Exploring the Differences between Successful and Unsuccessful Startups

Arandra Kalista Malkan, Andra

Award date:
2023

Awarding institution:
Tilburg University

Exploring the Differences between Successful and Unsuccessful Startups

Bachelor Thesis

Arandra Kalista Malkan (1590111)

Supervised by:

Dr. Werner Liebregts

Full Thesis

Eindhoven, June 2023

Abstract

Startups play a crucial role in driving economic growth, making them a subject of importance. Despite extensive research conducted in this field, it remains highly relevant due to the ever-changing ecosystem and the persistent rate of startup failure. This paper aims to investigate the differences and factors of successful and unsuccessful startups. Firstly, the definitions of successful and unsuccessful startups were provided. Focusing on early-stage startups, successful startups were defined as startups who have secured funding in the growth stage. Unsuccessful startups were defined as startups who failed to progress beyond the early stage. By utilizing Dealroom, a startup data platform, data on successful and unsuccessful startups were obtained. Adopting an exploratory approach and conducting analysis in the form of hierarchical regression, insights on startup success and failure were gained.

1. INTRODUCTION

Every startup wants to achieve success and become a household name in its industry. However, the road to success is never easy, and many startups experience failure along the way. In fact, it was reported that 9 out of 10 startups fail (Kotashev, 2022). According to the U.S. Bureau of Labor Statistics, roughly 20% of new businesses fail within their first year. The chances of reaching unicorn status is even slimmer, with only 1% of startups achieving success as massive as Uber and Airbnb (*The Venture Capital Funnel*, 2018).

The topic of startup success has been researched for many decades. Researchers have tried numerous approaches and obtained a variety of results in their attempts to determine the factors that lead to startup success. Research has focused primarily on the performance of surviving firms, limiting our understanding of startup failure (D. Shepherd & Wiklund, 2006). However, given the prevalence of startup failures, it is crucial to examine the disparities between success and failure in order to gain valuable insights.

It is also crucial to acknowledge that startups progress through various stages in the startup lifecycle, each requiring different sets of skills and considerations. For example, obtaining a seed funding has different criterias than achieving a series funding. This paper focuses on analyzing early-stage startups and identifying which factors affect their progression and failure to the subsequent stage. Something that early stage startups often struggle with is the “Valley of Death”. During this phase, a startup tends to operate without any revenue from products and services, relying mostly on their initial invested capital (*How Can Start-Ups Overcome the “Valley of Death”?*, 2020). Surviving this valley is a major milestone for startups.

Commonly, the next step for them is to accumulate capital, which has been proven to be one of the major hurdles for early-stage startups. Without adequate funding, ideas developed by entrepreneurs that have potential to be groundbreaking will not be successful (Bauer et al., 2023). According to CB Insights (CB Insights, 2021), the largest reason why startups fail (38% of cases) is due to failing to raise a new capital. Furthermore, fewer than 10% of startups that raise a seed round successfully raise a Series A investment (Ruby, 2022). Acquiring funding from investors in this stage is crucial. Investors are usually known to bet on the founders rather than the business plan (Gladstone & Gladstone, 2002; Harrison, 2017). Irrespective of the horse (product), horse race (market), or odds (financial criteria), it is the jockey (entrepreneur) who fundamentally determines whether the venture capitalist will place a bet at all (Macmillan et al., 1985). But how true is this statement? Alternatively, other studies found that investors put more weight on the opportunity rather than the founder (Kaplan et al., 2007).

Lastly, the field of entrepreneurship research has been experiencing significant growth and expansion owing to big data availability and the rise of the start-up ecosystem. Data platforms such as Crunchbase and Dealroom provide information and analysis on millions of startups and related companies. These types of platforms have led to breakthroughs in startup research, allowing for attainment of longitudinal information, identification of patterns and trends, prediction of venture funding using machine learning among other things (Kaplan et al., 2007).

1.1. **PROBLEM STATEMENT AND RESEARCH QUESTION**

The emergence of innovative companies are one of the engines of global economic growth, especially in developing countries. Due to technological advances and economic globalization, there exists a conducive environment where entrepreneurs can more readily develop innovative products and influence change in the market. However, the sad reality is that although it's easier to start a business today, it's not easy to attain long-term success. There are thousands of startups that get lost in the dark and never see the light of day. Only one in 1000 firms achieves a successful financial exit such as an IPO or high-value acquisition, and a very small fraction of firms is ultimately responsible for the vast majority of growth in revenues and employment (Catalini et al., 2019). It's difficult for startups to keep up with the challenges posed by the ever-changing environment. Furthermore, persistent issues such as building a viable business model, having a cohesive team, and ensuring a product-market fit are very much still prevalent problems startups face. This underscores the ongoing relevance and importance of researching startup success and failures. Therefore, the goal of this study is to provide a comprehensive comparison between startup success and failure. This led to the main research question, supported by two sub-questions:

MAIN RESEARCH QUESTION: To what extent do successful startups differ from unsuccessful startups?

SUB QUESTION 1: How can successful and unsuccessful startups be defined and measured?
SUB QUESTION 2: What are the factors affecting successful and unsuccessful startups?

This study contributes and extends to the literature in at least three ways. First, a comparison of factors distinguishing unsuccessful and successful startups was done. Most previous studies focus on either failed startups or successful startups, with more research on successful startups. These studies came with an inbuilt limitation on generalizability (Pattanayak & Kakati, 2021). Furthermore, the majority of research excludes failed cases from analysis. Second, a startup data platform was used to obtain the data used for this study. Surveys, interviews and case studies have always been methods at the forefront when it comes to analyzing startup success (Baum & Silverman, 2021; Macmillan et al., 1985). This paper adopts an exploratory approach that primarily relies on the dataset rather than pre-selecting specific factors for analysis. Lastly, this study is not limited to location or industry unlike most other research papers regarding this topic.

Ultimately, the findings of this research will provide valuable insights for entrepreneurs, investors, and policymakers seeking to improve the success rate of startups.

2. THEORETICAL BACKGROUND

2.1. Startup Definition

What exactly is a startup? Is it merely a new company? Is it a company that is driven by innovation and technology? The definition of a startup has long been debated and discussed. Although a growing number of scholars have been trying to capture and describe the unique specific characteristics of the startup phenomenon in recent years, there is still no general agreement as to what a startup is—both among scientists and business support institutions and among entrepreneurs themselves (Skala, 2018; Kézai et al., 2020).

The concept of startups has evolved over time, with different definitions emerging as the world of business and technology has changed. Before the term "startup" became more widespread in the second half of the 20th century, Schumpeter used the term "innovative entrepreneurs" in his seminal work "The Theory of Economic Development" (Schumpeter, 1934) to explain the phenomena of entrepreneurs who introduced new products, services, and methods that disrupted the status quo and created new opportunities for economic activity. In the original sense, the word “startup” meant any form of business in its early stage of development (Skala, 2018). The beginnings of this change dated back to the 1970s. The word ‘startup’ started being associated with new businesses in the high-technology fields (Lebret, 2016).

The most popular definition of a startup, widely cited in industry publications and scientific literature, was formulated by Steve Blank. According to Blank, a startup is an organization formed to search for a repeatable and scalable business model. A business model describes how the company creates, delivers and captures value. Most startups change their business model multiple times until it reaches the one to scale (Blank, 2010). Blank emphasized the importance of experimentation and iteration in the startup process and the need for startups to be flexible and adaptable in response to changing market conditions. Another well-known definition of startups was formulated by Paul Graham, who gave a short and concise definition on startups: “A startup is a company designed to grow fast” (Graham, 2012). This can mean a rapid increase in employees, revenue, or customers which will ultimately lead to an increase in the company's value.

Although there are numerous perspectives on what defines a startup, research is consistent to a certain degree. There seems to be common themes that emerge when discussing startups, which can be summarized in the following sentence: A startup is a young, high-risk company that develops innovative products and is scalable and fast-growing. The definition of start-ups in this paper is discussed in **section 3.1.**

2.2. Startup Stages

Many experts have proposed various frameworks and models that outline the stages a startup goes through. While there might be differences depending on the context, several key phases can be concluded from these frameworks.

The first stage of a startup is the *concept stage*. This stage begins with an idea of a product from the entrepreneur. The entrepreneur checks if there's a problem/solution fit and a market for their idea. This can be done by conducting research and talking to potential customers. They will then validate their product and business concept by making early prototypes. The prototype does not need to be functional nor viable (*CEMEX Ventures*, 2021). The organization of a startup at this stage is typically informal, loosely structured, and fluid. Resources are also limited. At this stage, the forms of financing are usually bootstrapping, FFF (friends, family, and fools) and angel investors.

The next stage of a startup is the *seed stage*. At this stage, a minimum-viable-product (MVP) is built to further develop the product, validate the idea, and find a product-market fit. A MVP is a development technique in which a new product or website is developed with sufficient features to satisfy early adopters (Członkowski, 2023). Feedback from the early adopters will be used to improve and iterate the product. Building an MVP is an important step of the lean startup methodology, which is an approach to starting a new business that emphasizes speed, experimentation, and learning (Ries, 2023). This stage is usually where a startup has its first sales and expands their customers. The forms of financing for this stage are angel investors, seed funding from venture capitalists, accelerators and incubators.

The third stage of a startup is the *growth stage*. A startup may be referred to as a scale-up at this stage. This stage can be categorized into two phases: Early growth and late growth. Early growth is usually indicated by Series A funding. Series A funding is the first venture capital backed funding that allows angel investors to exit the startup. Late growth is indicated by the subsequent series funding after Series A. These two phases will be combined into the growth stage. The general objective of this stage is to scale its operations and consolidate growth in both revenues and employees. The entrepreneurs must add significant resources to profitably scale the startup. Consistent profitability is required to provide a return for investors and fund the drive to market leadership. The loosely-structured organization is not effective any longer, and will be replaced with a structured and disciplined organization (Picken, 2017). The venture capitalists investing in series funding take center stage during this phase. Additionally, growth equity are also forms of funding in this stage.

The last stage of a startup is the *mature stage*. In the mature stage, startups have become well-established companies and have a stable position in the market. Private equity is a form of investment at this stage. At this stage, a startup is completely staffed and has reached significant growth and are looking for expansion opportunities. Commonly, a successful exit in the form of IPO, private sale, or Merger & Acquisitions is the next step as it is required to harvest the value accumulated by the venture for the benefit of the entrepreneur and investor. Many startups consider this as the *ultimate* success. However, most startups don’t ever reach this achievement.

This paper focuses on analyzing factors affecting startup failure at the seed stage and compares it to startups who are in the growth stage. This is relevant due to the “Valley of Death'' early-stage startups face. The ‘Valley of Death’ describes a gap between low-ticket business angel investors and more institutional investors, such as Venture Capitalists (Hermann, 2022) (Figure 1). According to one estimate, only 40% of early stage internet start-ups are able to secure required Series A funding (Spiegel et al., 2016). Given the high failure rate among startups at this stage, it is necessary to investigate the factors that contribute to success and failure in progressing to the growth stage.

The measures of startup success and failure chosen for this paper will be based on this scope.

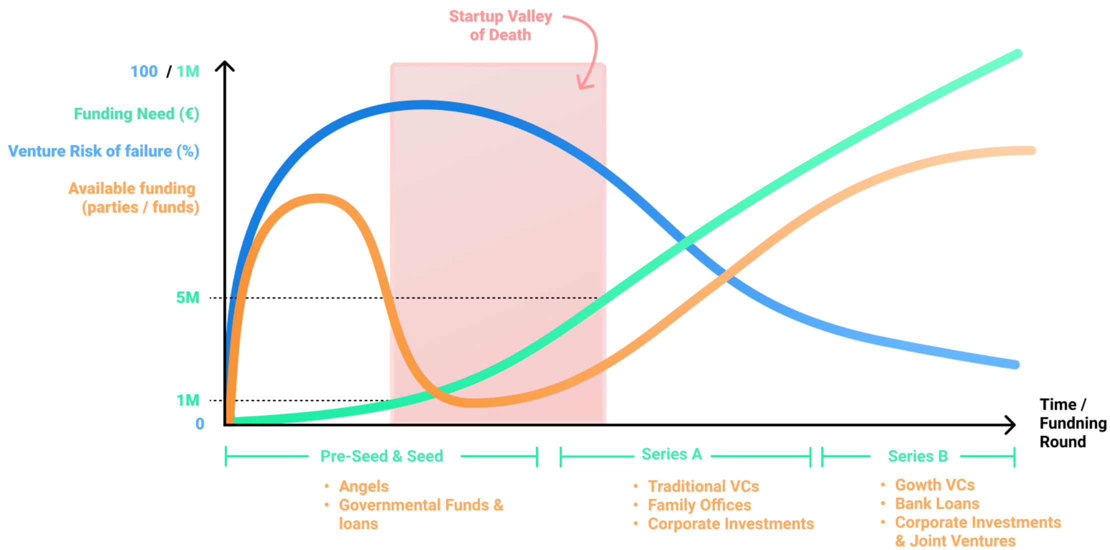

Figure 1. The Valley of Death (Hermann, 2022)

2.3. Measures of Startup Success and Failure

2.3.1. Startup Success

Researchers need to define what constitutes a startup success when doing an analysis. There are many different factors that could be used to evaluate startup success. It's difficult to get an objective measurement of startup success as entrepreneurs and investors have various objectives and goals in mind. In this section, an overview of measurements of startup success will be provided, and what constitutes a startup success for this study will be given.

As discussed in the previous section, an initial public offering of success is a crucial moment for a startup. It generates interest and can be a signal to the top talent in the industry that the company has made it (*How the IPO Process for Startups Works*, n.d.). Ultimate success implies the "cashing out" of the investment through an IPO or an acquisition (Croce et al., 2016). However, not every entrepreneur has grand goals such as reaching unicorn status.

Two distinct generational profiles of entrepreneurs who hold contrasting views on what constitutes a startup success are identified. Entrepreneurs born after the 20th century emphasized more on personal and societal satisfaction. They define success as having freedom to do what you want, being your own boss, and spending time with your family (McGowan, 2015). Making a positive impact on other people's lives through their startup is also a priority for them (Santisteban & Mauricio, 2017). On the other hand, older entrepreneurs stuck to a more traditional view of success, defining success as earning money and growing a scalable business.

Additionally, some entrepreneurs describe startup success as achieving the company's goals and having effective management (Santisteban et al., 2021). Furthermore, researchers have measured startup success through financial performance, market share, number of jobs a company created, and number of customers a company attracts and retains (Azimzadeh et al., 2013).

For this paper, measures such as personal satisfaction and social goals are hard to measure and do not relate to the study at hand. Furthermore, while company achievements may serve as a potential metric for evaluating startup success, they are often subject to challenges related to their objectivity and measurability. Measuring startup success based on funding rounds is more observable, feasible, and objective. Due to the fact that this study focuses on early-stage startups advancing to the growth stage, the measure of start-up success for this study is: A start-up will be deemed as successful if it has secured funding in the growth stage (Series funding and growth equity VC).

2.3.2. Startup Failure

Similar to startup success, the definition of startup failure is elusive and lacks a definitive consensus. In this section, an overview of measurements of startup failure will be provided, and what constitutes a startup failure for this study will be given.

Some researchers see failure as a temporary phase in an ongoing entrepreneurial process which can be used as a valuable source of learning and improved self-awareness (Politis & Gabrielsson, 2009). This perspective gives the idea that repeated failures occur naturally in the process of a new venture creation. For example, a software startup going through iterations of technical challenges. However, this definition is extremely subjective and not measurable.

Three definitions of startup failures that are more objective can be identified: discontinuity of business, bankruptcy, and discontinuity of ownerships. Bankruptcy is a well-known definition of startup failure due to its clarity and observable nature. There are rarely any cases where firms recover from bankruptcy. On the other hand, the ‘discontinuity of ownership’ definition is vague since it includes every business that was sold, including businesses that were sold due to success. Alternatively, Shepherd (2003) proposed a definition where a startup failure ‘occurs when a fall in revenues and/or a rise in expenses are of such magnitude that the firm becomes insolvent and is unable to attract new debt or equity funding; consequently, it cannot continue to operate under the current ownership and management’.This definition does not include firms that were sold due to success but rather due to failure in attracting funding. The definition of Amankwah-Amoah (2016) will be used in this paper: ‘Organizational failure is defined as a situation where the firm ceases operations and loses its identity due to inability to respond and adapt to changes in the external environment in a timely fashion’. This definition is chosen due to it being observable, clear, and can be implemented to the data with the most ease. Lastly, to align with the scope of this paper, we analyzed startups that ceased operations during the seed stage.

2.4. Startup Performance Factors

In the case of this paper, we forgo presenting formal hypotheses to be tested. Since the measurement of success and failure is dependent on fundings from investors, this section will give a general overview of startup performance factors based on the investors’ and founders’ perspective. **Section 3.4** further shows the factors we have chosen, and the ‘Discussion’ section provides a comparison between the results and previous research.

2.4.1. Investors’ Perspective

As stated before, investors usually bet on the founders rather than the product. Numerous research supports this statement. For example, Mason et al., (2016) found that the entrepreneur/team is the key reason for rejecting investment opportunities. Both interviews and online surveys are used to examine the reason and were conducted with business angels. They found that it is crucial for the entrepreneurs to be open, straightforward, believable, trustworthy, honest, and appear knowledgeable. They should not be unrealistic regarding valuation and equity share. These kinds of soft decision factors play an important role in the BA investment process because of the hands-on-role BAs take in the investee venture (Granz, Henn, & Lutz, 2019). Overall, there is a clear consensus that the reasons for rejecting investment opportunities are widely associated with perceived weaknesses in the entrepreneur and management team. Meanwhile, venture capitalists focus more on an experienced management team. With an experienced management team, future failure risk of an investment can be moderated and increase potential of generating higher returns on the venture capital's investment. Franke, Gruber, Harhoff, & Henkel (2008) conducted a conjoint experiment with 51 German and Austrian professionals in VC firms. Their results indicated that industry experience, educational background, and leadership experience are the three most important team characteristics. Another study (Fried & Hisrich, 1994) also concluded that founders are an important factor for venture capital investors. Investors pay significant attention to the manager's leadership, flexibility, and past experience, and the manager's ability to identify risk and make realistic forecasts. These skills are usually cultivated through experiences.

On the other hand, KAPLAN et al.,(2009) produced different findings. They concluded that the business is more important than the founder's characteristics. Based on the IPO sample, firms that go public rarely significantly change from their initial business idea or line of business. An initial strong business, although not sufficient, appears to be necessary for a company to succeed. Meanwhile, it is common for firms to replace their founders and still go public. Another more recent study (Bernstein et al., 2014) has similar findings. Based on a sample of 4500 investors in young companies, they stated the importance of human capital assets for the success of early stage firms. The founding team is crucial at the beginning for differentiating the company, for developing the products and the market. However, as the firm develops and undergoes a standardization phase, human assets are replaceable and principal business lines such as patents, technology, and physical assets are the key to success. Agrawal et al., (2019) also showed that founders who registered in Delaware and acquired a patent and trademark protection were "278 times more likely to achieve an equity growth event than firms that are associated with none of these choices".

These findings further emphasized that different stages in the startup life cycle require different criterias. Investors pay significant attention to founders and the managerial team in the early stage. However, the business operations matter more once you have passed the early stage.

Lastly, market and growth potential are the next important factors after founders and managerial teams. Silva (2004) focused on venture capitalists' decision-making process, revealing that their attention primarily centers around the entrepreneur, the business idea, its sustainable advantages, and growth potential. Interestingly, the study highlights that financial projections of the prospects do not appear to significantly influence the selection of early-stage startups.

2.4.2. Founders’ Perspective

There are less studies about the founder’s perspective compared to the investor’s perspective. Similar to investors, research shows that the founding team is the most important factor for startup success, specifically for startups who are in the early stages. Prohorovs et al., (2018) questioned founders of 40 startups in Latvia and Russia and concluded that startups who are successful in attracting seed funding are managed by entrepreneurs who have had previous experience in creating business entities and are capable of building a team with employees who have appropriate experience, specialized education and high-level management skills. Furthermore, Zaheer et al. (2009) drew key insights from 12 digital start-ups in the form of interviews and public documents. They found that combined experiences and wisdom on human and social capital has a positive impact on the success of start-ups. Spiegel et al. (2015) also emphasized on high social capital of the founders in early stage internet start-ups as important factors for startup success.

Another study by HBR obtained input from 141 Harvard Business School alumni founders, most of whom lead venture capital–backed technology startups. Results show that assembling a founding team is the highest-priority skill for a future technology leader. The founding team is the key asset as they are the ones responsible for developing the idea and product together and having the motivation and skills to manage the startup. However, during the early stage, founders emphasized on product scalability and financial resources to ensure future growth. The amount of financial resources the founders have at their disposal is an important factor for investment decisions according to founders. This is contradictory to the investments' perspective, especially for venture capitalists who are not limited in financial resources. Furthermore, founders see product scalability as barriers in the early stage whereas investors see business model, high competition and managerial skills as barriers.

3. METHODOLOGY

3.1. Data Collection

Dealroom is the main source of data in this study. Dealroom is a data provider on startups, growth companies, and tech ecosystems around the globe. They combine machine learning and data engineering with robust verification processes and a strong network of ecosystems to collect their data (*How Dealroom Collects Data*, n.d.). Dealroom uses Paul Graham's definition of a startup, which is “A startup is a company designed to grow fast”. Furthermore, they qualify startups based on the following characteristics:

1. Rapidly scaling/scalable
2. Founded after 1990
3. Innovative by design: the product and/or business model are innovative. In most cases, the company is tech-enabled: proprietary tech/software or business processes heavily enabled by tech.

Their definition is similar to what we've concluded in the start-up definition section. However, we stated that a start-up is a young company. We deviate from their second point, and only limit companies from 2013 up until 2023. Older startups may not serve as an accurate reflection of current startups, given the technological advancements that have emerged.

By creating multiple accounts and using Dealroom's free trial, we were able to download samples of startups through an iterative process. The data can be exported 25 rows at a time. Dealroom contains an “advanced filter” option that is helpful for us to get a match on our definition of startup success and failure. Due to the nature of the “advanced filter” option and dataset, a stratified random sampling approach was chosen. For startup success, the “Last Funding Round” column was filtered to series funding and growth equity VC, and the “Company Status” column was filtered to “operational”. For startup failure, the “Last Funding Round” column is filtered to seed and angel investors, and the “Company Status” column is filtered to “closed”. For both types of startups, the age of the company is limited to companies who launched from 2013 until 2023. Lastly, the columns were filtered in a way that minimizes missing values for both types of startups.

After analyzing the first iteration of exporting startup failures, we found that the closed startups are not all accurate as some are still operating. To tackle this problem, we checked if the startups are actually shut down by verifying with other reliable sources. Occasionally, the startup and/or the startup founder's own platform/social media (LinkedIn, Twitter, etc) explicitly stated that they have shut down their operations. If this is not the case, the diagram below gives the process of investigating whether the startup has shut down or not (Figure 2).

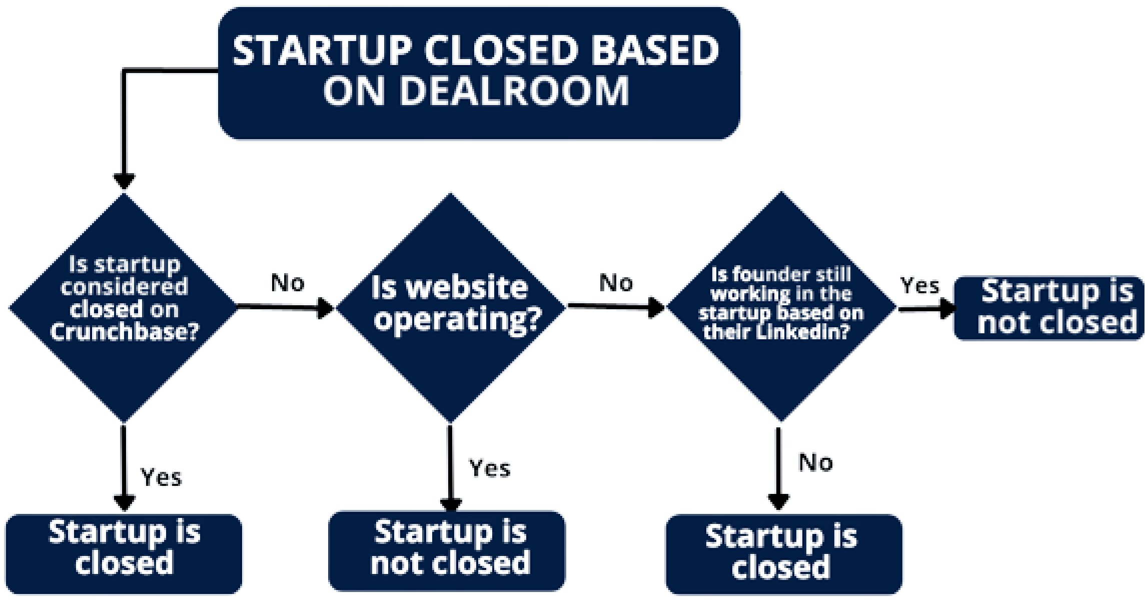

Figure 2. Diagram outlining the verification of closed startups

We manually verified 10% of the dataset for the columns containing organizational and business model characteristics with other startup data platforms and the startup's own website and found 96.25% accuracy. Dealroom seems to not have a problem in obtaining accurate information for these columns. On the other hand, we found 79% accuracy in 10% of the dataset for the column containing the founders' names. Dealroom is not exhaustive in listing the founders of a startup. However, we argue that this accuracy is still acceptable.

3.2. Data Preprocessing

The process of acquiring and cleaning the data was a non-trivial task. Although using powerful algorithms to gather extensive and real-time data is beneficial for exploratory analysis, it also comes with disadvantages. Inaccuracies, incompleteness, and biases may arise with this method. In this section, we provide the preprocessing of the data.

3.2.1. Founder Characteristic Columns

The 'FOUNDERS' column, indicating the name of the founders, contained some duplicates of founders' names. However, after further analysis, removing exact duplicates won't entirely resolve the issue. Some columns contained the same founder names more than once, but with variations such as the inclusion or omission of their middle names. To solve this problem, the Levenshtein distance was used to measure how different two founder names are. It is defined as the minimum number of changes required to convert string a into string b, which is done by inserting, deleting or replacing a character in string a (*The Levenshtein Distance Algorithm*, n.d.). The smaller the Levenshtein distance, the more similar the strings are. Before calculating the Levenshtein distance, the strings were put into lowercase first. If the strings have a similarity score of 75 or more, only one of the names was regarded. Finally, a 'num_founders' column for the total number of founders was calculated.

The 'FOUNDERS BACKGROUND' column contained information about the founders' higher education background. This column often had at least one information about a founder in a startup missing. It also came to our attention that the founder's background is relatively recent. This posed a challenge as we aimed to compare a startup that has ceased operations and a startup that is still active. In order to ensure meaningful comparisons, it is crucial to get data points that are comparable between the failed and successful startups. The founders' backgrounds at the time of founding the startup serves as a pertinent data point. To solve this problem, we manually fill this column by going through the founders' linkedins. The values of this column are limited to 'Business', 'Technical','Others', and 'Unknown'. Founders with no higher education were not taken into account.

The 'FOUNDERS FIRST DEGREE' column also lacked complete data. Furthermore, Dealroom provides data of the first degree the founder achieved. It is more relevant to include the degree the founder possesses at the time of founding the startup. To solve this problem, we again manually filled this column by going through the founders' linkedins. The values of this column were limited to 'High School Diploma', 'Bachelors', 'Masters', and 'PhD'.

3.2.2. Additional Columns

First, utilizing the 'OWNERSHIPS' column, we examined the investor types of the startups. A dedicated column was created specifically for accelerators, having 1 if the 'OWNERSHIPS' column contains an accelerator, and 0 if not.

Third, Dealroom contains a 'ALL TAGS' column that provides detailed information about the startups. One of the most used tags in the platform is 'sustainable development goals'. Furthermore, Dealroom also contains a 'SDG' (sustainable development goals) column describing the specific SDG goal a startup focuses on. Leveraging these two columns, we derived a binary variable describing whether a startup addresses a sustainable development goal. A value of 1 was assigned if the startup had a 'sustainable development goals' tag or the SDG column value for that startup is not null, we put 0 otherwise.

Fourth, the Dealroom data also contains information on the headquarter country of a startup. Making this column into dummy variables will result in too many variables, rendering the data sparse, especially since some countries only have 1-2 occurrences within the whole dataset. Therefore, we resorted to generating another column that discerns if the country is developing or developed. This is relevant due to the fact that developed and developing countries play a significant role in the startup ecosystem. We used Worlddata's list of developing countries. Worlddata serves a comprehensive database for geographic, climatological and demographic data (*List of 152 Developing Countries*, n.d.).

Lastly, GEM data (*GEM Global Entrepreneurship Monitor*, n.d.) was used to obtain the internal market dynamics and entrepreneurship finance of a country. Internal market dynamics refers to the level of changes in markets from year to year and entrepreneurship finance refers to the availability of any financial resources for new and growing firms. GEM obtains their data through questionnaires administered to a minimum of 36 carefully chosen experts, who are asked to respond to a series of statements on a Likert scale, rating them from completely false to completely true. GEM provides a score from these questionnaires every year. The data was merged with our dataset on the country and launch year of the startup.

3.3. Descriptive Statistics [1]

Launch Year
Figure 2 displayed the amount of successful and unsuccessful startups by launch year. The period between 2013-2016 contained a high proportion of unsuccessful startups. However, unsuccessful startups started to consistently decrease from 2017.

Industries
There were a total of 28 industries in the whole dataset, with 1 occurrence being null. A large part of the startups were active in the Fintech and Enterprise Software Industry. To avoid sparse data, only the industries that had an occurrence of more than 10 were selected (see Figure 3).

Revenue Model
There were a total of 184 filled values and 22 null values in this column. SaaS (Software-as-a-Service) had the most occurrences with 109 values, with marketplace & ecommerce coming in second (51) and manufacturing (40) last. The null values were changed to 'unknown'. It is important to note that the missing values can indicate two things. First, Dealroom cannot obtain information for the startup. Second, since Dealroom is limited to only three types of revenue model, the startup's revenue model might not align with the available values of Dealroom's database.

Income Streams
For this column, there were 170 filled values and 36 null values. 'Commission' (90) and 'subscription' (72) were the leading values followed by 'selling own inventory' (11) and 'advertising' (3). Due to the 'advertising' value only having 3 occurrences, of which are all failures, it was dropped to avoid large standard errors. The null values were changed to 'unknown'. The missing values in this column can also be interpreted in the same logic as the revenue model column.

HQ Region
Headquarter region Europe (105) contained the most number of startups, followed by North America (74) and Asia (18). Europe and North America were kept due to the large amount of values. The two regions encompassed 86.9% of the startups in the dataset. The rest of the regions were put as 'others'. Dealroom was founded and operates in Europe, which might give a reason as to why Europe is the most frequent region.

[1] Some startups have more than one type of revenue model, income stream and industry

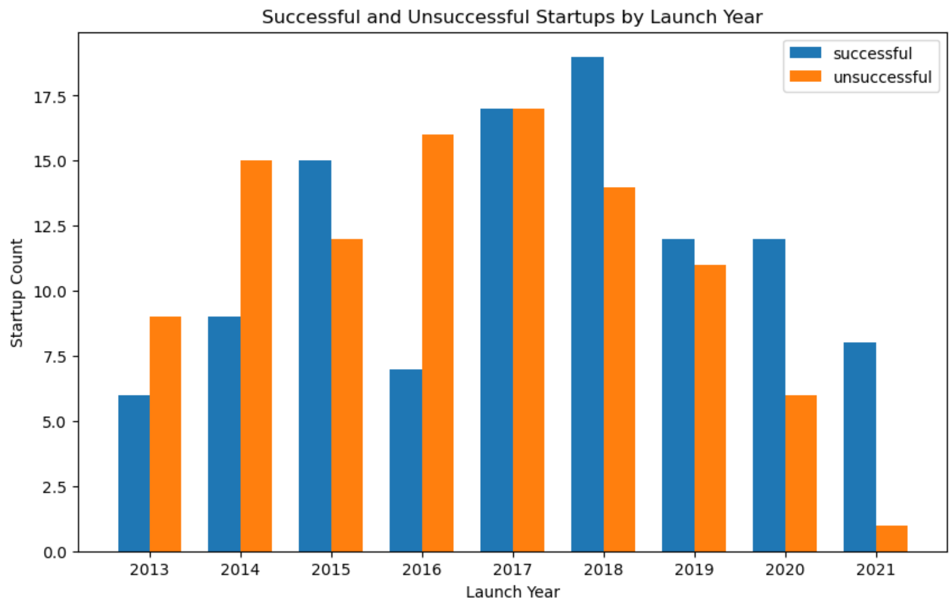

Figure 3. Bar Graph of Successful and Unsuccessful Startups by Launch Year

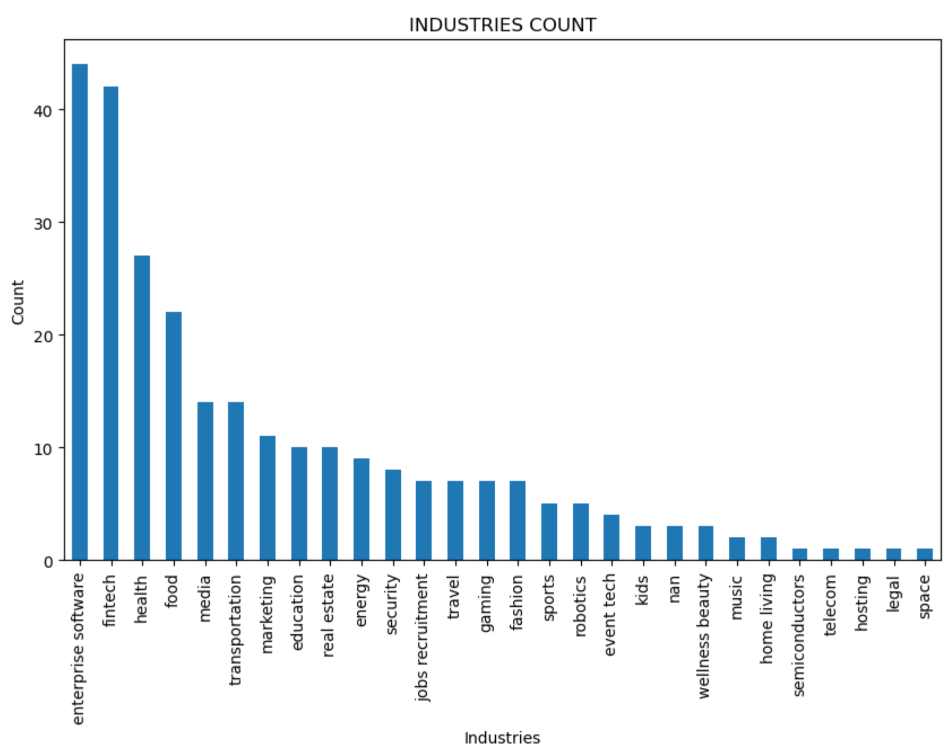

Figure 4. Bar Graph of Startup Industries

3.4. Variables

While Dealroom offers an extensive amount of data, careful consideration was done to mitigate the potential issue of reverse causality. For example, the number of employees, profit and twitter followers can be influenced by the startup's progression and scaling rather than them causing success in funding rounds. Although there is historical data for these features, it is limited to a few years, making it unsuitable for startups that were founded before that period. Furthermore, financial features like revenue and profit have mostly missing values, as startups do not typically disclose this information widely. Lastly, columns were investigated beforehand to ensure accuracy and determine the extent to which they can be used for a fair comparison between unsuccessful and successful startups. Taking into account these considerations, selecting and adding features based on the Dealroom data was done carefully, and the summary of the final model variables can be seen in Table 1 below. The summary statistics of the model variables are given in Appendix A. The dataset consisted of 101 unsuccessful startups and 105 successful startups.

Column Name	Column Description	Input	Type	Characteristics
B2B/B2C	Whether the startup is business-to-business or business-to-consumer	Made into dummy columns: 'business', 'consumer', 'business;consumer'	Categorical, Independent	Business Model
REVENUE MODEL	The revenue model of the startup. Dealroom limits the values to 'saas, manufacturing, and marketplace&ecommerce'	Made into dummy columns: 'saas', 'manufacturing', marketplace&ecommerce', 'unknown'	Categorical, Independent	Business Model
INCOME STREAMS	The income stream of the startup. Dealroom limits the data to 'commission, advertising, and subscription'.	Made into dummy columns: 'commission', 'advertising', 'subscription', 'unknown'	Categorical, Independent	Business Model
NUMBER OF FOUNDERS	The amount of founders the startup has	-	Numerical, Independent	Founder
IS FOUNDERS FIRST COMPANY	The founder's previous founding experience	Made into a binary variable: 1 if at least one founder has founded a company previously, 0 otherwise	Categorical, Independent	Founder
FOUNDERS GENDER	The gender of the founder	Made into a binary variable: 1 if at least one	Categorical, Independent	Founder

		founder is female, 0 otherwise		
FOUNDERS DEGREE	The academic degree of the founder at the time of founding the startup	Made into an ordinal variable: 0 as 'High School or under', 1 as 'Bachelor or under', 2 as 'Masters or under', and 3 as 'PhD or under'	Categorical, Independent	Founder
FOUNDERS BACKGROUND	The higher-education background of the founder at the time of founding the startup.	Made into dummy columns: 'business', 'technical', 'others', 'unknown'. A value of 1 was assigned if at least one founder has the background, 0 otherwise	Categorical, Independent	Founder
SUSTAINABILITY	Whether or not the startup addresses sustainable development goals	Made into a binary variable: 1 if startup has a sustainable development goal, 0 otherwise	Categorical, Independent	Others
ACCELERATOR	Whether or not the startup had support from an accelerator	Made into a binary variable: 1 if startup had support by an accelerator, 0 otherwise	Categorical, Independent	Others
INDUSTRIES	The industries the startup operates in	Industries with 10 or more occurrences are made into dummy columns	Categorical, Control	External
COUNTRY STATUS	Whether the startup's headquarter is in a developing or developed country	Made into a binary variable: 1 if country is developed, 0 if a country is developing	Categorical, Control	External
HQ REGION	The headquarter region of the startup	Made into dummy columns: 'North America', 'Europe', 'Others'	Categorical, Control	External
INTERNAL MARKET DYNAMICS	Levels of changes in a market from year to year in a country	Scoring based on Likert scale (1-9)	Numerical, Control	External
FINANCING FOR ENTREPRENEURS	Availability of financial resources for new and growing firms in a country	Scoring based on Likert scale (1-9)	Numerical, Control	External
TARGET VARIABLE	Whether the startup is successful or not	Made into a binary variable: 1 if a startup is successful, 0 if startup is unsuccessful	Categorical, Dependent	-

Table 1. Summary of Model Variables

3.5. Model

Since the dependent variable was binary and the aim was to interpret the relationship of the independent and dependent variables, logistic regression was a viable option for the study. The equation of the logistic regression is as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots..$$

- p is the predicted probability of the outcome
- β_o is the intercept of the model
- $\beta_1, \beta_2 \ldots.$ are the coefficients assigned to each predictor $x_1, x_2 \ldots$
- $x_1, x_2 \ldots$ are the predictor variables

Since we have many variables, a hierarchical logistic regression was utilized. Hierarchical regression is used to investigate the incremental contribution of predictor variables and see if the addition of variables helped improve the overall explanatory power of the model on the likelihood of startup success. First, the control variables - external factors - were included in the first model. Next, the founder characteristics were inserted. The third model block included the business model characteristics. Lastly, the variables ‘sustainability’ and ‘accelerator’ were included for the final full model.

4. RESULTS

4.1. Hierarchical Logistic Regression

Before running the logistic regression, certain assumptions need to be met and tested beforehand.

First, the multicollinearity assumption was checked by first creating a correlation matrix and then evaluating the Variance Inflation Factor (VIF) of every predictor variable in the model (see Figure B.1 in Appendix B). 'B2B_business' and B2B_consumer' were highly correlated with a score of 0.81. 'HQ_REGION_NORTH_AMERICA' and 'HQ_REGION_EUROPE' were also correlated with a score of 0.76. 'HQ_OTHERS' and 'country_status' were correlated with a score of 0.71. Hence, 'B2B_consumer', 'country_status' and 'HQ_REGION_NORTH_AMERICA' were dropped. Next, we tested multicollinearity with VIF. Most commonly, a VIF of 10 is recommended as the maximum value (Marquardt, 1970). The 'Income Streams' dummies seemed to be heavily correlated as their VIF ranges from 15-21. Therefore, we dropped the 'income_streams_unknown' variable, which thereafter resulted in a VIF of less than 10 for all predictors, with the highest VIF being 5.4 (see Figure B.2 in Appendix B).

Second, the assumption of a linear relationship between the continuous independent variables and the log-odds of the dependent variable was tested using the Box-Tidwell test. The null hypothesis states that there is a linearity between the independent variable and the logit of the outcome variable. The continuous variables in the data were 'num_founders', 'Financing for entrepreneurs', and 'Internal market dynamics' and their p-values are 0.89, 0.89, and 0.064 respectively. This indicated that the null hypothesis was not rejected at a 5% significance level and the assumption was satisfied.

Third, outliers (see Figure B.3 in Appendix B) were detected by constructing a box plot for the numerical variables. The median and mean for the 'num_founders' column were 2 and 2.09 respectively. One outlier was detected with the value of 7. The outlier was inspected and found to be correct and was not removed. Since 'Financing for entrepreneurs' and 'Internal market dynamics' are scoring based on a Likert scale of 1-9, the outliers were also kept.

After assumptions were satisfied, we conducted the hierarchical regression. With this result, we were able to answer the second sub-question: *What are the factors affecting successful and unsuccessful startups?*

The goodness-of-fit of the models were analyzed with the McFadden R-squared and AIC. The lower the AIC, the better the model. McFadden's pseudo R-squared value between 0.2 to 0.4 indicates excellent fit. Comparison between the current model block and the previous model block was conducted with the likelihood ratio test. Lastly, the coefficients were exponentiated for ease of interpertation.

Model 1, consisting of the control variables, were introduced as predictors. The significant variables in this model were ‘HQ REGION_Europe’ at a 10% significance level, ‘INDUSTRIES_enterprise.software’ and ‘INDUSTRIES_fintech’ at a 5% significance level and ‘Financing for entrepreneurs’ and ‘INDUSTRIES_health’ at a 1% significant level. For every unit increase in ‘Financing for entrepreneurs’, the odds of ‘startup success’ increased by 2.81. Furthermore, ‘HQ REGION_Europe’ was also significant and negatively associated with startup success. Since we dropped ‘HQ REGION_America’ from the data, it became the reference group. The significant negative coefficient in the ‘HQ Region Europe’ means that startups with a headquarter in Europe have approximately 0.50 times lower odds of startup success compared to those with a headquarter in North America. Next, the health industry variable had the highest coefficient and lowest p-value. This indicated that startups who are operating in the health industry have a 6.63 times higher odds of being successful than those operating in non-health industries. The Enterprise Software and Fintech industry were also positively significant. This model had a McFadden R-squared of 0.19, which indicated that the goodness-of-fit is adequate.

In model 2, the founder characteristics were added. We can see that the variables in model 1 were still significant positive predictors. Other positive predictors included the ‘Number of Founders’ and ‘founders_background_unknown’ on a 1% significance level. This indicated that for every unit increase in ‘Number of Founders’, the odds of ‘startup success’ increase by 2.37. Furthermore, the ‘founders_background_unknown’ variable was hard to interpret because it can indicate that the founder’s educational background was not disclosed or that the founder had no higher education. A negative coefficient for this variable indicated that either a lack of higher education or the absence of information of a founder’s background decreased the odds for ‘startup success’. However, due to uncertainty, no conclusive statement can be provided regarding this variable. The McFadden R-squared for this model was 0.32, indicating a much better fit than the first model ($\Delta R^2 = 0.13$ and $p < 0.001$ for the likelihood ratio test). The AIC was also reduced from 256.23 in model 1 to 233.94 in this model. Both models perform better than the null model, which had an AIC of 287.50.

Business model characteristics were added to model 3. The significance of the previous predictors were still present except for 'INDUSTRIES_enterprise.software'. B2B was a significant positive predictor at a 5% significance level. This means that B2B startups are 2.70 times more likely to succeed in comparison to only B2C startups (B2C is the reference group). Lastly, income streams subscription was also positively significant at a 5% significance level. The McFadden R-squared for this model is 0.39 ($\Delta R^2 = 0.06$ and $p < 0.05$ for the likelihood ratio test), indicating a good fit. The AIC was only reduced from 233.94 to 233.39.

Finally, model 4, the model with all variables, was introduced. 'Sustainability' and 'accelerator' were input into the model. 'Accelerator' had a significant negative coefficient, indicating that participating in an accelerator program decreases the odds of 'startup success' by 0.31. Moreover, the 'founders_degree' variable became significant at a 10% significance level. This means that for every unit increase in 'founders_degree', the odds of 'startup success' increased by 1.79. Model 4 had the lowest AIC and highest R-squared, displaying that this model had the best fit out of all the models, although it did not differ by much from the last model ($\Delta R^2 = 0.03$ and $p < 0.05$ for the likelihood ratio test).

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Variables	β	SE	β	SE	β	SE	β	SE
Financing for Entrepreneurs	1.03***	0.31	0.94***	0.34	0.93**	0.36	0.95**	0.38
Internal Market Dynamics	0.28	0.27	0.31	0.32	0.24	0.34	0.17	0.35
HQ Region Europe	-0.69*	0.38	-0.89**	0.45	-1.22**	0.50	-1.39***	0.53
HQ Region Others	0.51	0.60	0.93	0.70	0.85	0.81	0.95	0.85
Enterprise Software Industry	0.87**	0.42	0.91**	0.46	0.25	0.54	0.11	0.54
Fintech Industry	0.95**	0.42	1.17**	0.48	1.17**	0.54	1.28**	0.55
Food Industry	0.23	0.51	0.57	0.66	1.16	0.80	1.16	0.86
Health Industry	1.89***	0.58	2.04***	0.66	2.07***	0.76	2.22***	0.80
Marketing Industry	0.3	0.75	0.17	0.90	-0.13	0.89	-0.35	0.94
Media Industry	-0.98	0.68	-0.90	0.74	-0.13	0.83	-0.22	0.83
Transportation Industry	-0.18	0.68	-0.18	0.79	-0.21	0.90	-0.16	0.98

Number of Founders			0.86***	0.27	0.94***	0.30	0.98***	0.31
Founder’s Gender			-0.29	0.48	-0.29	0.53	-0.27	0.55
Founder’s Founding Experience			0.13	0.42	0.23	0.46	0.17	0.47
Founder’s Technical Background			0.14	0.50	-0.18	0.54	-0.08	0.57
Founder’s Business Background			-0.32	0.43	-0.46	0.50	-0.50	0.53
Founder’s Others Background			0.57	0.44	0.30	0.47	0.35	0.48
Founder’s Background Unknown			-2.46***	0.90	-2.17**	0.93	-2.4**	0.95
Founder’s Degree			0.36	0.29	0.52	0.33	0.58*	0.34
Income Stream Subscription					1.23**	0.56	1.27**	0.58
Income Stream Commission					0.34	0.53	-0.03	0.58
Income Stream Selling own Inventory					-0.83	1.14	-1.02	1.19
Revenue Model SaaS					0.59	0.88	0.59	0.91
Revenue Model Marketplace & ecommerce					0.18	0.84	-0.07	0.88
Revenue Model Manufacturing					0.95	0.97	0.57	1.05
Revenue Model Unknown					-0.49	1.06	-0.93	1.12
B2B					0.99**	0.48	0.89*	0.50
B2B and B2C					1.14	0.76	1.06	0.80
Sustainability							0.71	0.68
Accelerator							-1.17**	0.47
R^2	0.187		0.321		0.385		0.411	
ΔR^2			0.134		0.064		0.026	
AIC	256.23		233.94		233.39		230.10	

*p<0.1, **p<0.05, ***p<0.01. SE = standard error of coefficient

Table 2: Hierarchical Regression

To assess the robustness of the model, a 5-fold cross validation was performed. The coefficients of each fold were investigated, revealing approximately consistent signs across the folds. Furthermore, the average coefficient of the cross-validation was compared with the full logistic regression model. There were very minor differences in the signs of the coefficients and the magnitudes were comparable. This proved that the model is sufficiently robust.

4.2. Additional Analysis

This section further analyzed the differences and similarities between successful and unsuccessful startups.

Firstly, the time to seed funding is an important factor to determine startup success. Startups that can secure funding more quickly have a better chance of positioning themselves for long-term success. Dealroom provides the launch year and seed year which enabled us to count the time to seed funding of startups. Null values in the 'launch year' and 'seed year' columns were dropped, resulting from 206 rows to 167 rows. Next, the time to seed funding values for successful and unsuccessful startups were compared to see if there's a statistical difference between them using the independent sample t-test. Before conducting the t-test, assumptions were tested. When testing for normality using the Shapiro-Wilk test, normality was rejected. This means that the independent sample t-test cannot be conducted, and Mann-Whitney U test - a nonparametric test - was used instead. The null hypothesis states that there's no difference between the two groups. The results of the Mann-Whitney U test was 0.21, indicating that there's not enough statistical significance to reject the null hypothesis.

Next, we were curious to see if the founder backgrounds are significantly different between successful and unsuccessful startups . From the hierarchical regression above, none of the founder's backgrounds (except for founders_background_unknown) were significant. This compelled us to further investigate and assess the individual contributions of these variables. A chi-squared test was employed for each founder's background variable to see whether they have an association with startup success/failure. The chi-squared assumption of independence was fulfilled due to random sampling. Furthermore, the chi-square test 'expected value of cells should be 5 or greater in at least 80% of cells' assumption was tested to see if the chi-square test can be applied. The assumption was met for all variables. Only founders_backgrounds_business was not significant. Table 2-3 shows the contingency table and p-values of the significant variables.

	FOUNDERS BACKGROUND TECHNICAL	
TARGET_VAR	No	Yes
Unsuccessful	36	65
Successful	20	85
Chi-square statistic	6.35	
P-value	0.01	

Table 3: Chi-square Test of Founder’s Technical Background

	FOUNDERS BACKGROUND OTHERS	
TARGET_VAR	No	Yes
Unsuccessful	75	26
Successful	64	41
Chi-square statistic	3.57	
P-value	0.06	

Table 4: Chi-square Test of Founder’s ‘Others’ Background

From observing the contingency table and confirming by conducting a univariate regression, it can be concluded that the founder’s technical and ‘others’ background are more strongly associated with successful startups. It is important to note that when accounting and controlling for other variables in a multivariate regression model, these variables were insignificant, suggesting that they may have less predictive power compared to other variables in the model.

Lastly, we investigated the top industries that successful and unsuccessful startups operate in. Figure 4 and 5 shows the bar graphs of the industries of each startup. Interestingly, fintech and enterprise software were the top 2 industries for both unsuccessful and successful startups. In the hierarchical regression, fintech was significant in the first block, showing that accounting for the external control variables that affect the target outcome, the fintech industry was significantly associated with startup success. On the other hand, health is the third most common industry for successful startups whereas it is the tenth most common for unsuccessful startups.

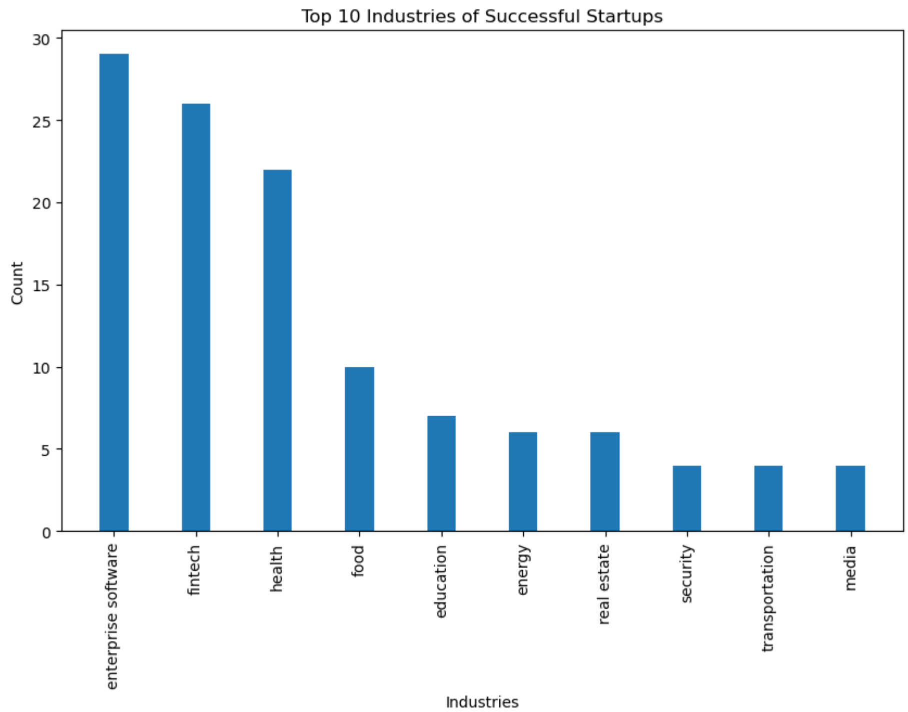

Figure 5: Bar Graph of Top Industries in Successful Startups

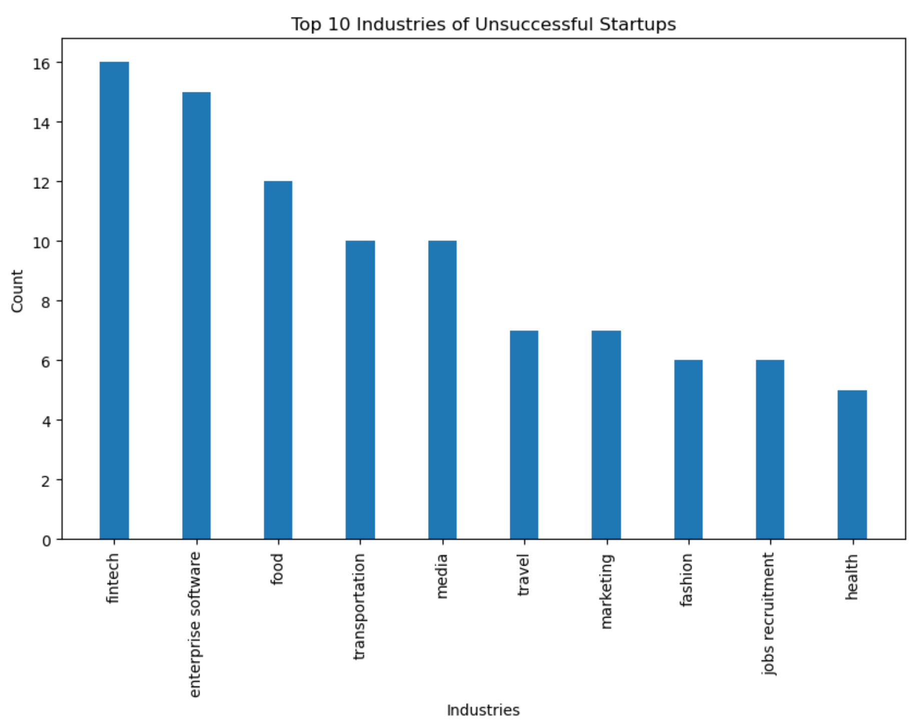

Figure 6: Bar Graph of Top Industries in Unsuccessful Startups

5. DISCUSSION

This study utilized a startup data platform, namely Dealroom, to assess the similarities and factors of unsuccessful and successful startups. We first provided the theoretical background in which the definition of a startup, the startup lifecycle, the measurement of startup success and failure, and startup performance factors according to the founder's and investor's perspective were discussed. We then collected data from Dealroom and performed data preprocessing where variables were transformed, reduced, and added. A hierarchical logistic regression was then utilized to analyze the relationship between the features and outcome of startup success/failure. Afterwards, cross validation was performed to check for robustness. Lastly, additional analysis in the form of univariate and bivariate analysis and visualizations was conducted. Below, a discussion of the results and the connections to previous research were presented.

Since this study did not limit on location, it was highly relevant to control for external factors. The location in which a startup operates can significantly impact its performance. The results showed that availability of funding in a country and having the startup's headquarter in Europe were positive and negative significant variables in the model respectively. Increased availability of funding can provide startups with essential resources to fuel growth. Inadequate access to finance remains a major obstacle for many aspiring entrepreneurs, particularly in developing countries ("5: Improving Access to Finance | UNCTAD," n.d.). Furthermore, this study also showed the disparity between Europe and North America, specifically the US. More than half of start-ups in the US operate in the "Superhubs" which have significantly impacted their success due to the concentration of entrepreneurs, tech talent, and investors ("Entrepreneurship beyond Silicon Valley | McKinsey," n.d.). Superhubs in Europe have not reached the level of the US, and only 30% of European startups have their headquarters in a tech superhub. Furthermore, health and fintech industry are positively related to startup success. Only five out of twenty-two startups in the health industry failed in our data. This can be attributed to the fact that the health industry typically requires more funding due to research and development, regulatory compliance among other things. Lastly, Fintech companies have grown rapidly in recent years and are attracting increased interest from venture capitalists.

For the founder characteristics, we expected that the bigger the founding team and the higher the academic degree of the founders, the higher probability of success. We also expected that founders who possessed a business or technical background and founders who have founded a company previously will also increase the probability of startup success. The founder characteristics that are significant in our model were ‘founder's background unknown’, the ‘number of founders’ and the ‘founder’s degree’. Excluding ‘founder’s background unknown’, these characteristics are supported by previous research. Startups with more founders benefit from variation in experiences, more resources and network (Delmar & Shane, 2006) which impacts performance. For the founder’s academic degree, there have been some discrepancies in results. Some research found that higher education does not improve start-ups’ access to funding (Ratzinger, Amess, Greenman, & Mosey, 2017). However, our results are more similar to Arenius and De Clercq (2005) where the highly educated founders are more likely to receive funding. Moreover, founders with prior firm-founding experience appear to raise more venture capital (Zhang, 2009). However, our results did not yield significance for this in the hierarchical regression and corresponds to Song, Podoynitsyna, Van Der Bij, & Halman (2007) where prior start-up experience has no significant effects on venture performance. Lastly, previous research stated that the founder’s technological background (Paik & Woo, 2017) is important for venture capital funding. Our model produced this finding only at a bivariate level.

In the model, two business model characteristics were found to positively affect startup success — B2B and income stream subscription. Firstly, it’s important to identify the advantages and disadvantages of B2B and B2C. The B2C model exhibits fast scalability due to its expansive market and lower barriers to entry. However, this can also be seen as a disadvantage. B2B, while taking longer to establish a presence in the market, can gain a competitive advantage once it’s in the growth phase. They generate steady revenue, which is appealing for investors. Moreover, B2C companies are getting hard to scale and have a higher risk of failure (Streletzki & Schulte, 2013). Income stream subscription was positively significant in our model. During these recent years, the subscription e-commerce market has grown extensively and is fueled by venture capitalists (“Global Subscription E-Commerce Market Report 2023-2028: Market to Grow at a CAGR of 71% - Growing Use among Millennials, Increasing Internet Penetration, & Low Costs Driving Growth,” 2023). Investors prize subscription models due to it generating more predictable long-term revenue flows than conventional one-off transaction models (Koenigsberg, 2023). From this, it can be concluded that both B2B and subscription models are shown to generate steady and predictable revenue. This appealing characteristic is likely to attract investors.

Next, getting support from accelerators is negatively associated with startup success in our model. Not much research has been done on whether accelerator programs are proved to be effective. However, this was hard to interpret because participating in an accelerator tends to be linked to more success in fundraising, revenue, and employee growth (Miller & Bound, n.d.). The accelerator’s main goal is to prevent the startup from failing and acquire funds. More research needs to be done to achieve a conclusive statement. Lastly, although seed funding time is an important factor on startup performance, no statistically significant difference was found between seed funding time in successful and unsuccessful startups.

6. LIMITATIONS AND FUTURE WORK

One of the biggest limitations in this study was the data itself. Although Dealroom provides extensive data with significant amounts of startups, the bad structure, incompleteness, and inaccuracies bring limitations to this study. Many features were unable to be used, and preprocessing, creating multiple Dealroom accounts, and filling in missing values took a great portion of our time. Furthermore, although this study took an exploratory approach on startup success and failure, many factors were not taken into account, producing omitted variable bias. Dealroom mostly provides general data, and more specific data such as adaptability to market conditions and founder's personality traits would be beneficial to the analysis. Next, this study only analyzed correlation between the dependent and independent variables, not allowing any causal relationship statements to be made. Lastly, due to the availability of the data, we did not limit our scope to a certain location or industry. This increases generalizability. However, this is also a limitation in our study as it becomes challenging to draw definitive conclusions in specific contexts. There may be variations in the contextual factors that could impact the results. Furthermore, it is worth noting that our study is based on a sample size of 206, which, given the broad scope of our analysis, can be considered relatively small.

For future work, including investor characteristics will give more insights and accuracy to the study. Dealroom provides data on Venture Capitals which can be utilized to further analyze startup success and failure. If data allows, exploring the factors in each stage of the startup lifecycle will also be beneficial and insightful. Comparing startups in different stages allows for more complete and accurate analysis. Lastly, when filling in missing values of the founders, it came to our attention that the founders whose unsuccessful startup we analyzed often have already founded another startup. It would be interesting to do a longitudinal study on these startup founders and track the progress and outcomes of these founders across multiple startups over time.

7. CONCLUSION

The goal of this research was to find the factors discerning successful and unsuccessful startups. Defining what constitutes a successful and unsuccessful startup was the first step in our research. We focused on early-stage startups and factors affecting success in the next stage. For this study, we utilized the startup data platform Dealroom to collect data on successful and unsuccessful startups. Afterwards, a hierarchical regression was performed to find factors contributing to startup success and failure. Factors that were found to be associated with the probability of success are highly educated founders, having more than one founder, being in the health and fintech industry, having a B2B and subscription-based business model, and being in a country with significant availability of financial resources. Factors which appear to be to reduce the probability of success are participation in an accelerator program, being headquartered in Europe, and unknown founder background (which can mean a lack of information of founders or no higher education of founders). Furthermore, univariate and bivariate analysis on certain factors were conducted. It was found that founders with technical and non-business and non-technical backgrounds were more associated with startup success, whereas there was no difference between successful and unsuccessful startups for founders with a business background. We also analyzed whether the time to seed funding differs between successful and unsuccessful startups, and results showed that there was no difference. Lastly, it was found that both successful and unsuccessful startups have fintech and enterprise software as their top industries. From this research, we can conclude that external, founder, and business model characteristics all play a role in startup performance. We believe this research will help future research in further exploring and enhancing our understanding of startup success and failure.

References

5: Improving Access to Finance | UNCTAD. (n.d.). Retrieved October 5, 2021, from unctad.org website: https://unctad.org/topic/enterprise-development/entrepreneurship-policy-hub/5-Access-to-Finance

Agrawal, A., Guzman, J. A., & Stern, S. (2019). *Passive Versus Active Growth: Evidence from Founder Choices and Venture Capital Investment*. https://doi.org/10.3386/w26073

Amankwah-Amoah, J. (2016). An integrative process model of organisational failure. *Journal of Business Research*, *69*(9), 3388–3397. https://doi.org/10.1016/j.jbusres.2016.02.005

Azimzadeh, S. M., Pitts, B., Ehsani, M., & Kordnaeij, A. (2013). The Vital Factors for Small and Medium Sized Sport Enterprises Start-ups. *Asian Social Science*, *9*(5). https://www.academia.edu/69080725/The_Vital_Factors_for_Small_and_Medium_Sized_Sport_Enterprises_Start_ups

Bauer, D., Junge, S., & Reif, T. (2023). May the resources be with you: a systematic review and framework of startup funding options. *Management Review Quarterly*. https://doi.org/10.1007/s11301-023-00336-6

Baum, J., & Silverman, B. (2021). *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. https://www.semanticscholar.org/paper/Picking-winners-or-building-them-Alliance%2C-and-as-Baum-Silverman/da39c389570e612cbf9d8473ee5b90590e6678f3

Bernstein, S., Korteweg, A. G., & Laws, K. (2014). Attracting Early Stage Investors: Evidence from a Randomized Field Experiment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2432044

Blank, S. (2010, January 25). *Steve Blank What's A Startup? First Principles.* Steve Blank. https://steveblank.com/2010/01/25/whats-a-startup-first-principles/

Catalini, C., Guzman, J., Stern, S., Thank, W., Amabile, T., Babina, T., Cockburn, I., Greenstein, S., Hopenhayn, H., Kogut, B., Lerner, J., Nanda, R., Phillips, D., Puri, M., & Simcoe, T. (2019). *NBER WORKING PAPER SERIES PASSIVE VERSUS ACTIVE GROWTH: EVIDENCE FROM FOUNDER CHOICES AND VENTURE CAPITAL INVESTMENT*. https://www.nber.org/system/files/working_papers/w26073/w26073.pdf

CB Insights. (2021, August 3). *Why Startups Fail: Top 12 Reasons I CB Insights*. CB Insights Research. https://www.cbinsights.com/research/report/startup-failure-reasons-top/

Croce, A., Guerini, M., & Ughetto, E. (2016). Angel Financing and the Performance of High-Tech Start-Ups. *Journal of Small Business Management*, *56*(2), 208–228. https://doi.org/10.1111/jsbm.12250

Członkowski, R. (2023, April 5). *MVP development - what, why & how*. Www.boldare.com. https://www.boldare.com/blog/mvp-what-why-how/

Delmar, F., & Shane, S. (2006). Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization*, *4*(3), 215–247. https://doi.org/10.1177/1476127006066596

Entrepreneurship beyond Silicon Valley | McKinsey. (n.d.). Retrieved June 18, 2023, from www.mckinsey.com website: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/entrepreneurship-beyond-silicon-valley

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2008). Venture Capitalists' Evaluations of Start-Up Teams: Trade-Offs, Knock-Out Criteria, and the Impact of VC Experience.

Entrepreneurship Theory and Practice, *32*(3), 459–483. https://doi.org/10.1111/j.1540-6520.2008.00236.x

Fried, V. H., & Hisrich, R. D. (1994). Toward a Model of Venture Capital Investment Decision Making. *Financial Management*, *23*(3), 28. https://doi.org/10.2307/3665619

Gladstone, D., & Gladstone, L. (2002). Venture Capital Handbook: An Entrepreneur's Guide to Raising Venture Capital. In *Google Books*. FT Press. https://books.google.nl/books/about/Venture_Capital_Handbook.html?id=i3kGlIVysFwC&redir_esc=y

Graham, P. (2012). *Startup = Growth*. Paulgraham.com. http://paulgraham.com/growth.html

GEM Global Entrepreneurship Monitor. (n.d.). Retrieved from GEM Global Entrepreneurship Monitor website: https://www.gemconsortium.org/wiki/1599

Global Subscription E-commerce Market Report 2023-2028: Market to Grow at a CAGR of 71% - Growing Use Among Millennials, Increasing Internet Penetration, & Low Costs Driving Growth. (2023, January 30). Retrieved June 18, 2023, from Yahoo Finance website: https://finance.yahoo.com/news/global-subscription-e-commerce-market-152000316.html

Granz, C., Henn, M., & Lutz, E. (2019). Research on Venture Capitalists' and Business Angels' Investment Criteria: A Systematic Literature Review. *Contemporary Developments in Entrepreneurial Finance*, 105–136. https://doi.org/10.1007/978-3-030-17612-9_5

Harrison, R. (2017). Backing the horse or the jockey? Due diligence, agency costs, information and the evaluation of risk by business angel investors. *Www.academia.edu*. https://www.academia.edu/88046803/Backing_the_horse_or_the_jockey_Due_diligence_agency_costs_information_and_the_evaluation_of_risk_by_business_angel_investors

Hermann, J. (2022, October 3). *What is the "Startup Valley of Death"?* Www.linkedin.com. https://www.linkedin.com/pulse/what-startup-valley-death-jonas-hermann/?trk=pulse-article

How can start-ups overcome the "Valley of Death"? (2020, July 10). Nexus. https://nexusleeds.co.uk/insights/blog/how-can-start-ups-overcome-the-valley-of-death/

How Dealroom collects data. (n.d.). Knowledge.dealroom.co. Retrieved May 17, 2023, from https://knowledge.dealroom.co/knowledge/how-dealroom-collects-data

How the IPO Process for Startups Works. (n.d.). Www.rocketlawyer.com. Retrieved November 7, 2022, from https://www.rocketlawyer.com/business-and-contracts/business-operations/corporate-finance/legal-guide/how-the-ipo-process-for-startups-works

Kaplan, S. N., Sensoy, B. A., & Strömberg, P. (2007, August 1). *Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=657721

KAPLAN, S. N., SENSOY, B. A., & STRÖMBERG, P. (2009). Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies. *The Journal of Finance*, *64*(1), 75–115. https://doi.org/10.1111/j.1540-6261.2008.01429.x

Kézai, P. K., Fischer, S., & Lados, M. (2020). Smart Economy and Startup Enterprises in the Visegrád Countries—A Comparative Analysis Based on the Crunchbase Database. *Smart Cities*, *3*(4), 1477–1494. https://doi.org/10.3390/smartcities3040070

Koenigsberg, O. (2023, February 9). New Threats to the Subscription Model. Retrieved from MIT Sloan Management Review website: https://sloanreview.mit.edu/article/new-threats-to-the-subscription-model/

Kotashev, K. (2022, January 9). *The Ultimate Startup Failure Rate Report [2019]*. Failory.com. https://www.failory.com/blog/startup-failure-rate

Lebret, H. (2016, May 5). *When was the word "start-up" first used? | Start-Up*. https://www.startup-book.com/2016/05/22/when-was-the-word-start-up-first-used/

Mason, C., Botelho, T., & Zygmunt, J. (2016). Why business angels reject investment opportunities: Is it personal? *International Small Business Journal: Researching Entrepreneurship*, *35*(5), 519–534. https://doi.org/10.1177/0266242616646622

Macmillan, I. C., Siegel, R., & Narasimha, P. N. Subba. (1985). Criteria used by venture capitalists to evaluate new venture proposals. *Journal of Business Venturing*, *1*(1), 119–128. https://doi.org/10.1016/0883-9026(85)90011-4

Marquardt, D. W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, *12*(3), 591. https://doi.org/10.2307/1267205

McGowan, E. (2015, January 15). *The Definition of a Successful Startup Is Surprisingly Complex*. Www.startups.com. https://www.startups.com/library/expert-advice/the-definition-of-a-successful-startup-is-surprisingly-complex

Miller, P., & Bound, K. (n.d.). *The Startup Factories*. Retrieved from http://www.bioin.or.kr/InnoDS/data/upload/policy/1310018323687.PDFMcGowan, E. (2015, January 15). *The Definition of a Successful Startup Is Surprisingly Complex*. Www.startups.com.

Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: a Handbook for visionaries, Game changers, and Challengers*. Wiley.

Paik, Y., & Woo, H. (2017). The Effects of Corporate Venture Capital, Founder Incumbency, and Their Interaction on Entrepreneurial Firms' R&D Investment Strategies. *Organization Science*, *28*(4), 670–689. https://doi.org/10.1287/orsc.2017.1133

Pattanayak, S., & Kakati, M. (2021). An Empirical Study on Key Entrepreneur Motivations and Their Influence on Enterprise Success. *SEDME (Small Enterprises Development, Management & Extension Journal): A Worldwide Window on MSME Studies*, *48*(1), 7–27. https://doi.org/10.1177/09708464211061895

Picken, J. C. (2017). From startup to scalable enterprise: Laying the foundation. *Business Horizons*, *60*(5), 587–595. https://doi.org/10.1016/j.bushor.2017.05.002

Politis, D., & Gabrielsson, J. (2009). Entrepreneurs' attitudes towards failure. *International Journal of Entrepreneurial Behavior & Research*, *15*(4), 364–383. https://doi.org/10.1108/13552550910967921

Prohorovs, A., Bistrova, J., & Ten, D. (2018). Startup Success Factors in the Capital Attraction Stage: Founders' Perspective. *Journal of East-West Business*, *25*(1), 26–51.

Ranjith, V. K. (2016). Business Models and Competitive Advantage. *Procedia Economics and Finance*, *37*, 203–207. https://doi.org/10.1016/s2212-5671(16)30114-9

Ratzinger, D., Amess, K., Greenman, A., & Mosey, S. (2017). The impact of digital start-up founders' higher education on reaching equity investment milestones. *The Journal of Technology Transfer*, *43*(3), 760–778. https://doi.org/10.1007/s10961-017-9627-3

Ries, E. (2011). *The lean startup* (1st ed.). Crown Business.

Ries, E. (2023). *The Lean Startup | Methodology*. Theleanstartup.com. https://theleanstartup.com/principles

Ruby, D. (2022, October 31). *106+ Startup Statistics of 2022 (The Extensive List)*. Demandsage. https://www.demandsage.com/startup-statistics/

Santisteban, J., & Mauricio, D. (2017). Systematic Literature Review of Critical Success Factors of Information Technology Startups. *Academy of Entrepreneurship Journal*. https://www.semanticscholar.org/paper/Systematic-Literature-Review-of-Critical-Success-of-Santisteban-Mauricio/a9d7e3873543e9a862cc588e8ed7bcf9d3a540cb

Santisteban, J., Mauricio, D., & Cachay, O. (2021). Critical success factors for technology-based startups. *International Journal of Entrepreneurship and Small Business*, *42*(4), 397. https://doi.org/10.1504/ijesb.2021.114266

Schumpeter, J. (1934). *The Theory of Economic Development — Joseph A. Schumpeter*. Harvard.edu. https://www.hup.harvard.edu/catalog.php?isbn=9780674879904

Shepherd, D. A. (2003). Learning from Business Failure: Propositions of Grief Recovery for the SelfEmployed. *The Academy of Management Review*, *28*(2), 318–328. JSTOR. https://doi.org/10.2307/30040715

Silva, J. (2004). Venture capitalists' decision-making in small equity markets: a case study using participant observation. *Venture Capital*, *6*(2-3), 125–145. https://doi.org/10.1080/13691060410001675974

Skala, A. (2018). The Startup as a Result of Innovative Entrepreneurship. *Digital Startups in Transition Economies*, 1–40. https://doi.org/10.1007/978-3-030-01500-8_1

Song, M., Podoynitsyna, K., Van Der Bij, H., & Halman, J. I. M. (2007). Success Factors in New Ventures: A Meta-analysis*. *Journal of Product Innovation Management*, *25*(1), 7–27. https://doi.org/10.1111/j.1540-5885.2007.00280.x

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., & Schoder, D. (2015). Business model development, founders' social capital and the success of early stage internet start-ups: a mixed-method study. *Information Systems Journal*, *26*(5), 421–449. https://doi.org/10.1111/isj.12073

Streletzki, J.-G., & Schulte, R. (2013). Which venture capital selection criteria distinguish high-flyer investments? *Venture Capital*, *15*(1), 29–52. https://doi.org/10.1080/13691066.2012.724232

The Levenshtein distance algorithm. (n.d.). Educative: Interactive Courses for Software Developers. https://www.educative.io/answers/the-levenshtein-distance-algorithm

The Venture Capital Funnel. (2018, September 6). CB Insights Research. https://www.cbinsights.com/research/venture-capital-funnel-2/

What Are The Six Stages Of A Startup? | CEMEX Ventures. (2021, February 17). Www.cemex ventures.com. https://www.cemexventures.com/startup-stages-phases/

World Data. (2022). List of the 152 Developing Countries. Retrieved from Worlddata.info website: https://www.worlddata.info/developing-countries.php

Zaheer, H., Breyer, Y., Dumay, J., & Enjeti, M. (2019). Straight from the horse's mouth: Founders' perspectives on achieving "traction" in digital start-ups. *Computers in Human Behavior*, *95*, 262–274. https://doi.org/10.1016/j.chb.2018.03.002

Zhang, J. (2009). The advantage of experienced start-up founders in venture capital acquisition: evidence from serial entrepreneurs. *Small Business Economics*, *36*(2), 187–208. https://doi.org/10.1007/s11187-009-9216-4

APPENDIX

APPENDIX A

TABLE A.1: Summary Statistics of Categorical Model Variables

Variables	Number of startups (N=206)	Percentage
COUNTRY STATUS		
Developed	192	93.2
Developing	14	6.8
FOUNDERS DEGREE		
PhD	35	16.99
Masters	98	47.57
Bachelor	67	32.52
High School Diploma	6	2.91
FOUNDERS GENDER		
Female	48	23.3
Male	148	76.7
FOUNDERS FIRST COMPANY		
Yes	124	60.19
No	82	39.81
Sustainable		
Yes	33	16.02
No	173	83.98
Accelerator		
Yes	71	65.53

No	135	34.47
Income Stream Commission		
Yes	90	43.69
No	116	56.31
Income Stream Selling own Inventory		
Yes	11	94.66
No	195	5.34
Income Stream Subscription		
Yes	72	34.95
No	134	65.05
Income Streams Unknown		
Yes	36	17.48
No	170	82.52
Revenue Model Manufacturing		
Yes	40	19.42
No	166	80.58
Revenue Model Marketplace & Ecommerce		
Yes	51	24.76
No	155	75.24
Revenue Model SaaS		
Yes	109	52.91
No	97	47.09
Revenue Model Unknown		
Yes	22	10.68
No	184	89.32
HQ Region Europe		
Yes	106	51.46
No	101	48.54

HQ Region Others		
Yes	26	12.62
No	180	87.38
HQ Region America		
Yes	79	38.35
No	127	61.65
Founders Background Technical		
Yes	150	72.82
No	56	27.18
Founders Background Business		
Yes	128	62.14
No	78	37.86
Founders Background Others		
Yes	67	32.52
No	139	67.48
Founders Background Unknown		
Yes	17	8.25
No	189	91.75
Business-to-Business		
Yes	108	52.43
No	98	47.57
Business-to-Consumer		
Yes	86	41.75
No	120	58.25
Business-to-Business and Business-to-consumer		
Yes	20	9.71
No	186	90.29
Industries Enterprise Software		

Yes	44	21.36
No	162	78.64
Industries Fintech		
Yes	42	20.39
No	164	79.61
Industries Food		
Yes	22	10.68
No	184	89.32
Industries Health		
Yes	27	13.11
No	129	86.89
Industries Media		
Yes	14	6.80
No	192	93.20
Industries Transportation		
Yes	14	6.80
No	192	93.20
Industries Marketing		
Yes	11	5.34
No	195	94.66

TABLE A.2: Summary Statistics of Numerical Model Variables

Numerical variables	**Mean**	**Min**	**Max**	**Standard Deviation**
Number of Founders	2.09	1	7	1.16
Internal Market Dynamics	5.1	3.36	7.25	0.64
Financing for entrepreneurs	4.55	2.1	6.2	0.69

APPENDIX B

FIGURE B.1: Correlation Matrix of Independent Variables

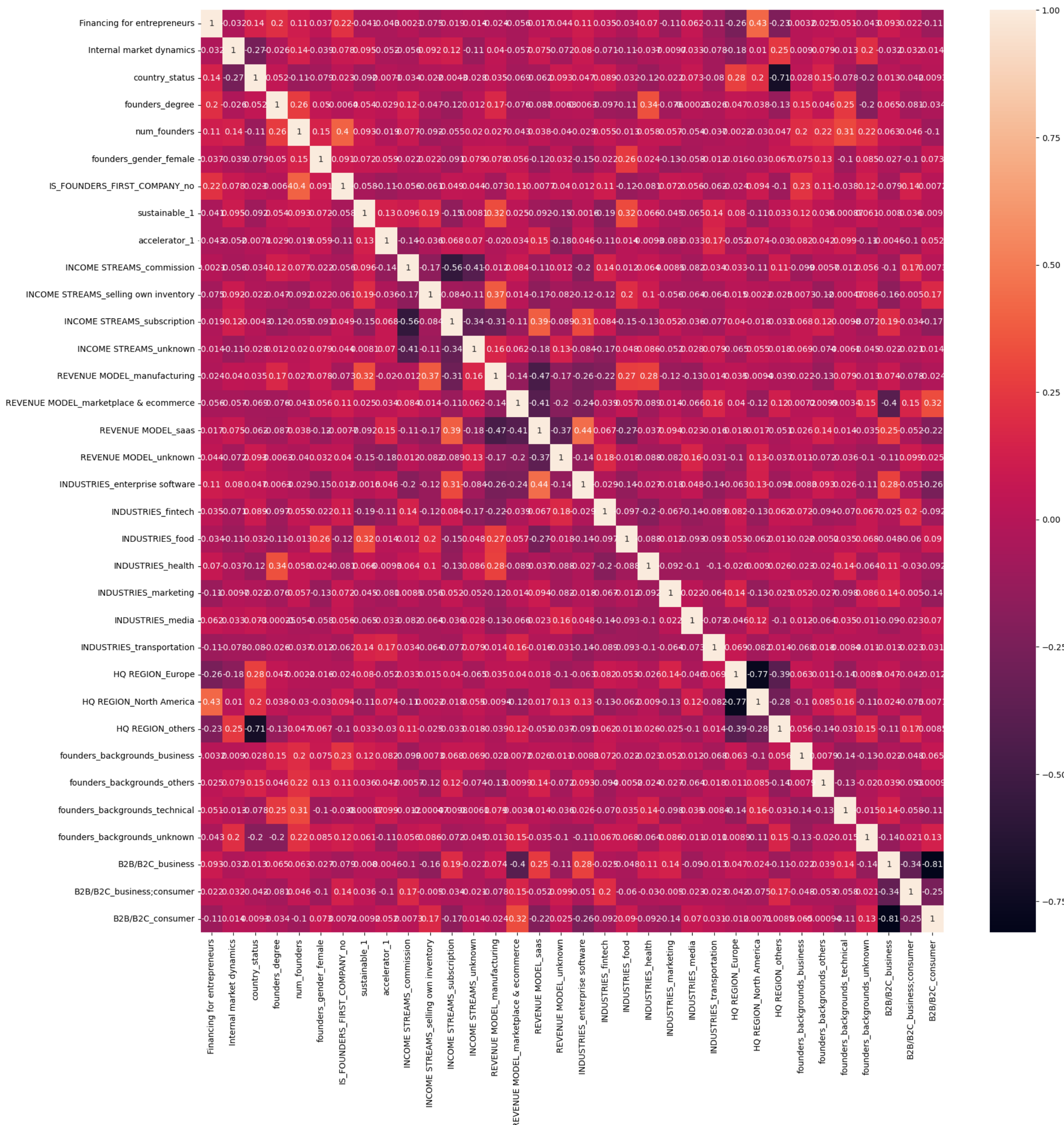

TABLE B.2: VIF of Final Model Variables

Financing.for.entrepreneurs 1.645486	Internal.market.dynamics 1.356357
founders_degree 1.615144	num_founders 2.371998
founders_gender_female 1.495354	IS_FOUNDERS_FIRST_COMPANY_no 1.413112
sustainable_1 1.598556	accelerator_1 1.387721
INCOME.STREAMS_commission 2.121309	INCOME.STREAMS_selling.own.inventory 1.347613
INCOME.STREAMS_subscription 2.160008	REVENUE.MODEL_manufacturing 4.031070
REVENUE.MODEL_marketplace...ecommerce 3.841632	REVENUE.MODEL_saas 5.458526
REVENUE.MODEL_unknown 3.005779	INDUSTRIES_enterprise.software 1.495010
INDUSTRIES_fintech 1.401931	INDUSTRIES_food 1.761171
INDUSTRIES_health 1.571474	INDUSTRIES_marketing 1.346419
INDUSTRIES_media 1.254890	INDUSTRIES_transportation 1.510079
HQ.REGION_Europe 1.880992	HQ.REGION_others 2.277819
founders_backgrounds_business 1.708028	founders_backgrounds_others 1.332277

founders_backgrounds_technical 1.731193	founders_backgrounds_unknown 1.552471
B2B.B2C_business 1.677497	B2B.B2C_business.consumer 1.314402

TABLE B.3: Box Plot of Numerical Variables