STARTUP SUCCESS PREDICTION. EXAMPLE FROM THE US.

by

Anna Melnychuk

A thesis submitted in partial fulfillment
of the requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2021

Thesis Supervisor: ______________________ Professor Tymofiy Brik

Approved by ____
Head of the KSE Defense Committee, Professor [Type surname, name]

____

____

____

Date ______________________________________

ACKNOWLEDGMENTS

I would like to express my special thanks of gratitude to my advisor Professor Timofiy Brik for his expert advice and encouragement throughout this work, as well as Oleksandr Gorbach, Kun Nhan Nhuien, and Yelyzaveta Tikhonovska for their constant support and guidance in tough moments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND TERMS

Startup young company founded to develop a unique product or service, bring it to market and make it irreplaceable for customers

AUM Active Assets Under Management

Series A startups the first venture capital-backed funding that allows angel investors to exit the startup

IPO Initial Public Offering

CAGR Compound Annual Growth Rate

Venture Capital a specific case of Private Equity, where the investment is made in the very early stages of a company's life

Traction Financial performance of a company during its lifetime

Cash burn rate rate at which company uses up its cash reserves or cash balance. A measure of negative cash flow, typically recorded as a monthly rate

EV, Enterprise Value measure of company's total value. Enterprise value includes the market capitalization of a company, as well as its short-term and long-term debt

YoY Year Over Year

Unicorn company, which has its Enterprise Value of more than 1 Billion USD

CHAPTER 1. INTRODUCTION

Initially, the idea of this work is to define the factors that influence the private company's probability to become successful. In this work, a successful company is the one that achieved unicorn status. The term \`unicorn\` was firstly mentioned by Aileen Lee (Brown and Wiles, 2015) and outlines companies that:

- raised money via at least one funding round;
- have always been privately held;
- are not a division of publicly-traded companies;
- have their market value of one billion USD or more.

Often \`unicorns\` are the companies, which provide disruptive technology in their industry sectors and, thus, offer the market a product, which solves a critical problem. For instance, in August 2013 Uber - the transportation network company, which matches drivers and riders, raised $319M of private market funding and achieved a $3.7B market valuation. It was the company's third financing round when it got such a high valuation estimated by top investors while remaining privately held. This transaction demonstrates and is a bright example of a financial market trend that has been evolving over the past decade. It describes a shift, which happened in a way of how private technology firms are raising growth capital. It could be observed that a big number of startups refrain from going public and prefer to stay privately held. Companies, growing at a higher pace have historically made small funding rounds, where they received initial capital for their growth from angels, venture capital (VC) investors, and private equity investors before deciding to go to the public market and make an Initial public offering (IPO) to get larger sums of funding, which was critical to financing their long-term needs. Nowadays, rapidly growing companies can raise growth capital in the private markets to the extent that earlier was accessible only from public markets (Brown and Wiles, 2015).

The private markets are represented by Venture Capital investors and this market is booming. The second quarter of 2021 demonstrated the highest activity of venture capital ever: $156B was invested in startups by VCs, which represents 157% YoY growth in VC funding (CB Insights, 2021). The growth in funding led to a record in the number of unicorns emerging globally: in the second quarter of 2021 136 new unicorns were born, which represents 491% YoY growth (CB Insights, 2021).

The beautiful thing about Venture Capital investment is that the earlier investor identifies a potentially successful company – the higher return they will make from an exit from it. An exit happens, when an investor decides to get rid of an investment in a business venture and can be a result of IPOs, acquisitions, buyouts, or bankruptcy (Investopedia, 2021). For instance, when Uber became a public company, it made 755x return for its early investor Benchmark capital. Alibaba made Softbank 3000x return, while Facebook made Accel (an early-stage investor) 709x return (Pitchbook, author's calculations). Being a top return instrument, Venture capital early investments make, on average, 21.2% yearly return (Cambridge Associates, 2020). It is 1.55x higher than the 13.6% annual average return of S&P (Business Insider, 2021).

The problem is private companies tend to keep financial and operational data private (because they have no requirements to make it public, unlike publicly traded companies) and this makes the process of companies' valuation even more complicated to investors. Traditional models, which are used to evaluate business performance do not apply to startups, because they often lack financial information, as well as data on growth and sometimes even market size. Venture capitalists and private investors tend to give a very subjective evaluation to startups at the early stages of their growth because they majorly focus on the background of entrepreneurs, uniqueness of the idea, potential market share the start-up could gain (Bonaventura et al, 2019). But this process is labor-intensive and takes time and good intuition to conduct successfully.

Overall, there are 6 types of stakeholders in business, and the majority of them (employees, investors, suppliers, and managers) can benefit from building prediction models for business success. Investors can focus better on financial returns while identifying more accurately the perspective ventures to put money into and minimize the risk of funding potential failures. Employees can make the right choice regarding career planning and provide themselves with more stable income and safety (reduce risk of joining the company with bankruptcy potential). Also, suppliers could rely on constant revenue generation. Managers could detect early signals of bankruptcy and take measures to survive while saving human and financial resources.

Therefore, some metrics for the detection of potentially successful startups would be beneficial for their stakeholders. In this research, we will study how to decide on future startup market valuation concerning information given on basic stages of their development.

The research question is formulated as follows: how to detect a unicorn at the early stages of the private company's development with limited or no information regarding the company's operational and financial traction. Thus, the author takes the factors, which potentially can influence the future market value of a startup and analyzes what impact they have made historically.

The author uses 3 logistic regressions to see which factors influence the probability of a startup to become a unicorn positively and negatively. The original data was obtained from the world's largest database on private equity companies Pitchbook. Additional variables for regressions were created using Social Network Analysis tools.

The model showed that startups, which operate in the health industry are 7.8% less likely to become a unicorn, while startups, which provide software as a service platform, are 4.8% more likely to become a unicorn. We have also found that each additional investor increases the odds of becoming a unicorn by 3%, while the additional link between companies of different investor clusters increases the odds by 0.03%. The last finding indicates that the bigger the number of times the company is connected to investors of different clusters, the higher probability it has to become a unicorn. We have found that, for example, a company with 7000 connections has a 50% chance to become a unicorn, while a company with 10000 connections has approximately 80% chance. We have also discovered that with each additional $1M the company raises, the odds of becoming a unicorn increase by 2%. The model also demonstrates that if the company raises more than $1000M of funding it is a unicorn, which is a logical conclusion.

By giving answers to the research question, the contribution of this work is important for practical usage. Firstly, we make a contribution to the existing literature about startups valuation and in particular on determinants for venture valuation. Second, we offer a new understanding of how investors' characteristics impact the probability for startups to become a unicorn and in this way identify critical influences for venture success. Thereby, our findings can be used by the stakeholders of a venture to make more strong assumptions on its future success at the early stage of the company's development.

CHAPTER 2. INDUSTRY OVERVIEW AND RELATED STUDIES

2.1 Current state of Startup ecosystem and the role of Venture Capital

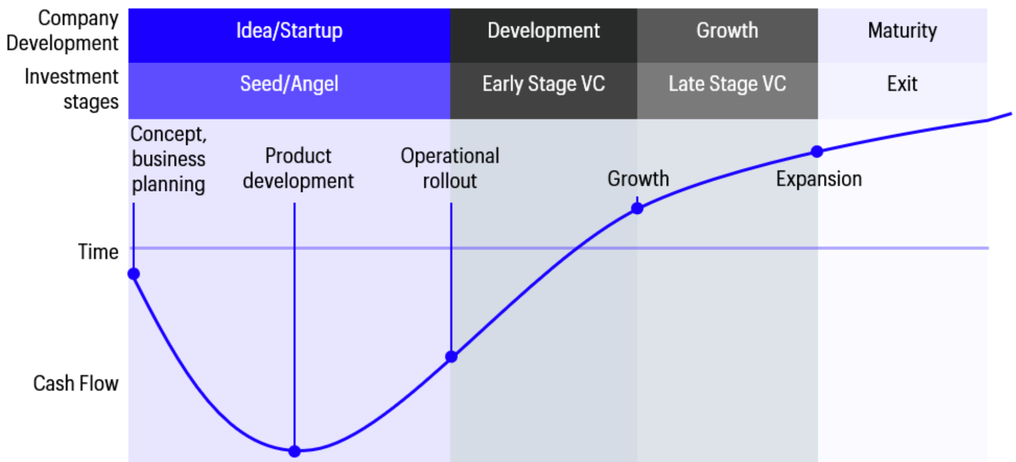
According to Forbes (2021) definition, startups are young companies founded to develop a unique product or service, bring it to market and make it irreplaceable for customers. Startups are rooted in innovation, addressing the deficiencies of existing products or creating entirely new categories of goods and services, thereby disrupting entrenched ways of thinking and doing business for entire industries. That's why many startups are known within their respective industries as "disruptors".

In this paper, the startup's definition will be adjusted to the data used. Thus, US companies, which are privately held, registered for IPO, acquired or merged and venture capital-backed will be considered as startups.

Private Equity is an investment made by a financial institution – Private Equity Investor (PEI) in the equity of a non-listed or private company. Venture Capital (VC) is a specific case of Private Equity, where the investment is made in the very early stages of a company's life (Private Equity and Venture Capital Course, Bocconi University, 2020).

VC plays a vital role in a tech ecosystem and allows the development of new startups since it finances and assists the startups from the product development stage to the moment when it becomes a self-sufficient company with partial or full ability to finance itself independently (Figure 1).

Figure 1: Startup financing cycle



Source: Private Equity and Venture Capital Course, Bocconi University, 2020; author's visualization

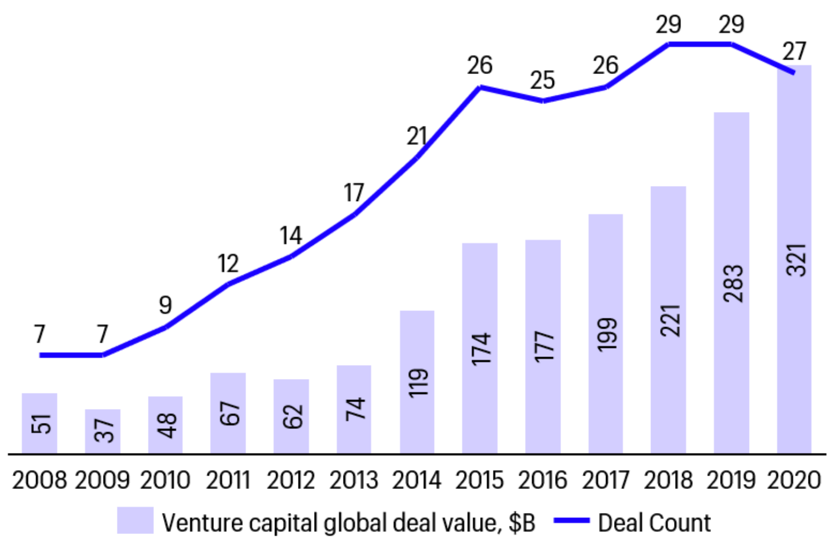
The investors' motivation to invest lies in high return potential or the opportunity to make more than 100x return on money invested. For example, for an early-stage investment (Series A) Facebook made Accel 709x return, Snapchat made Benchmark Capital Partners 237x return, Alibaba made Softbank 3000x return, Coinbase made Union Square Ventures 920x return, Uber made Benchmark Capital Partners 755x return[1].

Currently, the Venture Capital market is booming, because it has experienced a gigantic expansion in the number of unicorns. As of October 2021, there are 851 unicorns globally and their cumulative valuation amounts to $2.7T. Among them, 366 reached unicorn status between January and October 2021 (CB Insights, 2021). Notably, the US is home to most of the top entrepreneurs with 402 unicorns with 58% of the new unicorns created in 2021 globally (The Asian Banker, 2021).

The global VC deal value increased 2 times since 2015 and grew at a higher pace each year (Figure 2).

[1] Calculated by the author from publicly available source (Pitchbook)

Figure 2: Venture Capital Global deal value



Source: Pitchbook, author's visualization

Even though venture investment is a top return instrument, it brings risks associated with investing in early-stage, such as limited traction or its absence, unprofitability and a high level of cash-burn, absence of product working in its full capabilities, and lower liquidity of investments compared to other markets. That is why early investors should rely more on advanced metrics while assessing early-stage startups.

2.2 What is considered to be a successful startup

The success of a particular company is often defined by its financial indicators. Initially, they allow comparing a firm's performance with a wide range of similar companies by outlining industry benchmarks. On the other hand, financial performance indicators deliver a tracking of a company's historical conduct over its lifetime.

For example, Lussier (1995) considers a business to be successful if it has made at least industry-average profits for the previous three years. The performance also may be viewed as profitability, i.e., a company's ability to generate revenue and profit in terms of its current labor, assets, and capital stock, as well as marketability, i.e., a company's performance in the stock market by its revenue and profit generated (Joe Zhu, 1998). Butler and Fitzgerald (2000) determine success as having a competitive advantage over similar companies.

However, to evaluate startups, the traditional `valuation` methods mentioned above cannot be applied. These companies usually lack enough financial data on their historical performance and sometimes do not have direct peers to compare with. Lastly, it is common that startups even do not generate or disclose their revenue, implying that their profitability may be negative even after some years of their ongoing operations.

Therefore, other definitions of the success of a company are needed, when we talk about startups. For instance, 'second round' or 'Series A' funding (a major investment by venture capitalists to support growth) is widely considered as a factor that distinguishes successful from unsuccessful startups (Abbassi, Schlagwein, Fischbach, 2016). García-Ochoa (2020) links high performance with the probability of getting financed and the size of startup financing. Bosma (2004) defines 3 measures of entrepreneurial performance of startups: profit (though can be misleading), cumulative employment created, and survival time (in months). Caseiro and Coelho (2019) include non-financial measures of performance, such as product-market outcomes (market share, introduction of new products and marketing effectiveness, and internal process outcomes). These operational factors can contribute to financial performance in the long run.

2.3 Startups' success prediction models

Lussier model (2001) states that the success or failure of a firm is defined with 15 variables: capital, record keeping, and financial control, industry experience, management experience, planning, professional advisors, education, staffing, product/service timing, economic timing, age of owner, partners, parents owned a business, minority, and marketing skills. To predict the success of startups logistic regression model was used. The results showed that the imposing role belongs to well-educated human resources, the development of employees, and timely professional advice.

In recent years machine learning techniques are applied to predict success. For instance, Dellerman et al (2017) used the Hybrid Intelligence Method that combines the strength of both machine and collective intelligence (humans' capabilities) to predict the success of startups. The need for such a combination arises because humans and machines both have advantages over one another and can make a unique contribution to the process of predicting a thriving business. For example, humans are capable of coming up with intuitive predictions, using their gut feeling, processing rare information, and making judgments based on their previous experience. It has been proven by Huang and Pearce, (2015), that decision-makers usually develop experience-based schemas to help them identify the opportunities for extraordinary returns. But a human's area of knowledge is often pretty narrow, that is why `calibration` is needed when qualified experts evaluate the idea. For example, Andrew Chen, a general partner at Andreessen Horowitz, whose area of expertise is majorly ads, analytics, consumer communication, and publishing, when making an investment decision always makes a discount on the fact that he might not know something in the area where his judgment is shaky. Also, if he gives advice to entrepreneurs and finds himself repeating the same delivery to a bunch of entrepreneurs, Andrew tries to refine and rethink the product to consider its specifics and peculiarities (Andrew Chen, Newsletter). Indeed, usually, people are biased, since the human mind uses cognitive shortcuts to solve complex problems under constraint. Indeed, Shah and Oppenheimer (2008) saw, that heuristics reduce cognitive effort, which, consequently, helps to use less mental effort to solve perplexing problems or make important decisions. That is where machines should help collective intelligence. Machine, on the contrary, can process hard data to identify the pattern most efficiently. The statistical models are unbiased and are not subject to social contingence. Moreover, the empirical statistical models become more accurate if the data processed grows. However, such a way of predicting success also has drawbacks, among which the most important is an inability of machines to explain the random error in distributions of data used (for example, company's age, number of employees). This random error term arises due to the unexpected risk as well as various signals of new ventures, like well-qualified c-level management, innovativeness, demand for a particular product at a particular point in time.

Dellerman et al (2017) used both machine learning (Logistic regression, Artificial Neural Network, Random Forests, Support Vector Machine, and Naives Bayes) and collective intelligence (opinions of non-experts and experts) to predict the success of Series A venture capital-backed startups. The signals for the prediction were divided for hard (industry, firm age, business model, competition, revenue model, capital raised, team size, entrepreneurial experience, financial support) and soft (product innovativeness, scalability, entrepreneurial vision, proof of value) signals. Thus, hard and soft signals were evaluated with a machine and humans respectively. However, this model hasn't been tested yet to provide real outcomes.

Diego Martinez (2019), who defined a successful startup as the one, which has its total funding above €1M, proved that significant predictors of a startup are five variables: university ranking (whether employees studied at the universities listed in QS world university rankings), time dedication (full-time vs part-time), team size, societal relevance of business based on Likert scale and employee incentives as a part of equity. He used a logistic model for the prediction incentives. Also, if success is revenue growth with an average annualized return of at least 20% in the past 3 years, then such variables as time dedication and employee incentives proved to be crucial. Lastly, if success has more than 10 employees, then the number of founders, usage of data obtained from the market, and employee incentives are significant as variables.

Francisco Ramadas (2017) employed logistic regression, support vector machines, and random forest to predict successful startups in the USA. According to his definition, a successful startup is one that either has IPO by going to a public stock market or is merged with/ acquired by another company. It was found, that the majority of successful startups are located in California (18%), New York (12%), and Massachusetts (20%). Success also depends on the industry with the highest probability of triumph in Healthcare, Entertainment, Hardware, and Software technologies.

In his book `Super Founders: What Data Reveals about Billion-Dollar Startups` Ali Tamaseb identifies factors that set apart startups that become unicorns. Among the statistically significant factors are:

- Founders are alumni of top-ranked schools
- Founders have worked for themselves or tier 1 companies
- Repeat founder
- The product is a pain killer, rather than a vitamin
- Company raised from tier 1 VCs
- Company raised within a 1 year of founding
- Company based in Silicon Valley.

Among the factors, which appeared to be not statistically significant are:

- Founders' years of experience
- Founders' past relevant experience
- Having technical founders, even for enterprise businesses

Although there are quite a few studies about predicting for startups success, most of which are based on machine learning techniques, such as random forest models, support vector machines, and logistic regression (as the most common predictive tool), there is still space for different types of approaches focused on various stages of startups' life and with such a data-rich platform as Pitchbook, it would be an interesting exercise to discover new patterns and compare them to previous findings.

CHAPTER 3. METHODOLOGY

3.1 Research hypotheses description

The major question, which the author wants to answer in this research is how to detect a startup with a potentially high valuation, in other words, a unicorn (enterprise value of more than $1B) concerning the information given on such stages of the development, when they are still considered as privately held companies. Thus, the author will take the factors, which potentially can influence the future value of a startup and analyze what impact they have made historically. The following research hypotheses will be answered using econometric analysis.

Hypothesis 1: The number of investors of the company increases the probability of becoming a unicorn more than the number of years the company exists.

Hypothesis 2: The company, which provides software as a service has higher chances of becoming a unicorn than other companies.

3.2 Startup success factors determination

Financial indicator

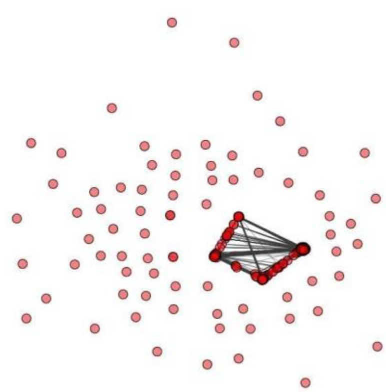
Since there is little or no data on financial performance or key financial indicators of privately held companies, even though some of them are already well-known brands or even unicorns, the author will use only 1 common metric, which is disclosed by major companies and thus, publicly available. This is the total capital raised by a startup.

Recognition from investors

To add to already existing models on startup valuation and success determination described earlier, guided with the data available, the author complemented her model with the factor that explains investor influence on performance. The number of investors in each of the businesses was included. Since other variables describing investors are categorical, it would be a mistake to treat them in the same way and further interpret them as simple numerical variables.

Additional 3 variables were created to understand the impact of the investing activity on a company's opportunity to become a unicorn. Initially, a startup-investors graph data model was built to understand how strong the companies are connected and whether they experience common investment activities (receive money from the same investors). Vertices represent the companies and nodes, which connect them, show the links between them. For instance, if two companies were funded by the same investor, they will have a node between them (Figure 3).

Figure 3: Startup-investors graph data model. Nodes (red dots) are the companies, connected with joint investors



Source: R-output, author's visualization

3 Variables were created with Social Network Analysis instruments using programming software R Studio. In particular, the author used centrality measures, which are used to understand graphs or networks:

1) Degree Centrality - an importance score based on the number of links held by each node. This measure explains how many direct connections each node

has to other nodes in the network. It is used to find very connected individuals, popular individuals, individuals who are likely to hold most information, or individuals who can quickly connect with the wider network (Cambridge Intelligence, 2020)

2) Betweenness centrality - the number of times a node lies on the shortest path between other nodes. This measure demonstrates which nodes are 'bridges' in a network (Cambridge Intelligence, 2020).

The final list of variables created is the following:

- Number of connections - the number of common investors of a particular company with other companies
- Unique investor – binary variable, which shows whether a particular company has common investors with other companies or not
- Betweenness – the variable which identifies companies that are common for many investors (if we assume, that there are clusters of investors interested, for example, in various industries, there will be companies that connect those clusters, i.e., attractive for various types of investors).

Industry influence

According to statistics, out of all privately held, later-stage VC globally, 25.32% comprise Software as a Service company, whereas 6.2% are Life sciences and healthcare startups (Pitchbook, 2021). In 2019 SaaS revenue was estimated at $171B, while it is projected to reach $369.4B by 2024, growing at a 17% CAGR (Forbes, 2021). 80% of the current market share or $186B of revenue generated by SaaS belongs to 770 unicorns (InnoFuture, 2021). Those two industries are among the most popular 5 among entrepreneurs, who decide to launch startups. The author will try to conclude whether starting a venture in a specific industry increases the chances to have a billion-dollar valuation. Thus, a combination of 2 dummies was created, which indicate whether a startup is a SaaS and whether it solves health problems: provides therapeutics, medical care, develops drugs, assists humans in diagnostics, or develops pharmaceuticals.

Age of a startup

The author decided to test whether the years of operating activity of a company influence its chances of becoming a unicorn. This variable was included in the analysis of Dellerman et al (2017), however, the model hasn't been tested yet, and thus it is interesting to find out if the age of the company is important for the company's success.

We include all of these variables in a logit model as the factors to predict the success of a startup.

3.3 The Logit Model

The author used the Logit Model to estimate how the factors mentioned above influence the probability of the US privately held companies reaching \$1B+ valuation. The author defines that if the company is worth more than \$1B (becomes a unicorn) it is successful. The following approach was chosen as it is the most appropriate model when predicting a value of a dependent variable following the given set of explanatory variables. Also, this model was used for the same research purposes previously. The author uses the biggest database on private equity companies and this way she makes her contribution to the existing findings. The Logit Model could answer the following question of this research: what factors influence the probability of becoming a unicorn?

The choice of variables was initially motivated by the literature review by the author and the data on startups available. The last known valuation (how much the company is worth) of a company was considered as a dependent variable. The author made another dependent variable on its basis, which was used in the model. This is a binary variable, which indicates the success of a startup: 1 (success) - if the startup's enterprise value is more than \$1B and 0 (failure) - if it is less than \$1B.

The model is formalized as below:

$$P(Yi = 1|Xi) = F(\beta_1 + \beta_2 Xi),$$

where Yi is the occurrence of success (a company becomes a unicorn {1}) depending on the company's characteristics Xi.

Each variable was prescribed with the expected sign, based on the common sense and expertise of the author (Table 1).

Table 1: Expectations regarding the sign of independent variables

Variable Name	Description	Expected sign
Total raised	Total capital raised by a company through funding	+
Number of Active Investors	Number of active investors	+
Number of Connections	Number of common investors of a particular company with other companies	+
Betweenness	A variable that identifies how popular the company is for different clusters of investors	+
Isolates	Dummy for a company = 1, if a startup has unique investors, 0 otherwise	-
Years since founding	Years since founding	+
Health	Dummy for an industry =1, if it is startup's industry, 0 otherwise	+
Business/Productivity Software	Dummy for an industry =1, if it is startup's industry, 0 otherwise	+
SaaS	Dummy for an industry =1, if it is startup's industry, 0 otherwise	+

Source: Pitchbook data, author's assumptions

CHAPTER 4. DATA

4.1 Preliminary analysis of data

The initial data was obtained from the PitchBook platform, which is the best private market data provider. It consists of 1518 observations: those are startups, which have their headquarters located in the United States and undergo venture capital financing, in particular Series A, Series B, and Series C stages of financing. The intuition behind choosing startups at these particular stages of financing was to select the companies at the early and growth stages of their development to predict the highly valued companies when they are still considered to be young but have a full-fledged product. The initial data set was edited for ease of usage and more accurate results of the logistic regression. Steps, which were taken to clear the data set are described as follows:

1) Startups with the last founding round earlier than 2018 were deleted from the list (the motivation behind is to process the newest information on the startup ecosystem)
2) Startups, which underwent Initial public offering (IPO) were deleted from the list (the financial valuation of these startups can no longer be compared to the financial valuation of privately held companies because they experience the influence of the stock price fluctuations due to the external factors on the public market)
3) Startups with financial valuation and total money raised of less than $10k were deleted from a list since they would significantly distort the results. Moreover, they are outside of our scope of interest, because startups with such financing sums usually are pre-seed companies with no minimum viable product at that stage
4) The dataset consisted of some N/A for the important variables, so these companies were deleted from the observations list

Finally, 1487 companies and their characteristics were used by the author as a dataset for the research conducting. 201 Companies from the list are unicorns and the remaining 1286 haven't achieved this status yet (Table 2).

Table 2: General statistics summary of variables

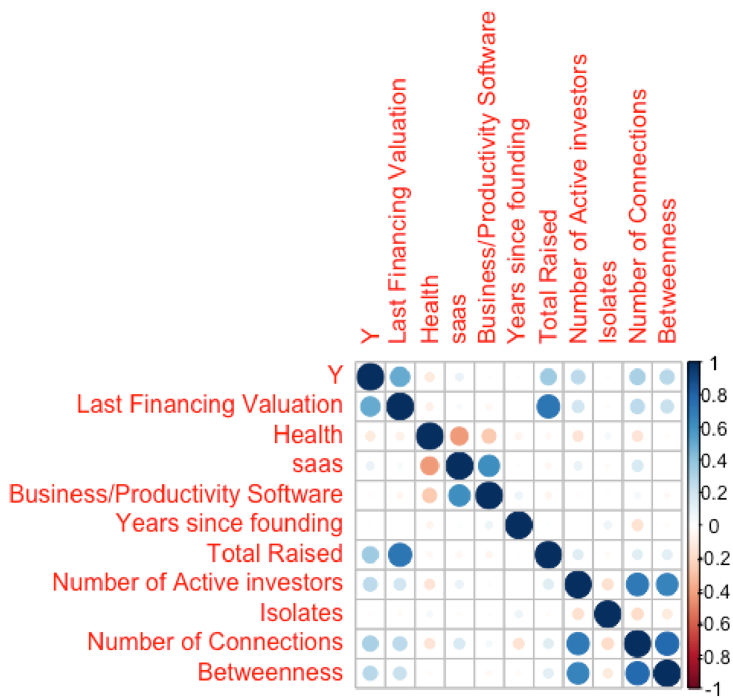
Variable	Obs	Min	Mean	Median	Max	Std. Dev.
Y (Unicorn)	1487	0	0.135	0	1	0
Total raised	1487	1	142	58	14980	542
Number of Active Investors	1487	1	13	11	90	11
Number of Connections	1487	0	90	73	455	80
Betweenness	1487	0	892	451	10806	1227
Isolates	1487	0	0.026	0	1	0
Years since founding	1487	0	7	6	38	4
Health	1487	0	0.234	0	1	0
Business/Productivity Software	1487	0	0.186	0	1	0
SaaS	1487	0	0.373	0	1	0

Source: Pitchbook, author's calculations

The table represents the summary statistics of the data used for the regression analysis. It can be observed that 13.5% of the selected companies have an enterprise value of more than $1B (they are unicorns), 23% operate in the health industry, 18.6% provide business or productivity software, and 37.3% offer Software as a Service as a core product. Notably, on average startups raised $142M from VC investors. However, we should consider the median value, since the data includes some unicorns, which raised significant sums of money. For example, JUUL - manufacturer of e-cigarettes and nicotine products intended to provide an alternative to tobacco smoking – raised $14,980M, which is not a common practice for a VC-backed firm. The average quantity of active investors for all the companies is 13, while the average number of common investors (number of connections) with other companies is 90. Only 2.6% of startups were financed by unique investors and thus have no connections to other companies (they are isolates). On the median, the one company connects the different clusters of investors 451 times, meaning that even though each investor has its preferences (to industries/ business models/ types of products) the majority of them are versatile in their investment strategies and they fund different types of startups. The higher this betweenness is, the more attractive a startup should be to different types of investors. Also, on average, the selected startups have been operating for 7 years.

To assess the presence of highly correlated variables in the model, which can cause multicollinearity problems, we used a correlation matrix (Figure 4).

Figure 4: Correlation matrix



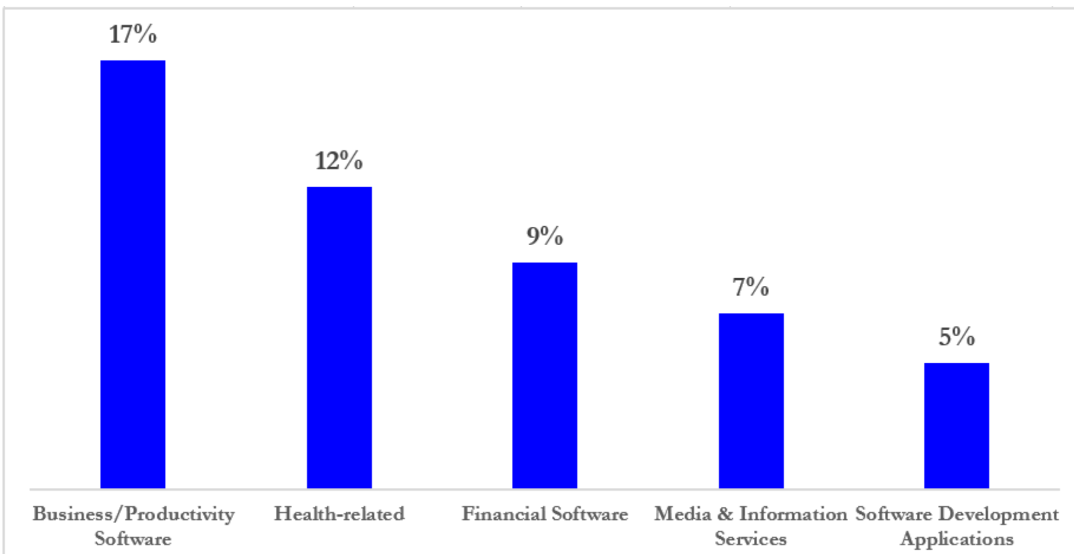
Source: Pitchbook, author's calculations

Initially, the dependent variable \`Y\` correlates with the independent variables collected for the studies, in particular: negatively with \`Health\` and positively with \`SaaS\`, \`Total Raised\`, \`Number of Active Investors\`, \`Number of Connections\` and \`Betweenness\`. All of the correlation coefficients lie in the range of -0.11 to 0.35, which is acceptable for further analysis. Yet, there is a correlation of 0.72 between \`Last Financing Valuation\` and \`Total Raised\`, which can cause a multicollinearity problem in the model, since the dependent variable \`Y\` was created from \`Last Financing Valuation\` (it is 1 if

`Last Financing Valuation` is >$1B and 0 otherwise). Also, the correlation coefficient (-0.42) between `Health` and `SaaS` should be taken into consideration). `SaaS` and `Business/Productivity Software` are positively correlated with a 0.62 coefficient. Notably, `Number of Active Investors` is positively correlated with `Betweenness` (0.66) and `Number of Connections` (0.7). `Betweenness` and `Number of Connections` have this coefficient equal to 0.77. Thus, it is very important to omit using these variables in the same model. Alternatively, the multicollinearity problem should be solved.

In our dataset, the most popular industry among startups is Business and productivity software, as it comprises 17% (Figure 5).

Figure 5. Unicorns' distribution among top 5 associated industries, in percent



Source: Pitchbook, own calculations

The second biggest industry cohort among unicorns is Health care services (12%). The Financial software industry comprises 9% and Media & Information Services comprise 7% of all the highly valued companies. This is in line with the global statistics mentioned above.

Business and productivity software means an application designed to create a smooth working experience for organizations, corporations, and individuals while enabling them to carry out their daily tasks efficiently. The common features are file sharing, task calendars, chat boxes, employee communications (Software world, 2021). The most highly valued among them in the dataset is Gong, enterprise value (EV) $7.25B. This is a revenue intelligence platform, which uses Artificial Intelligence to tell businesses necessary information on each customer like who is most likely to continue using the product and who is most likely to churn. The company was established in 2015 and raised $584M as of Jun'21 via 7 financing rounds. The success of this startup lies on 3 main pillars: required B2B product, which helps businesses to retain customers; tier 1 VC funds among investors (Sequoia Capital, Coatue, Tiger Global, Salesforce Ventures) and strong founding team (founders with experience at top tech companies, e.g., Microsoft). Among other interesting startups are Outreach, EV $4.4B (Machine Learning platform for sales engagement), Calendly, EV $3B (Software to schedule meetings without back-and-forth emails), Forter, $3B (Developer of a fraud prevention platform designed to help online retailers detect and eliminate transaction risks) and Thoughtspot, EV $2.4B (Developer of enterprise analytics platform designed to easily analyze complex, large-scale enterprise data with an automatic, relational search engine).

Another industry we were interested in is Healthcare Services, where among the companies with the highest worth are Roman Health Ventures, EV $5B (Care platform, which offers a personalized, end-to-end healthcare experience from diagnosis, to delivery of medication and ongoing care), Talkspace, EV $1.4B (Platform, which is intended to connect individual clients with a network of licensed therapists and allows clients to send their dedicated therapists' text, video, and voice messages, engages in live video sessions, enabling clients to avail psychiatry services, including prescription fulfillment, adolescent therapy, and couples counseling online). Notably, healthcare startups often have tough funding history with a lot of rounds raising small sums of money. Talkspace, for instance, raised its first financing round in 2012 and it took the company almost 10 years to become a unicorn. This may explain the negative sign of the correlation between our dependent variable and the healthcare startup dummy variable[2].

[2] Information about the companies' activities was taken from their websites and Pitchbook

5.1. Logit model output

It is needless to mention that eventually 2 logit models were built to make the results of the regressions more accurate and relevant. Both of them have a binary variable (unicorn/ not unicorn) as a dependent variable, but the industry factor, namely whether a startup is a SaaS or it is related to a Health care company has proved to show many useful results when considered independently. Such a conclusion was made due to the assessment of the pseudo-R^2 and Likelihood Ratio Test, as well as coefficients of the variables obtained from a joint model and two separate models (where variables \`Health\` and \`SaaS\` are the components of separate logistic regressions). For this reason, we will describe the outputs of the 2 regressions separately. Additionally, a model with the separate effect of the raised money was made.

Table 3. Logistic regression output (1)

		Y	
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.07	0.04 – 0.11	**<0.001**
Health	0.47	0.28 – 0.77	**0.004**
saas	1.28	0.91 – 1.80	0.152
Number of Active investors	1.03	1.01 – 1.04	**0.001**
Years since founding	1.00	0.95 – 1.05	0.980
Isolates	1.74	0.50 – 4.59	0.314
Betweenness	1.00	1.00 – 1.00	**<0.001**
Observations	1487		
R^2 Tjur	0.106		

Source: author’s calculations

Companies, which provide their solutions in the health and wellness industry reduce the odds of becoming a unicorn by 53% (Table 3). Additional investor increases the odds of becoming a unicorn by 3%, while the additional link between companies of different investor clusters increases the odds by 0.04%.

We also used the logit model average marginal effects to detect the influence of the dependent variables (Table 4).

Table 4. Binary outcome model marginal effects (1)

Is unicorn	Logit average marginal effects
Health	-0.0778*
saas	0.0258
Number of Active investors	0.0027*
Years since founding	0
Isolates	0.0577
Betweenness	0*

Source: author's calculations

Startups, which operate in the health industry are 7.8% less likely to become a unicorn. Probably, this is because companies in this industry normally have tough funding history and have to be able to suggest a disrupting solution to become highly valued among investors. For each additional investor, startups are 0.3% more likely to become a unicorn.

The overall goodness of fit of this model was measured by Pseudo-R^2 (14.5%) and the likelihood ratio test (123). The model is significant since the overall P-value is <0.0001.

Yet, the effect of the software industry is unclear from this model. Thus, the variable health was omitted to see whether SaaS has some effect on getting highly valued. The following model's goodness of fit was also measured by Pseudo-R^2 (13.4%) and the likelihood ratio test (113). Although both of the values are slightly lower than in the previous model, we can now assess the SaaS effect (Table 5).

Table 5. Logistic regression output (2)

	Y		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.05	0.03 – 0.08	**<0.001**
saas	1.59	1.16 – 2.18	**0.004**
Number of Active investors	1.03	1.01 – 1.05	**<0.001**
Years since founding	1.01	0.96 – 1.05	0.825
Isolates	1.66	0.48 – 4.37	0.352
Betweenness	1.00	1.00 – 1.00	**<0.001**
Observations	1487		
R^2 Tjur	0.100		

Source: author’s calculations

The interpretation is as follows: If the company offers software as a main service, it increases the odds of becoming a unicorn by 59%. Each additional investor increases the odds of becoming a unicorn by 3%, while the additional link between companies of different investor clusters increases the odds by 0.03%.

We also used the logit model average marginal effects to examine the influence of the dependent variables (Table 6).

Table 6. Binary outcome model marginal effects (2)

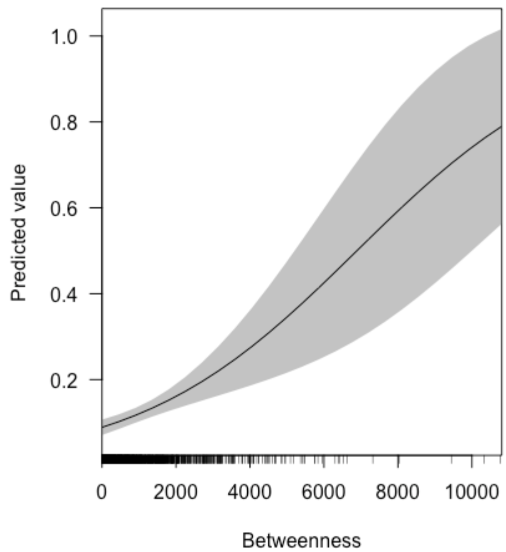
Is unicorn	Logit average marginal effects
saas	0.0486*
Number of Active investors	0.0031*
Years since founding	0
Isolates	0.0535
Betweenness	0,000035*

Source: author's calculations

Startups, which provide SaaS are 4.8% more likely to become a unicorn. For each additional investor, startups are 0.3% more likely to become a unicorn. Also, a startup with an additional link between companies of different investor clusters is 0.003% more likely to become a unicorn.

To make the effect of betweenness clearer to the reader it is shown in Figure 6. This is the prediction of the probability to become a unicorn with a specific quantity of connections passing through one company (the number of times the company is connected to investors of different clusters). It depicts that, for example, a company with 7000 connections has a 50% chance to become a unicorn, while a company with 10000 connections has approximately 80% chance.

Figure 6. Prediction of the unicorn with number of connections



Source: author's calculations

The two previous regressions do not demonstrate the influence of total money raised on the dependent variable. That is because the variable `Total Raised` is highly correlated with the company's valuation (the higher the sum of the raised money, the higher the enterprise value of the company). Due to this correlation `Total Raised`, if included in previous models, significantly distorts the results. Thus, the author tested the effect of the raised money on the company's valuation separately (Table 7). The model showed a high model fit (pseudo-R^2 is 63%) and the p-value of the model is <0.0001.

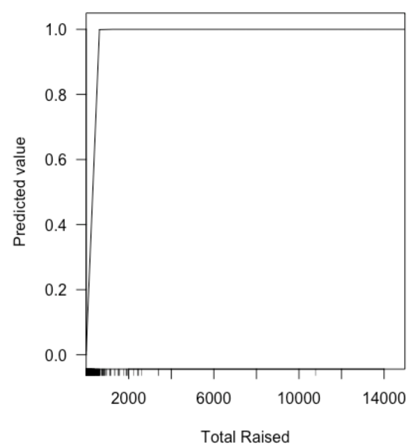
Table 7. Logistic regression output (3)

		Y	
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.01	0.01 – 0.02	**<0.001**
Total Raised	1.02	1.02 – 1.02	**<0.001**
Observations	1487		
R^2 Tjur	0.586		

Source: author's calculations

The result means that with each additional $1M raised, the odds of becoming a unicorn increase by 2%. The average marginal effect shows that with an additional $1M of funding the company is 1% more likely to become a unicorn. The model also demonstrates that f the company raises more than $1000M of funding it is a unicorn, which is a logical conclusion (Figure 7).

Figure 7. Prediction of the unicorn with the funding amount



Source: author’s calculations

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

This research primarily focuses on how to predict a startup's success at an early stage of its performance. The need for such work arises due to the high costs of all the stakeholders, who decide to invest their resources in a startup. Firstly, it takes a lot of time and energy to make a decision, while processing dozens of terabytes of information on startups and arriving at a conclusion with a range of risk constraints. Secondly, startup success prediction differentiates from corporate success prediction, as the most important financial data to evaluate business is often missing.

Literature studied mainly uses logistic regression and machine learning techniques to decide on future startup success. However, there is still a lot of room for various types of approaches focused on different stages of startups' life and with the help of the biggest platform on private equity firms – Pitchbook, the author discovered new patterns of success modeling. Hence, the author used 3 logistic regressions to see which factors influence the probability of a startup to become a unicorn positively and negatively.

The original data was obtained from the world's largest database on private equity companies Pitchbook. Finally, 1487 companies and their characteristics were used by the author as a dataset for the research conducting. 201 Companies from the list are unicorns and the remaining 1286 haven't achieved this status yet. 3 logit models with a binary variable (unicorn/ not unicorn) as a dependent variable, and industry factor, the number of years the company operates, the amount of funding it raised, and the number of investors as independent variables were built. To understand the impact of the investing activity on a company's opportunity to become a unicorn better, additional 3 variables were created with the help of Social Network Analysis and Graph Analysis techniques. The author used the centrality measures, such as degree centrality and betweenness centrality to explain how, for example, the number of common investors of a particular company with other companies influences its probability to become a unicorn.

The model showed that startups, which operate in the health industry are 7.8% less likely to become a unicorn. We assume this is because companies in this industry normally have tough funding history and have to be able to suggest a disrupting solution to become highly valued among investors. On the contrary, startups, which provide software as a service platform, are 4.8% more likely to become a unicorn. This finding is logical since 30% of all unicorns appear to be software platforms, which optimize business workflow or management of financial assets. We have also found that each additional investor increases the odds of becoming a unicorn by 3%, while the additional link between companies of different investor clusters increases the odds by 0.03%. The last finding indicates that the bigger the number of times the company is connected to investors of different clusters, the higher probability it has to become a unicorn. We have found that, for example, a company with 7000 connections has a 50% chance to become a unicorn, while a company with 10000 connections has approximately 80% chance. We have also discovered that with each additional $1M the company raises, the odds of becoming a unicorn increase by 2%. The model also demonstrates that f the company raises more than $1000M of funding it is a unicorn, which is a logical conclusion.

However, this model needs further improvements and development. First of all, more data on startups coming from various world regions would be helpful to enrich the dataset and make predictions more accurate. Secondly, more personalized data on the management of each company will be useful, although almost impossible to obtain, since it will take the extra effort of reaching each founder individually and making an interview. The alternative way is to look through each founder's page on LinkedIn to be able to get info on the previous career or educational background. Also, it would be still interesting to see how the industry defines further a company's success. Presumably, focusing on a specific industry and its branches, thus differentiating success benchmarks for various industries would help to determine the driving factors. Last but not least, discovering patterns in social networks between startups and VC funds could also add value if we want to understand how networking, the experience of founders, and even locations influence the companies' probability to become multibillion stories.

All in all, startup success prediction models are beneficial to all the stakeholders and it is essential to invent new models with the purpose of more accuracy and time efficiency while making an investment decision.

REFERENCES

Ali Tamaseb. 2021. *Super Founders: What Data Reveals about Billion-Dollar Startups*. Published in May 2021.
https://www.amazon.com/Super-Founders-Reveals-Billion-Dollar-Startups/dp/1541768426

Andrew Chen. 2020. Why are we so bad at predicting startup success? *Newsletter from a general partner at tier 1 VC funds*
https://andrewchen.com/why-are-we-so-bad-at-making-startup-predictions/

Andrew Disney. 2020. Social network analysis 101: centrality measures explained. Cambridge Intelligence. Published in January.
https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/

Anuj K. Shah and Daniel M. Oppenheimer. 2008. Heuristics Made Easy: An Effort-Reduction Framework. *Psychological Bulletin* 2: 207–222

Cambridge Associates 2020. Statistical information
https://www.cambridgeassociates.com/wp-content/uploads/2020/11/WEB-2020-Q2-USVC-Benchmark-Book.pdf

CB Insights. State of Venture Q2'21 Report.
https://www.cbinsights.com/research/report/venture-trends-q2-2021/

CB Insights. Statistical information.
https://www.cbinsights.com/research-unicorn-companies

Chalfant James, and Alston Julian. 1988. Accounting for Changes in Tastes. *Journal of Political Economy* 96 (April): 391-410.

Course on *Private Equity and Venture Capital*, Bocconi University. 2020. Course provided by Coursera
https://www.coursera.org/learn/private-equity

Diego Martinez. 2019. Startup success prediction in the dutch startup ecosystem. Delft University. Published in October
https://repository.tudelft.nl/islandora/object/uuid:1adc2972-db09-4583-b2da-05fd4e462941?collection=education

Dominik Dellerman, and Nikolaus Lipusch. 2017. Finding the Unicorn: Predicting Early Stage Startup Success through a Hybrid Intelligence Method. *SSRN Electronic Journal* (January)

Francisco Ramadas. 2017. Predicting Start-up Success with Machine Learning. University of Lisabon. Published in November
https://run.unl.pt/bitstream/10362/33785/1/TGI0132.pdf

García-Ochoa. 2020. How business accelerators impact startup's performance: Empirical insights from the dynamic capabilities approach. *Journal `Intangible Capital`* 16 (December): 107

Investopedia, Financial encyclopedia.
https://www.investopedia.com/terms/e/exitstrategy.asp

Joe Zhu. 1998. Multi-factor performance measure model with an application to Fortune 500 companies. *European Journal of Operational Research* 123 (2000): 105-124.
http://www.deafrontier.net/papers/EJORfortune.pdf

Keith C. Brown, and Kenneth W. Wiles. 2015. In Search of Unicorns: Private IPOs and the Changing Markets. *Journal of Applied Corporate Finance* 3 (summer 2015): 34-48.

Laura Huang and Jone Pearce. 2015. Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions. Johnson Cornell University. Published in July

Liz Knueven. 2021. The average stock market return over the past 10 years. Business Insider. Published in June 2021
https://www.businessinsider.com/personal-finance/average-stock-market-return

Margaret Manson. 2021. The Unicorn Seduction of the B2B SaaS Market. InnoFuture.Published in 2021
https://innofuture.com.au/b2b-saas-market-unicorn-seduction/

Moreno Bonaventura, Valerio Ciotti, Pietro Panzarasa, Silvia Liverani, Lucas Lacasa, and Vito Latora. 2019. Predicting success in the worldwide start-up network. Scientific Reports (January).
https://arxiv.org/pdf/1904.08171.pdf

Niels Bosma, Mirjam van Praag, Roy Thurik, and Gerrit de Wit. 2004. The Value of Human and Social Capital Investments for the Business Performance of Start-Ups. *Small Business Economics volume* (February): 227–236

Nuno Caseiro, and Arnaldo Coelho. 2019. The influence of Business Intelligence capacity, network learning, and innovativeness on startups performance. *Journal of Innovation and Knowledge* (April): 139-145

Pitchbook. Venture Capital, Private Equity, and M&A database. https://pitchbook.com/

Puja Abbassi, Daniel Schlagwein et a. 2016. Business Model Development, Founders' Social Capital and the Success of Early-Stage Internet Start-Ups: A Mixed-Method Study. *Information Systems Journal* 26 (September): 421-449

Rebecca Baldridge, and Benjamin Curry. 2021. What is a Startup? *Forbes* (April). https://www.forbes.com/advisor/investing/what-is-a-startup/

Robert N. Lussier. 2001. A Crossnational Prediction Model for Business Success. *Journal of small business management* (July): 228-239

T. Butler, and Brian Fitzgerald. 2000. Unpacking the systems development process: an empirical application of the CSF concept in a research context. *Journal of Strategic Information Systems* 8 (May): 351–371. https://www.brian-fitzgerald.com/wp-content/uploads/2013/10/but-fitz-99-jsis.pdf

The Asian Banker. Statistical information. https://www.theasianbanker.com/updates-and-articles/almost-40-of-current-unicorns-were-created-in-the-first-eight-months-of-2021

US Small Business Administration report. 2019. Published in April 2019. https://cdn.advocacy.sba.gov/wp-content/uploads/2019/04/23142719/2019-Small-Business-Profiles-US.pdf

Yi Gin. 2021. SaaS Trends to Watch In 2021. Forbes. Published in February https://www.forbes.com/sites/forbescommunicationscouncil/2021/02/24/saas-trends-to-watch-in-2021/?sh=3f26d6605385