007. [Valuation Estimator](#). Kauffman eVenturing. Ewing Marion Kauffman Foundation, Kansas City, Missouri, U.S.A.
- [37] Palma, W.; 2016. *Fundraising Tips: An Investor's Perspective*. Lisbon Investment Summit 2016. Beta-i, Lisbon, Portugal
- [38] Marion, T. J.; 2016. [4 Factors That Predict Startup Success, and One That Doesn't](#). Harvard Business Review - Founders. Harvard Business School Publishing, Cambridge, Massachusetts, U.S.A.
- [39] Shalman, W.A.; 1996. *Some thoughts on Business Plans*. Harvard Business School. Harvard Business School Publishing, Cambridge, Massachusetts, U.S.A.
- [40] 1000 Angels (A Onvest Company); 2016. [Signaling: What to look for and why it's important](#). 1000 Angels (A Onvest Company). 1000 Angels (A Onvest Company), New York, New York, U.S.A.
- [41] Brownlee J.; 2016. [Data Leakage in Machine Learning](#). Machine Learning Mastery. Machine Learning Mastery Pty Ltd., Vermont, Victoria, Australia
- [42] Fincham J.E.; 2008. [Response Rates and Responsiveness for Surveys, Standards, and the Journal](#). American Journal of Pharmaceutical Education. National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda, Maryland, U.S.A.
- [43] Provost F. and Fawcett T.; 2013. *Data Science for Business*. O'Reilly Media. O'Reilly Media, Inc., Sebastopol, California, U.S.A.
- [44] RapidMiner Documentation. [Decision Tree](#). RapidMiner Studio 7.6.000 Free Edition for Students. RapidMiner, Inc., Boston, Massachusetts, U.S.A.
- [45] Joshi, R.; 2016. [Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures](#). Exsilio Blog. Exsilio, Inc., Irvine, California, U.S.A.

Other Sources Reviewed

- Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data*. John Murray, London. U.K.
- Gompers, P.A. and Lerner, J. (2004). [*The Venture Capital Cycle*](#). The MIT Press, Cambridge, Massachusetts, U.S.A.
- Thomson, A. (2005). [*Entrepreneurship and Business Innovation*](#). Guildford, W.A. / Vineyard Publishing.
- Van Gelderen, M., Thurik, R. and Bosma, N. (2005). [*Success and Risk factors in the Pre-Startup Phase*](#), Marco van Gelderen. Small Business Economics, Massey University Institutional Repository. Springer International Publishing AG, Cham, Switzerland.
- Baum, J.A.C. and Silverman, B.S. (2004). [*Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups*](#). Journal of Business Venturing. Elsevier Inc., Amsterdam, Netherlands.
- GP Bullhound (2016). [*Europe's Tech Unicorns show sustained growth*](#). GP Bullhound. GP Bullhound, London, U.K.
- Scafton, L. (2014). [*What is a tech accelerator – and which are the best in Europe?*](#). The Guardian. The Guardian, London, U.K.
- Stuart R. and Abetti P.A. (1987). [*Start-up Ventures Towards The Prediction of Initial Success*](#). Journal of Business Venturing. Elsevier Inc., Amsterdam, Netherlands.
- Whitley, A. (2016). [*Hollywood turning to big data to write the next blockbuster*](#). Daily Herald. Paddock Publications, Inc., Arlington Heights, Illinois, U.S.A.
- April, L. (2016). [*Big Data Can Open Our Eyes to Talent*](#). Business 2 Community. Insider News Group, Malvern, Pennsylvania, U.S.A.
- Weisul, K. (2014). [*In Crowdfunding, All-or-Nothing Campaigns Are More Successful*](#). Inc. Mansueto Ventures LLC, New York, New York, U.S.A.
- Kellie, L. (2016). [*Study uses text-mining to improve market intelligence on startups*](#). EurekAlert!. American Association for the Advancement of Science (AAAS), Washington, DC, U.S.A.
- Fontana, A. and Gorenberg, M. (2016). [*Growing up in the intelligence era*](#). TechCrunch. AOL Inc., New York City, New York, U.S.A.
- Kamps, H. J. (2016). [*This AI-powered VC is smarter than your local startup cash dispensary*](#). TechCrunch. AOL Inc., New York City, New York, U.S.A.

Appendices

Appendix 1. Definition of Viability

Meaning	Source
Reaching objectives	The Illusions Of Entrepreneurship [23]
Durability (survival)	Entrepreneurial Finance [5]
Profitability (performance: sales & profits)	The Illusions Of Entrepreneurship [23]
(Continuous) Growth (sales, profits, capital / assets, employees and net value / worth / wealth)	The Illusions Of Entrepreneurship [23]
Certainty (reducing risk)	Entrepreneurial Finance [5]
(External) Investment	Entrepreneurial Finance [5]
Continuing to operate the venture	Entrepreneurial Finance [5]
Management Buyout (MBO)	Entrepreneurial Finance [5]
Sale of the business to employees / other members of the team	Entrepreneurial Finance [5]
Private sale of the venture to another firm (M&A)	Entrepreneurial Finance [5]
Going public (IPO)	Entrepreneurial Finance [5]
Closing business voluntarily (entrepreneurs & creditors don't lose their money)	The Illusions Of Entrepreneurship [23]
Contribution to society	The Illusions Of Entrepreneurship [23]
Survival is always the first step to success	Business for Punks [24]
Growth (Sales revenue, Net income, Cash flow available to investors)	Entrepreneurial Finance [5]
Meeting the objectives	Entrepreneurial Finance [5]
Lack of failure	Entrepreneurial Finance [5]
Lack of bankruptcy	Entrepreneurial Finance [5]
Success depends on building a competitive advantage that is sustainable	Entrepreneurial Finance [5]
The lifeblood of your business is cash. if you can't manage a cash flow, then you can't run a business	Business for Punks [24]
Success is the balance of both adaptation strategies: cost reduction and niche search	Adaptarse a la Marea [22]

Appendix 2. Viabale Framework - Step 1

Dimension	Area	Factor	Number of appearances in the Literature Review	Relevance in the Literature Review
Concept	General	Mission	4	High
		Vision	2	Low
		Social impact	2	Low
		Opportunities	1	Very Low
		Trendiness	1	Very Low
		Innovation	1	Very Low
		Location	3	Medium
	Product	Product vs. Service	0	Very Low
		Problem/Solution fit	3	Medium
		Product/Market fit	5	High
People	Founders	Age	4	High
		Gender	3	Medium
		Race	1	Very Low
		Academic level	9	High
		Relevant university	3	Medium
		Entrepreneurial experience	11	Very High

	Past (entrepreneurial) success / Exits	3	Medium
	Professional experience	13	Very High
	Industry experience	14	Very High
	Relevant company	3	Medium
	Network	8	High
	Reputation	5	High
	Influence	1	Very Low
	Motivation	9	High
	Passion	1	Very Low
	Opportunity cost of human capital	1	Very Low
	Wealth	1	Very Low
	Risk tolerance	1	Very Low
	Persistence	1	Very Low
	Resilience	1	Very Low
	Determination	2	Low
	Flexibility	3	Medium
	Coachability	3	Medium
	Capacity to be monitored	1	Very Low
	Adaptability	7	High
	Attitude	5	High
	Awareness / Sensibility	2	Low

		Part of the target demographic	1	Very Low
	Team	Structure	4	High
		Staffing	2	Low
		Staff retention	1	Very Low
		Staff turnover	1	Very Low
		Staff happiness	1	Very Low
		Labour ratios	1	Very Low
		Completeness	5	High
		Complementarity	3	Medium
		Commitment	8	High
		Team/Project fit	1	Very Low
		Advisors	1	Very Low
		Context	Environment	Political
Economical	4			High
Social	1			Very Low
Technological	3			Medium
Market	Go-to-market strategy		7	High
	Industry		5	High
	Geography		2	Low
	Size		9	High
	Growth		7	High

		Share	3	Medium	
		Barriers to entry	5	High	
		Focus	3	Medium	
	Competitors	Structure	7	High	
		Strength	7	High	
		Competitive advantage	5	High	
		Product differentiation	6	High	
		Defensibility	5	High	
	Partners	Structure	10	Very High	
		Incubator	5	High	
		Accelerator	5	High	
		Advisors	5	High	
		Media	3	Medium	
	Execution	Situation	Stage	6	High
			Technology readiness level	2	Low
Milestones			2	Low	
Time of operation			1	Very Low	
Size of the company			2	Low	
Tools			2	Low	
Uncertainty			1	Very Low	
Awards			1	Very Low	

	Strategy	Origin	1	Very Low
		Business model	10	Very High
		Supply / Demand	1	Very Low
		Scalability	6	High
		Risks/Contingency plan	2	Low
	Plan	Business plan	7	High
		Operations plan (Production, Sales...)	1	Very Low
		Marketing plan	6	High
		Financial plan	3	Medium
		Momentum	5	High
		Roadmap	1	Very Low
		Motion vs. Accomplishment	1	Very Low
	Operations	Suppliers	1	Very Low
		Development Lead Time	1	Very Low
		Shipment accuracy	1	Very Low
		Refunds	1	Very Low
		Wastage	1	Very Low
	Traction	Channels	4	High
		Clients	6	High
		Client satisfaction	2	Low
Client complaints		1	Very Low	

		Users	5	High	
		User satisfaction	2	Low	
		User complaints	1	Very Low	
		Referrals	1	Very Low	
		Additional customers	1	Very Low	
		Conversion rate	1	Very Low	
		Churn rate	1	Very Low	
	Online	Online engagement	0	Very Low	
		Alexa rank	0	Very Low	
		Facebook fans	0	Very Low	
		Twitter followers	1	Very Low	
		Google+ visits	0	Very Low	
	Financial	Revenues	Sales	1	Very Low
			Current	3	Medium
			Last 12 months	2	Low
Next 6 months			3	Medium	
Next 12 months			4	High	
Next 2 years			3	Medium	
Next 3 years			3	Medium	
In 5 years			4	High	
Expected growth			2	Low	

	Expenses	Cost of sales	1	Very Low
		Overheads	1	Very Low
		Current	1	Very Low
		Last 12 months	1	Very Low
		Next 6 months	1	Very Low
		Next 12 months	1	Very Low
		Next 2 years	1	Very Low
		Next 3 years	1	Very Low
		In 5 years	1	Very Low
		Cost of Production & Delivery	3	Medium
		Burn rate	3	Medium
		Doability quotient	1	Very Low
	Profits	EBITDA	1	Very Low
		Net profit	1	Very Low
		Current	1	Very Low
		Last 12 months	1	Very Low
		Next 6 months	1	Very Low
		Next 12 months	1	Very Low
		Next 2 years	1	Very Low
		Next 3 years	1	Very Low
In 5 years	1	Very Low		

	Cash	Cash flows	3	Medium
		Operating cash flows	1	Very Low
		Investing cash flows	0	Very Low
		Financing cash flows	0	Very Low
		Marginal gross margins	2	Low
		Marginal net margins	1	Very Low
	Capital	Working Capital	1	Very Low
		Efficiency	2	Low
		Capitalization	4	High
		Assets	3	Medium
		Equity	1	Very Low
		Debt	1	Very Low
		Creditor covenants	1	Very Low
		Budget allocated to R&D	1	Very Low
		Valuation of other similar startups	2	Low
		Valuation (NPV)	2	Low
		Tax status	1	Very Low
		Exchange rates	1	Very Low
		Investment	Breakeven point	2
	Time invested		1	Very Low
	Own amount invested		2	Low

		External amount invested (Track record / Cap table)	2	Low
		Previous external investors (Track record / Cap table)	5	High
		Previous external investors interested in investing now	5	High
		Credible external investors interested in investing now	6	High
		Amount required	6	High
		Purpose / Justification	2	Low
		Nature of the financing need	2	Low
		Urgency of the financing need	2	Low
		Round	2	Low
		Funding period	0	Very Low
	Metrics	Risk	1	Very Low
		Reward / Risk	1	Very Low
		CAC (Acquisition)	3	Medium
		CRC (Retention)	1	Very Low
		LTV	1	Very Low
		Average spend per transaction (Cost to produce and deliver the product or service)	2	Low
		Like-for-like sales growth	1	Very Low
		Sales-to-rent ratios	1	Very Low
		Cost to support a customer	1	Very Low
Capital required to support a dollar of sales		1	Very Low	

Legal	Incorporation	Form	3	Medium
	Deal	Terms & Conditions	1	Very Low
		Incentive effects of the allocation (risk vs. reward)	1	Very Low
		Simplicity	1	Very Low
		Fairness	1	Very Low
		Robustness	1	Very Low
		Lack of perverse incentives	1	Very Low
		Lack of foreclose valuable options	1	Very Low
		IP	Status	6
	Attorney	Lawyers firm	1	Very Low
		Intellectual property firm	2	Low

Appendix 3. Viable Framework - Step 2

Dimension	Area	Factor	Number of appearances in the Literature Review	Relevance in the Literature Review
Concept	General	Mission	4	High
		Vision	2	Low
		Social impact	2	Low
		Opportunities	1	Very Low
		Trendiness	1	Very Low
		Innovation	1	Very Low
	Location	3	Medium	
	Product	Product vs. Service	0	Very Low
		Problem/Solution fit	3	Medium
Product/Market fit		5	High	
People	Founders	Academic level	9	High
		Entrepreneurial experience	11	Very High
		Professional experience	13	Very High
		Industry experience	14	Very High
		Network	8	High
		Reputation	5	High
		Motivation	9	High
	Adaptability	7	High	
	Team	Completeness	5	High
		Commitment	8	High
Context	Environment	Economical	4	High
	Market	Go-to-market strategy	7	High
		Industry	5	High
		Size	9	High
		Growth	7	High
		Barriers to entry	5	High
	Competitors	Structure	7	High

		Strength	7	High	
		Product differentiation	6	High	
	Partners	Structure	10	Very High	
Execution	Situation	Stage	6	High	
	Strategy	Business model	10	Very High	
		Scalability	6	High	
	Plan	Business plan	7	High	
		Marketing plan	6	High	
		Financial plan	3	Medium	
		Momentum	5	High	
	Sales	Channels	4	High	
		Clients	6	High	
Users		5	High		
Financial	Revenues	Current	3	Medium	
		Next 12 months	4	High	
		In 5 years	4	High	
	Expenses	Burn rate	3	Medium	
	Cash	Cash flows	3	Medium	
	Capital	Capitalization	4	High	
	Investment	Previous external investors (Track record / Cap table)		5	High
		Previous external investors interested in investing now		5	High
		Credible external investors interested in investing now		6	High
Amount required		6	High		
Legal	Incorporation	Form	3	Medium	
	Deal	Terms & Conditions		1	Very Low
		Incentive effects of the allocation (risk vs. reward)		1	Very Low
		Simplicity		1	Very Low
		Fairness		1	Very Low
		Robustness		1	Very Low
		Lack of perverse incentives		1	Very Low
		Lack of foreclose valuable options		1	Very Low
	IP	Status	6	High	

Viable Framework 1.0, evaluating viability of startups by using a data based framework



	Attorney	Intellectual property firm	2	Low
--	----------	----------------------------	---	-----

Appendix 4. Investors survey - Viability from the Investor perspective

The link with the survey can be found at [Master Thesis - Viability from the Investor perspective](#).

The survey is divided into 2 sections:

- General Information
- Viability Factors

The section General Information contains the following questions and possible answers:

Investor profile

Answer	Type
Business Angel	Multiple Choice
Venture Capital	Multiple Choice
Corporation	Multiple Choice
Other	Multiple Choice

Stage of investment

Answer	Type
Pre-Seed	Multiple Choice
Seed	Multiple Choice
Series	Multiple Choice
Other	Multiple Choice

Geographical area

Answer	Type
West Europe	Multiple Choice
East Europe	Multiple Choice
Middle East	Multiple Choice
Asia	Multiple Choice
North America	Multiple Choice
South America	Multiple Choice



Oceania	Multiple Choice
Other	Multiple Choice

Industry vertical

Answer	Type
Automotive	Multiple Choice
Banking	Multiple Choice
Consumer	Multiple Choice
Education	Multiple Choice
Engineering	Multiple Choice
Energy	Multiple Choice
Fast-Moving Consumer Goods	Multiple Choice
Financial	Multiple Choice
Food and Beverage	Multiple Choice
Government	Multiple Choice
Healthcare	Multiple Choice
Insurance	Multiple Choice
Legal	Multiple Choice
Manufacturing	Multiple Choice
Media	Multiple Choice
Non-For Profit	Multiple Choice
Online	Multiple Choice
Real Estate	Multiple Choice
Religion	Multiple Choice
Retail	Multiple Choice
Technology	Multiple Choice
Telecommunications	Multiple Choice
Tourism	Multiple Choice
Transportation	Multiple Choice
Other	Multiple Choice

Your definition of viability has to do with...

Answer	Type
Reaching objectives	Multiple Choice
Durability	Multiple Choice
Profitability (performance: sales & profits)	Multiple Choice
(Continuous) Growth (sales, profits, capital / assets, employees and net value / worth / wealth)	Multiple Choice
Certainty (reducing risk)	Multiple Choice
(External) Investment	Multiple Choice
Continuing to operate the venture	Multiple Choice
Management Buyout (MBO)	Multiple Choice
Sale of the business to employees / other members of the team	Multiple Choice
Private sale of the venture to another firm (M&A)	Multiple Choice
Going public (IPO)	Multiple Choice
Closing business voluntarily (entrepreneurs & creditors don't lose their money)	Multiple Choice
Contribution to society	Multiple Choice
Survival is always the first step to success	Multiple Choice
Sales revenue, Net income, Cash flow available to investors	Multiple Choice
Lack of failure	Multiple Choice
Lack of bankruptcy	Multiple Choice
Success depends on building a competitive advantage that is sustainable	Multiple Choice
The lifeblood of your business is cash. if you can't manage a cash flow, then you can't run a business	Multiple Choice
Success is the balance of both adaptation strategies: cost reduction and niche search	Multiple Choice

The section Viability Factors contains the following questions and possible answers:

Define the importance of the following dimensions (sets of criteria) when evaluating an opportunity

Answer	Type
Concept (Idea, Product...)	Very Low, Low, Medium, High, Very High
People	Very Low, Low, Medium, High, Very High
Context (Market, Competition...)	Very Low, Low, Medium, High, Very High
Execution	Very Low, Low, Medium, High, Very High
Financial	Very Low, Low, Medium, High, Very High
Legal	Very Low, Low, Medium, High, Very High
Other	Very Low, Low, Medium, High, Very High

In terms of CONCEPT, define the importance of the following factors

Answer	Type
Location	Very Low, Low, Medium, High, Very High
Mission	Very Low, Low, Medium, High, Very High
Vision	Very Low, Low, Medium, High, Very High
Social impact	Very Low, Low, Medium, High, Very High
Opportunity	Very Low, Low, Medium, High, Very High
Trendiness	Very Low, Low, Medium, High, Very High
Innovation	Very Low, Low, Medium, High, Very High
Product vs. Service	Very Low, Low, Medium, High, Very High
Problem/Solution fit	Very Low, Low, Medium, High, Very High
Product/Market fit	Very Low, Low, Medium, High, Very High

In terms of PEOPLE, define the importance of the following factors

Answer	Type
Academic level	Very Low, Low, Medium, High, Very High
Entrepreneurial experience	Very Low, Low, Medium, High, Very High



Professional experience	Very Low, Low, Medium, High, Very High
Industry experience	Very Low, Low, Medium, High, Very High
Network	Very Low, Low, Medium, High, Very High
Reputation	Very Low, Low, Medium, High, Very High
Motivation	Very Low, Low, Medium, High, Very High
Commitment	Very Low, Low, Medium, High, Very High
Adaptability	Very Low, Low, Medium, High, Very High
Completeness of the team	Very Low, Low, Medium, High, Very High

In terms of CONTEXT, define the importance of the following factors

Answer	Type
Economical environment	Very Low, Low, Medium, High, Very High
Go-to-market strategy	Very Low, Low, Medium, High, Very High
Industry	Very Low, Low, Medium, High, Very High
Market size	Very Low, Low, Medium, High, Very High
Market growth	Very Low, Low, Medium, High, Very High
Barriers to entry	Very Low, Low, Medium, High, Very High
Competitors' structure	Very Low, Low, Medium, High, Very High
Competitors' strength	Very Low, Low, Medium, High, Very High
Product/Service differentiation	Very Low, Low, Medium, High, Very High
Partners' structure	Very Low, Low, Medium, High, Very High

In terms of EXECUTION, define the importance of the following factors

Answer	Type
Development stage	Very Low, Low, Medium, High, Very High
Business model	Very Low, Low, Medium, High, Very High
Scalability	Very Low, Low, Medium, High, Very High
Operations plan	Very Low, Low, Medium, High, Very High
Marketing plan	Very Low, Low, Medium, High, Very High
Financial plan	Very Low, Low, Medium, High, Very High



Momentum	Very Low, Low, Medium, High, Very High
Sales & Distribution channels	Very Low, Low, Medium, High, Very High
Number of clients	Very Low, Low, Medium, High, Very High
Number of users	Very Low, Low, Medium, High, Very High

In terms of FINANCIAL, define the importance of the following factors

Answer	Type
Current revenues	Very Low, Low, Medium, High, Very High
Revenues in the next 12 months	Very Low, Low, Medium, High, Very High
Revenues in 5 years	Very Low, Low, Medium, High, Very High
Burn rate	Very Low, Low, Medium, High, Very High
Cash flows	Very Low, Low, Medium, High, Very High
Capitalization	Very Low, Low, Medium, High, Very High
Previous external investors (Cap table)	Very Low, Low, Medium, High, Very High
Previous external investors interested in investing now	Very Low, Low, Medium, High, Very High
Credible external investors interested in investing now	Very Low, Low, Medium, High, Very High
Amount required	Very Low, Low, Medium, High, Very High

In terms of LEGAL, define the importance of the following factors

Answer	Type
Incorporation form	Very Low, Low, Medium, High, Very High
Terms & Conditions of the deal	Very Low, Low, Medium, High, Very High
Simplicity of the deal	Very Low, Low, Medium, High, Very High
Fairness of the deal	Very Low, Low, Medium, High, Very High
Robustness of the deal	Very Low, Low, Medium, High, Very High
Incentive effects of the allocation risk vs. reward	Very Low, Low, Medium, High, Very High
Lack of perverse incentives	Very Low, Low, Medium, High, Very High



Lack of foreclose valuable options	Very Low, Low, Medium, High, Very High
IP status	Very Low, Low, Medium, High, Very High
Intellectual property firm	Very Low, Low, Medium, High, Very High

Appendix 5. Crunchbase - Details

Name of the database: CrunchBase Excel Export.

Version: v3.22.

Date: 2016-08-31.

Agreement: By using the data contained in this document, you agree to the Data Access License Agreement and Data Access License Terms. You may not license, sublicense, sell, offer to sell, distribute or otherwise provide any CrunchBase data to any third parties without prior written approval from CrunchBase.

Structure:

Table “Funded Companies”

Field	Viability Framework Equivalence
company_name	N/A
domain	N/A
country_code	Location
state_code	Location
region	Location
city	Location
status	Stage
short_description	N/A
category_list	Industry
category_group_list	Industry
employee_count	(Staffing)
funding_rounds	Capitalization
funding_total_usd	Capitalization
founded_on	(Time of operation)
first_funding_on	(Time of operation)
last_funding_on	(Time of operation)
closed_on	(Time of operation)
email	N/A

phone	N/A
cb_url	N/A
twitter_url	N/A
facebook_url	N/A
uuid	N/A

Table “Rounds”

Field	Viabie Framework Equivalence
company_name	N/A
country_code	Location
state_code	Location
region	Location
city	Location
company_category_list	Industry
funding_round_type	Stage
funding_round_code	N/A
announced_on	N/A
raised_amount_usd	Amount required
raised_amount	Capitalization
raised_amount_currency_code	N/A
post_money_valuation_usd	(Valuation of other similar startups)
post_money_valuation	(Valuation of other similar startups)
post_money_currency_code	N/A
investor_count	Previous external investors (Track record / Cap table)
investor_names	Previous external investors (Track record / Cap table)
cb_url	N/A
company_uuid	N/A
funding_round_uuid	N/A
funding_round_uuid	N/A

Table “Acquisitions”

Field	Viabie Framework Equivalence
acquiree_name	N/A
acquiree_country_code	Location
state_code	Location
acquiree_region	Location
acquiree_city	Location
acquirer_name	Credible external investors interested in investing now
acquirer_country_code	N/A
acquirer_state_code	N/A
acquirer_region	N/A
acquirer_city	N/A
acquired_on	Stage
price_usd	(Valuation of other similar startups)
price	(Valuation of other similar startups)
price_currency_code	N/A
acquiree_cb_url	N/A
acquirer_cb_url	N/A
acquiree_uuid	N/A
acquirer_uuid	N/A
acquisition_uuid	N/A

Table “IPOs”

Field	Viabie Framework Equivalence
name	N/A
country_code	Location
company_state_code	Location
region	Location
city	Location



stock_exchange_symbol	N/A
stock_symbol	N/A
went_public_on	Stage
price_usd	(Valuation of other similar startups)
price	(Valuation of other similar startups)
price_currency_code	N/A
cb_url	N/A
ipo_uuid	N/A
uuid	N/A

Appendix 6. Investors survey - Viability from the Investor perspective - Answers

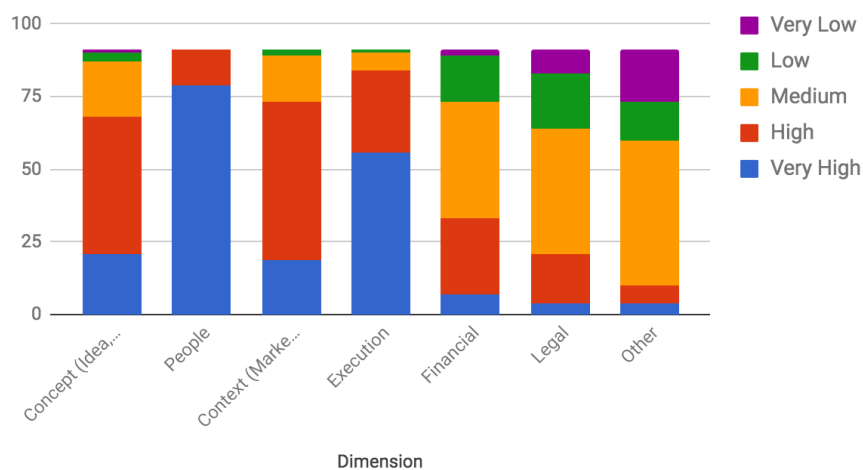
The graphical results of the survey Viability from the Investor perspective can be found [here](#).

The numerical results of the survey Viability from the Investor perspective can be found [here](#).

Importance of Dimensions

Dimension	Very High	High	Medium	Low	Very Low
Concept (Idea, Product...)	21	47	19	3	1
People	79	12	0	0	0
Context (Market, Competition...)	19	54	16	2	0
Execution	56	28	6	1	0
Financial	7	26	40	16	2
Legal	4	17	43	19	8
Other	4	6	50	13	18

Importance of Dimensions

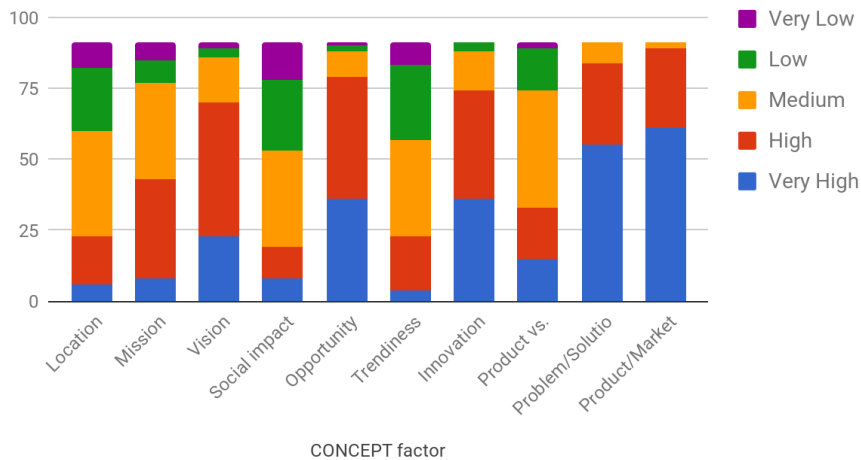


Importance of CONCEPT factors

CONCEPT factor	Very High	High	Medium	Low	Very Low
Location	6	17	37	22	9
Mission	8	35	34	8	6
Vision	23	47	16	3	2

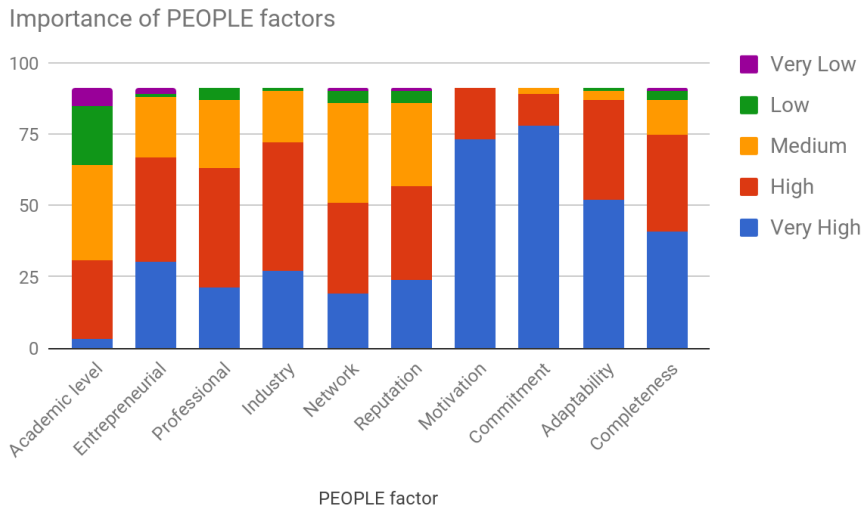
Social impact	8	11	34	25	13
Opportunity	36	43	9	2	1
Trendiness	4	19	34	26	8
Innovation	36	38	14	3	0
Product vs. Service	15	18	41	15	2
Problem/Solution fit	55	29	7	0	0
Product/Market fit	61	28	2	0	0

Importance of CONCEPT factors



Importance of PEOPLE factors

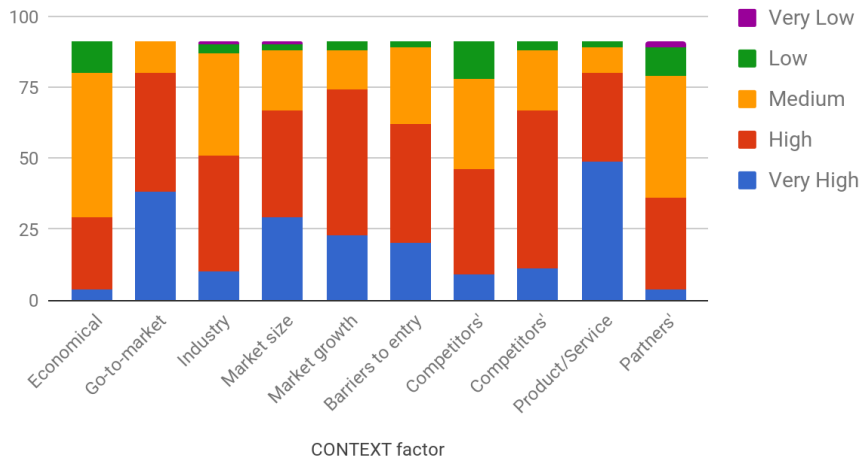
PEOPLE factor	Very High	High	Medium	Low	Very Low
Academic level	3	28	33	21	6
Entrepreneurial experience	30	37	21	1	2
Professional experience	21	42	24	4	0
Industry experience	27	45	18	1	0
Network	19	32	35	4	1
Reputation	24	33	29	4	1
Motivation	73	18	0	0	0
Commitment	78	11	2	0	0
Adaptability	52	35	3	1	0
Completeness of the team	41	34	12	3	1



Importance of CONTEXT factors

CONTEXT factor	Very High	High	Medium	Low	Very Low
Economical environment	4	25	51	11	0
Go-to-market strategy	38	42	11	0	0
Industry	10	41	36	3	1
Market size	29	38	21	2	1
Market growth	23	51	14	3	0
Barriers to entry	20	42	27	2	0
Competitors' structure	9	37	32	13	0
Competitors' strength	11	56	21	3	0
Product/Service differentiation	49	31	9	2	0
Partners' structure	4	32	43	10	2

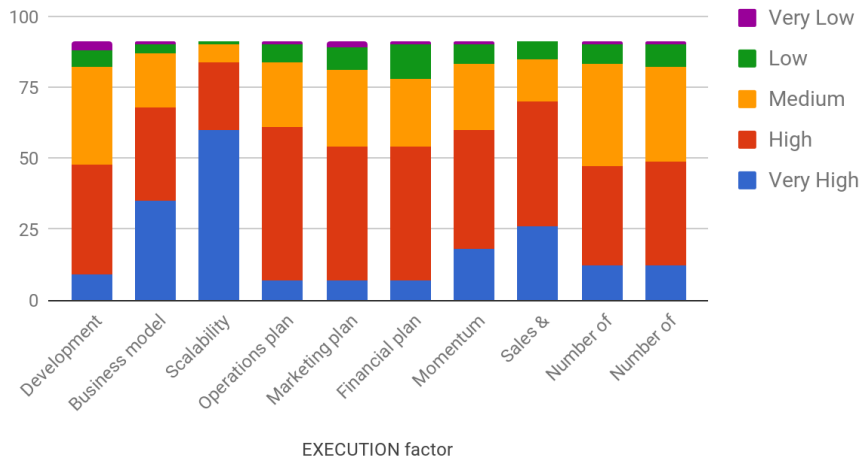
Importance of CONTEXT factors



Importance of EXECUTION factors

EXECUTION factor	Very High	High	Medium	Low	Very Low
Development stage	9	39	34	6	3
Business model	35	33	19	3	1
Scalability	60	24	6	1	0
Operations plan	7	54	23	6	1
Marketing plan	7	47	27	8	2
Financial plan	7	47	24	12	1
Momentum	18	42	23	7	1
Sales & Distribution channels	26	44	15	6	0
Number of clients	12	35	36	7	1
Number of users	12	37	33	8	1

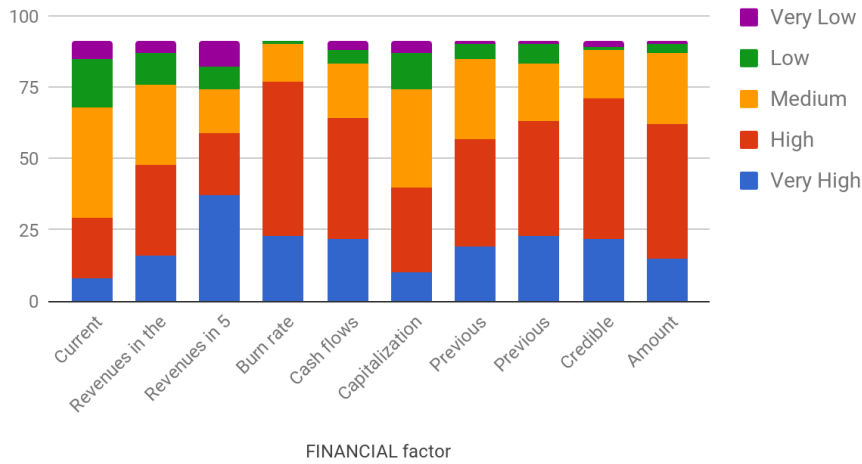
Importance of EXECUTION factors



Importance of FINANCIAL factors

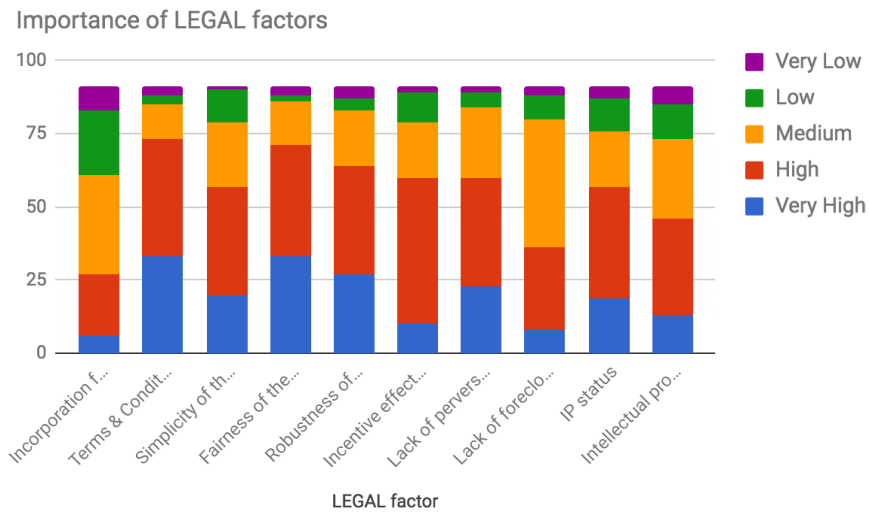
FINANCIAL factor	Very High	High	Medium	Low	Very Low
Current revenues	8	21	39	17	6
Revenues in the next 12 months	16	32	28	11	4
Revenues in 5 years	37	22	15	8	9
Burn rate	23	54	13	1	0
Cash flows	22	42	19	5	3
Capitalization	10	30	34	13	4
Previous external investors (Cap table)	19	38	28	5	1
Previous external investors interested in investing now	23	40	20	7	1
Credible external investors interested in investing now	22	49	17	1	2
Amount required	15	47	25	3	1

Importance of FINANCIAL factors



Importance of LEGAL factors

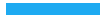
LEGAL factor	Very High	High	Medium	Low	Very Low
Incorporation form	6	21	34	22	8
Terms & Conditions of the deal	33	40	12	3	3
Simplicity of the deal	20	37	22	11	1
Fairness of the deal	33	38	15	2	3
Robustness of the deal	27	37	19	4	4
Incentive effects of the allocation risk vs. reward	10	50	19	10	2
Lack of perverse incentives	23	37	24	5	2
Lack of foreclose valuable options	8	28	44	8	3
IP status	19	38	19	11	4
Intellectual property firm	13	33	27	12	6



Appendix 7. Viable Framework - Step 3

Dimension	Area	Factor	Relevance in the Literature Review	Relevance in the Survey (Dimensions)	Relevance in the Survey (Factors)
Concept	General	Mission	High	High	Medium
		Vision	Low	High	High
		Social impact	Low	High	Medium
		Opportunities	Very Low	High	High
		Trendiness	Very Low	High	Medium
		Innovation	Very Low	High	High
	Location	Medium	High	Medium	
	Product	Product vs. Service	Very Low	High	Medium
		Problem/Solution fit	Medium	High	Very High
Product/Market fit		High	High	Very High	
People	Founders	Academic level	High	Very High	Medium
		Entrepreneurial experience	Very High	Very High	High
		Professional experience	Very High	Very High	High
		Industry experience	Very High	Very High	High
		Network	High	Very High	High
		Reputation	High	Very High	High
		Motivation	High	Very High	Very High
	Adaptability	High	Very High	Very High	
	Team	Completeness	High	Very High	High
		Commitment	High	Very High	Very High
Context	Environment	Economical	High	High	Medium
	Market	Go-to-market strategy	High	High	High
		Industry	High	High	High
		Size	High	High	High
		Growth	High	High	High
		Barriers to entry	High	High	High
	Competitors	Structure	High	High	Medium
Strength		High	High	High	

		Product differentiation	High	High	High	
	Partners	Structure	Very High	High	Medium	
Execution	Situation	Stage	High	Very High	Medium	
	Strategy	Business model	Very High	Very High	High	
		Scalability	High	Very High	Very High	
	Plan	Business plan	High	Very High	High	
		Marketing plan	High	Very High	High	
		Financial plan	Medium	Very High	High	
		Momentum	High	Very High	High	
	Sales	Channels	High	Very High	High	
		Clients	High	Very High	High	
		Users	High	Very High	High	
Financial	Revenues	Current	Medium	Medium	Medium	
		Next 12 months	High	Medium	Medium	
		In 5 years	High	Medium	High	
	Expenses	Burn rate	Medium	Medium	High	
	Cash	Cash flows	Medium	Medium	High	
	Capital	Capitalization	High	Medium	Medium	
	Investment	Previous external investors (Track record / Cap table)		High	Medium	High
		Previous external investors interested in investing now		High	Medium	High
		Credible external investors interested in investing now		High	Medium	High
		Amount required		High	Medium	High
Legal	Incorporation	Form	Medium	Medium	Medium	
	Deal	Terms & Conditions		Very Low	Medium	High
		Incentive effects of the allocation (risk vs. reward)		Very Low	Medium	High
		Simplicity		Very Low	Medium	High
		Fairness		Very Low	Medium	High
		Robustness		Very Low	Medium	High
		Lack of perverse incentives		Very Low	Medium	High
		Lack of foreclose valuable options		Very Low	Medium	Medium



	IP	Status	High	Medium	High
	Attorney	Intellectual property firm	Low	Medium	Medium

Appendix 8. Crunchbase - Definitions

Control Parameters

Concept	Definition
Criterion	<p>Selects the criterion on which Attributes will be selected for splitting. It can have one of the following values:</p> <ul style="list-style-type: none"> - information_gain: The entropies of all the Attributes are calculated and the one with least entropy is selected for split. This method has a bias towards selecting Attributes with a large number of values. - gain_ratio: A variant of information gain that adjusts the information gain for each Attribute to allow the breadth and uniformity of the Attribute values. - gini_index: A measure of inequality between the distributions of label characteristics. Splitting on a chosen Attribute results in a reduction in the average gini index of the resulting subsets. - accuracy: An Attribute is selected for splitting, which maximizes the accuracy of the whole tree.
Maximal depth	<p>The depth of a tree varies depending upon the size and characteristics of the ExampleSet. This parameter is used to restrict the depth of the decision tree. If its value is set to '-1', the maximal depth parameter puts no bound on the depth of the tree. In this case the tree is built until other stopping criteria are met. If its value is set to '1', a tree with a single node is generated.</p>
Apply pruning	<p>The decision tree model can be pruned after generation. If checked, some branches are replaced by leaves according to the confidence parameter.</p>
Confidence	<p>This parameter specifies the confidence level used for the pessimistic error calculation of pruning.</p>
Apply prepruning	<p>This parameter specifies if more stopping criteria than the maximal depth should be used during generation of the decision tree model. If checked, the parameters minimal gain, minimal leaf size, minimal size for split and number of prepruning alternatives are used as stopping criteria.</p>
Minimal gain	<p>The gain of a node is calculated before splitting it. The node is split if its gain is greater than the minimal gain. A higher value of minimal gain results in fewer splits and thus a smaller tree. A value that is too high will completely prevent splitting and a tree with a single node is generated.</p>
Minimal leaf size	<p>The size of a leaf is the number of Examples in its subset. The tree is generated in such a way that every leaf has at least the minimal leaf size number of Examples.</p>

The information above has been extracted from the RapidMiner Documentation on Decision Trees ([44] RapidMiner Documentation).

Performance Metrics

Concept	Definition
Accuracy	Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.
Precision	Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
Recall	Recall is the ratio of correctly predicted positive observations to the all observations in actual class.
ROC	A ROC graph is a two-dimensional plot of a classifier with false positive rate on the x axis against true positive rate on the y axis. As such, a ROC graph depicts relative trade-offs that a classifier makes between benefits (true positives) and costs (false positives).
AUC	The Area Under the ROC Curve (AUC). As the name implies, this is simply the area under a classifier's curve expressed as a fraction of the unit square. Its value ranges from zero to one. Though a ROC curve provides more information than its area, the AUC is useful when a single number is needed to summarize performance, or when nothing is known about the operating conditions.

The information above has been extracted from the Exsilio Blog ([45] Joshi, R.; 2016) and from the book Data Science for Business ([43] Provost F. and Fawcett T.; 2013).

Appendix 9. Viable Framework - Step 4

Dimension	Area	Factor	Relevance in the Literature Review	Relevance in the Survey (Dimensions)	Relevance in the Survey (Factors)	Relevance in the Data
Concept	General	Mission	High	High	Medium	Very Low
		Vision	Low	High	High	Very Low
		Social impact	Low	High	Medium	Very Low
		Opportunities	Very Low	High	High	Very Low
		Trendiness	Very Low	High	Medium	Very Low
		Innovation	Very Low	High	High	Very Low
		Location	Medium	High	Medium	Very High
	Product	Product vs. Service	Very Low	High	Medium	Very Low
		Problem/Solution fit	Medium	High	Very High	Very Low
		Product/Market fit	High	High	Very High	Very Low
People	Founders	Academic level	High	Very High	Medium	Very Low
		Entrepreneurial experience	Very High	Very High	High	Very Low
		Professional experience	Very High	Very High	High	Very Low
		Industry experience	Very High	Very High	High	Very Low
		Network	High	Very High	High	Very Low
		Reputation	High	Very High	High	Very Low
		Motivation	High	Very High	Very High	Very Low

	Team	Adaptability	High	Very High	Very High	Very Low
		Completeness	High	Very High	High	Very Low
		Commitment	High	Very High	Very High	Very Low
Context	Environment	Economical	High	High	Medium	Very Low
	Market	Go-to-market strategy	High	High	High	Very Low
		Industry	High	High	High	Very Low
		Size	High	High	High	Very Low
		Growth	High	High	High	Very Low
		Barriers to entry	High	High	High	Very Low
	Competitors	Structure	High	High	Medium	Very Low
		Strength	High	High	High	Very Low
		Product differentiation	High	High	High	Very Low
	Partners	Structure	Very High	High	Medium	Very Low
Execution	Situation	Stage	High	Very High	Medium	Very Low
	Strategy	Business model	Very High	Very High	High	Very Low
		Scalability	High	Very High	Very High	Very Low
	Plan	Business plan	High	Very High	High	Very Low
		Marketing plan	High	Very High	High	Very Low
		Financial plan	Medium	Very High	High	Very Low
		Momentum	High	Very High	High	Very Low
	Sales	Channels	High	Very High	High	Very Low

		Clients	High	Very High	High	Very Low
		Users	High	Very High	High	Very Low
Financial	Revenues	Current	Medium	Medium	Medium	Very Low
		Next 12 months	High	Medium	Medium	Very Low
		In 5 years	High	Medium	High	Very Low
	Expenses	Burn rate	Medium	Medium	High	Very Low
	Cash	Cash flows	Medium	Medium	High	Very Low
	Capital	Capitalization	High	Medium	Medium	Very High
	Investment	Previous external investors (Track record / Cap table)	High	Medium	High	Very Low
		Previous external investors interested in investing now	High	Medium	High	Very Low
		Credible external investors interested in investing now	High	Medium	High	Very Low
		Amount required	High	Medium	High	Very Low
		Funding period	Very Low	Medium	Very Low	Very High
Legal	Incorporation	Form	Medium	Medium	Medium	Very Low
	Deal	Terms & Conditions	Very Low	Medium	High	Very Low
		Incentive effects of the allocation (risk vs. reward)	Very Low	Medium	High	Very Low
		Simplicity	Very Low	Medium	High	Very Low
		Fairness	Very Low	Medium	High	Very Low
		Robustness	Very Low	Medium	High	Very Low



		Lack of perverse incentives	Very Low	Medium	High	Very Low
		Lack of foreclose valuable options	Very Low	Medium	Medium	Very Low
	IP	Status	High	Medium	High	Very Low
	Attorney	Intellectual property firm	Low	Medium	Medium	Very Low

Appendix 10. Viable Framework - Total

Dimension	Area	Factor	Relevance in the Literature Review	Relevance in the Survey (Dimensions)	Relevance in the Survey (Factors)	Relevance in the Data	Total Relevance
Concept	General	Mission	3	3	2	0	2
		Vision	1	3	3	0	2
		Social impact	1	3	2	0	2
		Opportunities	0	3	3	0	2
		Trendiness	0	3	2	0	1
		Innovation	0	3	3	0	2
		Location	2	3	2	4	3
	Product	Product vs. Service	0	3	2	0	1
		Problem/Solution fit	2	3	4	0	2
		Product/Market fit	3	3	4	0	3
People	Founders	Academic level	3	4	2	0	2
		Entrepreneurial experience	4	4	3	0	3
		Professional experience	4	4	3	0	3
		Industry experience	4	4	3	0	3
		Network	3	4	3	0	3

		Reputation	3	4	3	0	3
		Motivation	3	4	4	0	3
		Adaptability	3	4	4	0	3
	Team	Completeness	3	4	3	0	3
		Commitment	3	4	4	0	3
Context	Environment	Economical	3	3	2	0	2
	Market	Go-to-market strategy	3	3	3	0	2
		Industry	3	3	3	0	2
		Size	3	3	3	0	2
		Growth	3	3	3	0	2
		Barriers to entry	3	3	3	0	2
	Competitors	Structure	3	3	2	0	2
		Strength	3	3	3	0	2
		Product differentiation	3	3	3	0	2
	Partners	Structure	4	3	2	0	2
Execution	Situation	Stage	3	4	2	0	2
	Strategy	Business model	4	4	3	0	3
		Scalability	3	4	4	0	3
	Plan	Business plan	3	4	3	0	3
		Marketing plan	3	4	3	0	3

		Financial plan	2	4	3	0	2	
		Momentum	3	4	3	0	3	
	Sales	Channels	3	4	3	0	3	
		Clients	3	4	3	0	3	
		Users	3	4	3	0	3	
Financial	Revenues	Current	2	2	2	0	2	
		Next 12 months	3	2	2	0	2	
		In 5 years	3	2	3	0	2	
	Expenses	Burn rate	2	2	3	0	2	
	Cash	Cash flows	2	2	3	0	2	
	Capital	Capitalization	3	2	2	4	3	
	Investment	Previous external investors (Track record / Cap table)	3	2	3	0	2	
		Previous external investors interested in investing now	3	2	3	0	2	
		Credible external investors interested in investing now	3	2	3	0	2	
		Amount required	3	2	3	0	2	
		Funding period	0	2	0	4	2	
	Legal	Incorporation	Form	2	2	2	0	2

	Deal	Terms & Conditions	0	2	3	0	1
		Incentive effects of the allocation (risk vs. reward)	0	2	3	0	1
		Simplicity	0	2	3	0	1
		Fairness	0	2	3	0	1
		Robustness	0	2	3	0	1
		Lack of perverse incentives	0	2	3	0	1
		Lack of foreclose valuable options	0	2	2	0	1
	IP	Status	3	2	3	0	2
	Attorney	Intellectual property firm	1	2	2	0	1