MSc Business Analytics
Imperial College London
August 2020

Talbot Adrien – 01817798

A data-driven analysis of start-ups' success and survivability

Word count: 5,189

Abstract

Background: Quantitative methods can help investors such as venture capital funds in discovering new investment opportunities and help start-ups to secure funding depending on the market and country they are operating in. Due to the extensive domain knowledge requirement and the lack of access to exhaustive information, a combination of business acumen and quantitative methods would allow to collect, process and analyze data with the aim to answer the following question: what factors allow companies to become successful? Or given the available data prior to 2014, which specific funding sources allowed companies to survive? **Objective:** This report did not provide an established data-approach to spot the next Facebook, but rather looked at overall trends that companies which succeeded or survived share and attempted to discover insights on the success and survivability of companies given funding data in various markets and countries by using a set of data-related and statistical tools. We also tried to predict whether a company will be acquired or not by using supervised machine learning models. **Results**: Access to funding, and especially venture funding, was crucial for companies to survive. However, the success rate of companies was less dependent on funding raised than survival rate, hence the importance of strategic planning, the market as well as the country the start-up is operating in to determine success. In the past six years, companies in the biotechnology sector were particularly attractive in the eyes of investors. This strong interest from investors could be illustrated with the 600% increase in funding for companies in this sector from 2014 to 2020.

Table of Content

1. Introduction

1.1 Context and objectives of the report

This research report looked at data-driven methods to assess start-ups' survivability and success. Who can benefit from reading this report? First, companies looking to secure funding. Indeed, the most important variables we looked at throughout this analysis were the various funding sources companies can have access to. This enabled us to deduct if certain funding sources were better for companies to succeed or survive than others, while looking at specific markets and countries as well. Second, investors might also benefit from reading this research paper. For instance, investments funds such as venture capital funds need to develop reliable and quantitative methods to determine which company will generate excess returns since as much as half of the companies VC funds (VC) invest in fail on average (Kupor, 2019 [1]). Last, the economy as a whole would benefit from a quantitative approach to evaluate start-ups' potential success as, according to Staniewski, emerging start-ups play an important role in driving growth worldwide (Staniewski, 2012 [2]) .

To conduct an in-depth analysis on start-ups' survivability and success, this report mainly used data from more than 50,000 companies extracted from Crunchbase, a platform for finding business information about private and public companies, on December 2, 2014 such as total funding amount raised with a detailed breakdown of the type of funding, company status, industry, country, date founded and many other information. We looked at the relationships between numerical and categorical variables mentioned above with the aim to demonstrate which type of funding was determinant for companies to succeed and how it varied across markets and countries. In addition to that, we also leveraged from up-to-date data of around 1,100 companies retrieved for the purpose of this research using web scrapers in order to provide further insights on which companies have failed or increased their funding amount during the past six-year period (2014-2020). We compared our findings on the sample data with the results we obtained from the original data set.

As the research question that we addressed was: " What are the factors which affect survivability and success of start-ups", we first needed to define success, failure and survivability. The status of a company, which could either be "operating", "acquired" or "closed" gave a good indication of success or failure. Usually, investors which have equity stakes in a specific company generate a return when that company is acquired. Thus, we used the acquisition status as our main success metric in the analysis. On the other hand, a company that has closed suggested that the company failed to meet its goat. Hence, the closed status was a good proxy for failure. Last, being acquired or operating defined another key metric in our analysis: survivability.

1.2 Methodology

The report was structured into three main sections. In the first part we performed an exploratory data analysis on the data set using both descriptive and inferential statistics as well as conducted a binary logistics regression on the success binary variable. The logistic regression model was able to show which variables had an association with the target variable success given a 99% confidence level. In the second section, we looked at our sample data on 1,100 companies we obtained by scraping Crunchbase which allowed us to retrieve updated data on company status and funding amount raised. The objective was to obtain insights on funding change in the past six years at a country and market level, observe if market traction evolved and compare survivability with the period prior to 2014. Last, we developed machine learning models to predict if a company will be acquired using the original data set of 50,000 companies and compared performances to choose the most accurate predictive model.

2. Exploratory data analysis

2.1 Data-set description

Before diving into the funding analysis, it was important to provide general information of the data we manipulated. As mentioned in the introduction of the this report, the data set contained information on approximately 50,000 companies obtained from Crunchbase in December 2014 with details such as market, country, total funding by the end of 2014, number of funding rounds, details about the type of funding raised (equity, debt or VC round) and status. We also decided to remove 20,000 observations with key missing values in order to obtain more interpretable results. Additionally, some variables were created such as "company life duration", which is the time difference between the moment the data was extracted and the time a specific company was founded, and dummies for the country and market variables for the purpose of the logistic regression. Let's now explore the data set in more-depth. There was a strong presence of American companies as 69% of companies in the data set were based in the USA, Great Britain came second with only 5.8%, Canada had 3% and 2% of companies were based in France (see Appendix 2.1).

Furthermore, 11.2% of companies were operating in the software market, which was the most popular one, biotechnology was second with 8.69% and 5% of companies were operating in the Mobile sector (see Appendix 2.2).

There were two important pieces of information we mostly focused on during the analysis: company's status, which can either be "operating", "acquired" or "closed", and total funding amount. Note that as of December 2014, approximately 86% of the companies in the data set were operating, 10% were acquired and around 5% were closed (see Appendix 2.3). As for the market distribution for each status, we observed that in acquired companies, there was a greater proportion of companies operating in the software market (15%) than with closed companies (12%) and companies with "operating" status (see Appendix 2.4). By looking at the acquisition of the top 15 most frequent markets in the data set, we observed that software and enterprise software companies had high acquisition rates of 12% and 18% respectively. Additionally, the proportion of US-based and Canada-based companies in the "acquired" sample was greater than in the "operating" and "closed" samples (see Appendix 2.5) and we did not see a similar trends with other countries. The high acquisition rate of US and Canadian companies compared to other countries (see Appendix 2.6) also suggested that North American companies were more often successful than companies based in other countries. In the next section, we discussed the importance of access to funding in determining the success and survivability of companies across different countries and markets.

2.2 Funding analysis

As many start-ups grow in an environment where access to funding and especially venture funding is highly competitive (Kupor, 2019 [3]), it is crucial for companies to understand which funding sources are the best suited for their needs in order to survive and scale. The analysis offered insights on the relationship between the various funding sources and success or survivability by considering the different markets and countries as well. However, it was limited due to the lack of more granular data such as founders' motives and capabilities, which impacts the choice of source of the initial capital (Anirudh Gard, 2017 [4]).

First, let's define which funding sources were available to companies. Crunchbase separated seed investment and angel funding which were both essentially funding at an early stage from private investors who want to invest their capital into potentially successful businesses. An angel round is usually used to get a company off the ground by looking for funding from family and friends as well (Ramadani, Veland, 2012). Crunchbase defined seed round as being more consequential and ranged from USD 10,000 to USD 2 million and happened before Series A round, which we will explain later (Crunchbase, Glossary of Funding Types, 2020 [6]). Crowdfunding consists of raising money in exchange for equity or use of product from a large number of people notably via Internet. Convertible notes are short-term debt which converts to equity. With debt financing, an investor lends money to a company, and the company promises to repay the debt with added interests. Private equity funding is led by a private equity fund and it is a late stage round and the rounds are typically upwards of USD 50 million. A grant is when a company or institution provides funding without taking an equity stake in the company. Secondary market is a fundraising event in which one investor purchases stock in a company from an existing shareholder rather than from the company directly. Last, the data set also separates the different venture funding rounds (series A to H) which is essentially capital invested by VC funds into start-ups in exchange for equity in the business (Crunchbase, Glossary of Funding Types, 2020 [6]). Series A and Series B rounds range on average between USD 1 million and USD 30 million, while following Series rounds are for later stage and more established companies and are often much larger (Crunchbase, Glossary of Funding Types, 2020 [6]).

Overall, using the data from Crunchbase, we saw that acquired companies raised USD 24 million on average while operating companies raised USD 18 million and closed companies USD 10 million (see Appendix 2.7). Acquired companies also went through the most funding rounds (2.2 on average) whereas operating companies raised funding through 1.9 rounds and closed companies had 1.6 on average. Funding, which is necessary to grow, is good proxy for survivability (Anirudh Gard, 2017 [4]) and thus it was not surprising to see that companies which failed did not secure as much funding as successful companies which were acquired or still in operating. However, we investigated by looking at the breakdown of funding type across different status, which is what we will discuss next.

The funding distribution table summarized the proportions and average funding raised per funding source for each status (see Appendix 2.8). A couple of observations were worth further investigating here. The first one was that the proportion of companies raising seed was greater for closed companies (28%) than acquired ones (17%), although the average amount raised was greater for acquired companies, and we saw a similar trend with funding coming from angel investors (see Appendix 2.9). It could be partly explained by the fact that less successful companies needed to rely more on funding from non-institutional source such as friends, family or private investors as they did not manage to raise from VC funds hence the higher proportion. Second, we saw that companies which will close in the future proportionally secured less funding in series rounds, but those companies raised almost as much as successful companies in series A to D and even more in series E, F and G on average (closed companies raised USD 37 million on average through series E compared to USD 19 million for acquired companies). We decided to disregard series H as the sample data was too small (the data set only had four observations in total). Although we could clearly observe that the survival rate increased until series round D and then remained stable, the change in success rate followed more an inverse U-shape curve where the turning point appeared to be series round E (see Appendix 2.10). This observation and seeing also a higher average amount raised by closed companies in series E onwards implied that the relationship between venture funding and success was more contrasted. In his research paper, Anirudh Gard emphasizes on the idea that the reason the majority of companies which have raised initial funding fail is due to the lack of planning and proper vision from the founders (Anirudh Gard, 2017 [4]). Thus, we could argue that it was the combination of both proper execution and securing venture funding that determined which companies succeed or fail. Securing large amount of funding itself cannot necessarily translate into success and it might even lead to issues if the company lacks proper strategic planning and does not generate profits, which we will be discussing in more detail in the next section. Last, successful companies had a better diversification of their funding sources than companies with a "closed" status, which was seen when looking at private equity and debt financing. Indeed, 12% of acquired companies used debt to finance themselves compared to 9% for closed companies, and 3% of acquired firms made use of private equity funding while it is only 1% of closed companies (see Appendix 2.8). These results suggested that diversification enabled companies to secure funding from different sources and thus lowered the risk of failure.

The Crunchbase data contained 721 unique markets and we observed that the one which raised the most funding on average was the clean technology sector (USD 46 million) followed by health care and biotechnology, and it also secured the most venture funding. However, the clean technology only had an acquisition rate of 6% in the same time period, which was much lower than software market (12%). On the other hand, the survival rates of the most funded sectors were high and ranging between 93% and 97% (see Appendix 2.11). These results reinforced the idea that funding alone did not determine success while it was the case for survivability. To investigate on the specific funding needs across different markets, we looked at the funding disparities between two sectors: healthcare and mobile. We decided to choose these sectors as they differentiate from one another and thus, it would be clearer to see any patterns. Quite surprisingly, companies in the healthcare sector raised little funding from seed and angels investors compared to mobile businesses (see Appendix 2.12). Debt financing was one of the main funding sources for healthcare companies (23% of companies in the healthcare market raised debt funding compared to 9% for mobile firms) and healthcare companies secured more funding from series B round than from series A. This finding could either mean that healthcare companies struggled to raise initial funding from VC funds and thus looked for other ways to raise money, or start-ups in the healthcare sector favored other financing sources such as debt financing due to the absence of willingness to give away equity stakes in return of the investment. Mobile start-ups on the other hand more often raised funding from series A rounds (28%). However, healthcare companies raised more in each series round than health care companies on average and more often secured series C funding onwards regardless of the status. Therefore, despite the differences in funding needs and requirements across different markets, we also noticed that acquired and operating companies in both industries raised more venture funding than closed companies, which showed that it is crucial for both industries to access venture funding to grow and survive.

Finally, as discussed in section 2.1, the proportion of US-based companies in the acquired sample was greater than in the closed sample, and it was a trend we only observed with two countries: the US and Canada. Additionally, the average acquisition rate of US companies was 11%, while it was only 6% for non-US based companies. The closing rate was approximately the same for both US and non-US companies (c. 5%). So why were American companies more successful? We observed that companies founded in the US raised funding from VC funds more frequently and for larger amounts in series round A and B (see Appendix 2.13). The results from the logistic regression showed that series A and B rounds are statistically significant variables with a 99.9% confidence interval when trying to model the probability of a company to be acquired or not (see Appendix 2.14) while it was not the case for subsequent rounds. Thus, securing funding specifically from series A and B rounds was a major determinant for success in the US, which was not surprising given the importance of venture capital for the growth of the US economy (Strebulanev and Gornall, 2015 [7]). However, this apparent trend contradicted our previous findings as regard to the relationship between success and venture funding. However, the correlation between success and venture funding was not common to all countries. Indeed, China raised the most funding on average (USD 46 million) as well as more venture funding than companies in the US (USD 27 million vs USD 13 million) but it had a low acquisition rate of 2.4% compared to the US (12%). This was a trend we observed for countries outside of the United States and therefore, there was not an apparent relationship between success and funding in these countries.

3. Funding change analysis from 2014 to 2020

In this section, we shifted our interest to funding changes from the past six years (2014-2020) in order to study current trends. To do so, we scraped Crunchbase using the BeautifulSoup python package to obtain the total funding amount raised in August 2020, without the detailed breakdown per funding source, as well as the updated status of 1,100 randomly selected operating companies present in the original data set. Scraping Crunchbase was restricted and thus, we could not retrieve the up-to-date information of all the companies in the original data set. It was also important to mention that, contrary to the data set containing 30,000 companies, the sample set did not contain any companies with the "acquired" status. Therefore, the survivability metric we used in the previous section was modified as it only included operating companies in this section. We then used the funding change from 2014 to 2020 metric in the upcoming analysis to assess survivability and traction of specific markets.

From 2014 to 2020, approximately 9% of the companies which were operating in 2014 closed during the time period and 54% either closed or did not raise additional funding. We noted that the closing rate was higher in our sample data than in the original data set where the closing rate was 5%. Let's now look at funding change between closed and operating companies: in the past six years, 69% of the companies which closed did not secure additional funding whereas that proportion was only 49% for companies still operating (see Appendix 3.1). Although we cannot observe the changes in the various funding sources, these findings suggest that there was a clear positive relationship between funding increase and survivability in the past six years. It was also interesting to see the evolution of the closing rate as a function of funding change (see Appendix 3.2). We observed that the closing rate dropped when the funding change went from one (no funding change between 2014 and 2020) to three (funding raised tripled during the time period). However, we observed that the closing rate then rises and reaches a maximum value of 8.9% which showed that companies which saw their funding amount multiplied by 16 or more in the past six years (there were 52 companies in total) had a higher closing rate than companies which did not secure any funding. After that, the closing rate went back to a lower level and reached the lowest point when the funding change was equal to or greater than 26 (3.1%). These results implied that, as previously stated, there was a positive relationship between funding change and survivability. However, the recurrent spikes in closing rates observed on the graph as funding change increases indicated that some start-ups should not necessarily raise additional funding even if they can. Indeed, companies which wanted to scale and hence kept on raising funding might have found themselves in a situation where they used up all of their cash reserves and due to the lack of vision, a strong business model, profits and because of the failure to deliver based on investors' expectations, these companies failed despite large amounts of funding secured initially. What happened to WeWork, an American commercial real estate company that provides shared workspaces, in 2019 was a clear example of mismanagement and overfunding issue since the company secured large funding amounts from venture capital firms and notably from the Vision Fund, the biggest VC fund in the world, while incurred consequential losses due to poor strategic planning, which bankrupted the company (Platt and Edgecliffe-Johnson, 2020 [8]).

At a country-level, US-based companies saw their total funding amount multiply by 11 on average between 2014 and 2020 while Canadian and UK-based businesses had a lower funding amount change (see Appendix 3.3). As a reminder, we found that prior to 2014, US companies had the highest acquisitions rates, but we observed that average survival rate of US-based companies decreased between the period before 2014 (95%) and the period from 2014 to 2020 (90%). These results suggested that the funding change of 11 times should not be considered as high and since the survival rate decreased, we believe that the funding change could be lower than usual in the previous periods.

As for markets, the biotechnology, curated web and e-commerce sectors saw their funding amount multiply by six from 2014 to 2020 (see Appendix 3.4). It showed that general interest shifted compared to the period prior to 2014 where biotechnology companies only represented 5% of acquired companies and we saw that software companies, which had the most traction prior to 2014, only saw their funding amount grow by 300% on average during the same time period. This result confirmed the importance of biotechnology firms in the past decade in driving innovation notably with the growing interest of the genetic engineering field which has for example been used to develop new medicine. This trend is likely to continue in the future given the recent numerous initial public offerings of companies in the biotech sectors compared to the past decade (Blankenship, 2020 [9]).

4. Predicting acquired companies

In this section, we used classification models to predict if a company will be acquired or not given the 30,000 observations available in the Crunchbase data set prior to 2014. We decided to focus on predicting acquired companies rather than closing ones since we believed that the dependency on funding amount is greater when trying to predict closing companies than when predicting acquired ones. As we have seen in the report, there was a clear positive relationship between total funding raised and survivability. On the other side, we concluded that the relationship between success and funding raised is less apparent. An accurate prediction model would benefit investors by allowing them to make better investment decisions and notably for VC funds which generate a return when a company in their portfolio is acquired.

4.1 Feature engineering and pre-processing

Prior to implementing the models, pre-processing the data and choosing the right features were crucial in order to obtain the best-performing classifier. We created dummy variables for country and market in order to transform these categorical variables into numerical ones and use them for the purpose of predicting successful companies. There was also a clear imbalance in the target variable as only 10% of the companies in the data set were acquired. It was an issue as the predictions on the minority class "acquired" are the most important. To solve this problem, we decided to under-sample the data set, which consisted of randomly deleting observations in the majority class so that there was an equal distribution of acquired and non-acquired companies in the data set. Another important step in the pre-processing was to remove highly correlated variables by looking at the correlation matrix. We removed the "venture" variable as it was highly correlated with series rounds A to H. After transforming the data set, we were left with around 4,900 observations and 176 features which mostly contained all the different funding sources as well as binary features for the 100 most frequent markets and the 50 most frequent countries. Lastly, we separated the data set into a training, validation and test set. The training set, which was used to train the model, contained 3,000 observations. The models were then tested on the validation set consisting of 1,000 observations in order to provide an unbiased evaluation of the model fit on the training data. We decided which model to choose based on the performance on the validation set, and we re-trained using both the training and validation set to obtain the final model to use on the separated test set. This approach of isolating the train, validation and test sets allowed to provide an unbiased sense of model effectiveness (Kuhn and Johnson, 2013 [10]).

4.2 Classification model selection

Four supervised classification models were implemented: a nearest-neighbors classifier, a support vector machines classifier, decision tree and a random forest classifier, as well as a logistic regression. For each one of the models, we tested its performance on the validation set after tuning for the optimal parameters and normalizing the data if needed. The nearest-neighbors classifier measures the distances between pairs of samples to classify each observation into a class and thus, it is necessary to normalize the data because the distance calculation use features values. We implemented feature scaling using the Z-score normalization technique which is used to assign uniform weights to the features. According to Dr. Sebastian Raschka, refraining from normalizing the data would give some features more importance than others (Raschka, 2014 [11]). Then we had to find the optimal count value of the nearest neighbors. To do so, we assessed the performance of the model with k values ranging from 1 to 100. With this iterative process, the optimal k value we found was 72 with an accuracy score of 67% on the validation set. Similarly, the support vector machine classifier also deals with distances when trying to maximize the distance margin of the hyperplane. Therefore we normalized the data using the same process we performed on the nearest neighbors model. The results we obtained on the validation set were slightly less accurate than nearest-neighbors with a score of 64.8%. Next, we implemented a decision tree classifier. Contrary to previous models, the decision tree classifier does not require data normalization since it is partitioning the data set into subsets which does not need any distance calculations. We found an optimal tree depth of 42 by computing the accuracy score for each tree depth parameter ranging from 1 to 50 (see Appendix 4.1). The accuracy score on the validation set was 65.4% which was slightly better than the support vector machine model but lower than the performance of the nearest neighbors classifier. Not surprisingly, the decision tree accurately predicted 100% of observations on training set which shows a common practical issue of the decision tree: it is prone to overfitting (Kerdprasop, 2011 [12]). The random forest classifier is a collection of decision trees and it does not require normalization of the data for the same reason as the decision tree classifier. The hyper-tuning process found an optimal tree depth of 20 with 200 estimators with the following accuracy performance: 84% on the training data and 70.4% on the validation set. We saw that there was less overfitting than with the decision tree classifier thanks to the parameter-tuning process and the random forest classifier achieved so far the best prediction performance. Additionally, this model was the best since it overfitted less than the random forest classifier without the optimal parameters found with the tuning process (which predicted 100% on the training data), while performing better on the validation set. Last, we tried to run a logistic regression on the data, which models the probability of a company to be acquired or not by using a maximum likelihood estimation to estimate the optimal beta coefficients. In that case, it was necessary to normalize the data when implementing a logistic regression and we obtained an accuracy score of 66.1% on the validation set.

4.3 Performance report on test-set

By looking at the performance summary table (see Appendix 4.2), we see that the random forest classifier is the most accurate model on the validation data. Thus, we re-trained the model with both the training and validation sets using the same parameters we obtained from the tuning process. The accuracy of the random classifier was 83.3% on the combined training and validation sets and the prediction score on the test set was 71.9% with a generalization error of 11.4%. The classification report shows that the model does not show strong divergence across the difference performance metrics (see Appendix 4.3).

5. Limitations and recommendations

Our analysis enabled to analyze overall trends and patterns between funding sources and company's success or survivability, and prepared the ground for a more tailored analysis by country or sectors. Indeed, narrowing the analysis to specific countries aside from the US and particular sectors would have enabled investors to spot local disruptive start-ups and allow companies to discover the best funding sources available in the country they are based in. This focused approach would have provided additional value to start-ups and investors.

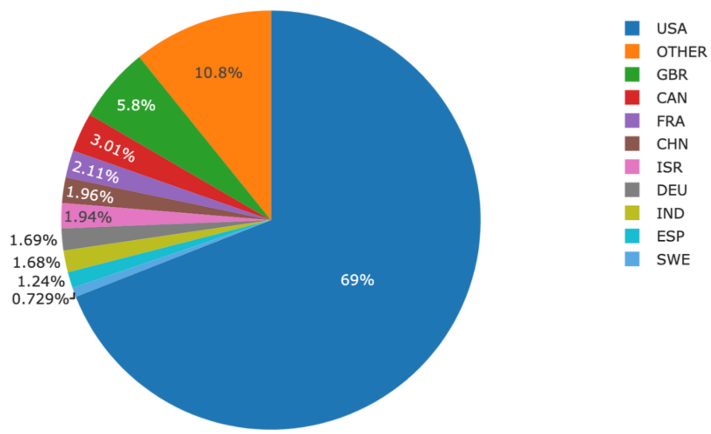
Regarding the prediction model, the accuracy score of 72% obtained using the random forest classifier on the test data is partially convincing due to the complexity of predicting if a company will be acquired given the numerous existing external factors such as the domain knowledge requirement or background of the team, which were information we did not have in the Crunchbase data set. Hence, the lack of data in order to predict success makes this model is hardly applicable in the real world. Additionally, the small size of the sample data we obtained using web scrapers limited the prediction analysis. Obtaining additional updated information would have allowed us to implement other predictive models with the aim, for instance, to determine the status or funding change six years from now. Unfortunately, we could not retrieve more data due to the restrictions placed on scraping Crunchbase. Also, the sample data had too few acquired companies, which limited our ability to link the results we found by analyzing the sample data with the insights obtained by conducting the exploratory data analysis on the original data set. Last, the proportion of closed companies in both the original and the sample data sets seem to be low (5% and 10% respectively). Companies which closed could have been removed from Crunchbase and thus, our data would not represent the population statistics. Similarly, most of the companies are based in the United States and the data set lacked companies from other countries. For instance, providing additional observations from companies based in Asia would have provided a better representation of the population data.

6. Conclusion

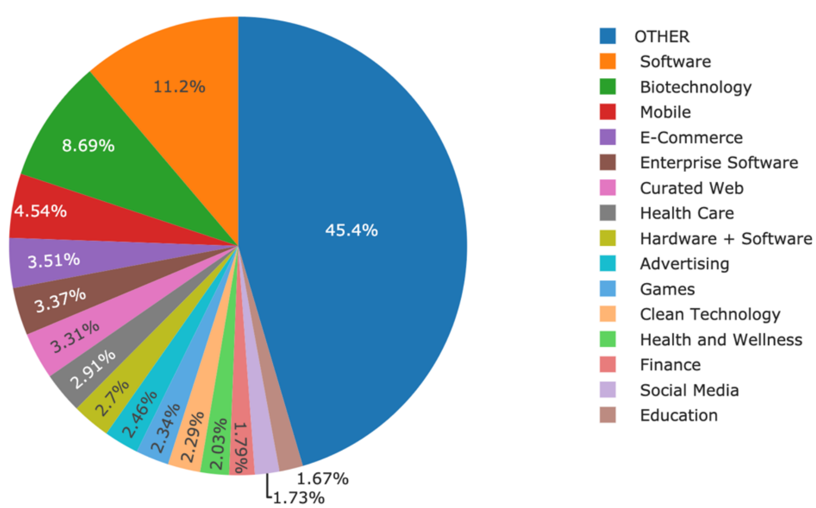
Our study aimed to assess the survivability and success of start-ups across various markets and countries prior to 2014 and between 2014 and 2020 by using the original Crunchbase data set and a sample data of randomly selected companies. We came to the conclusion throughout the report that funding and particularly financing secured from venture capital funds is crucial for companies to survive, which means being acquired or still operating. However, the success rate of companies was less dependent on funding raised than survivability, and thus, a multitude of other external factors needed to be taken into account to assess the potential of a start-up to get acquired. Additionally, we discovered that accessing venture funding was determinant for companies to succeed specifically in the US but other countries such as China did not show a similar trend, although Chinese companies raised the most venture funding on average. Biotechnology companies were particularly attractive in the past six years, which is reflected by the 600% increase in funding on average in this sector. Finally, we developed a classification model which predict which company will be acquired with a 72% accuracy. Although we lacked the data to build a more realistic model, it could still be used by investors to spot new investment opportunities.

7. Appendix

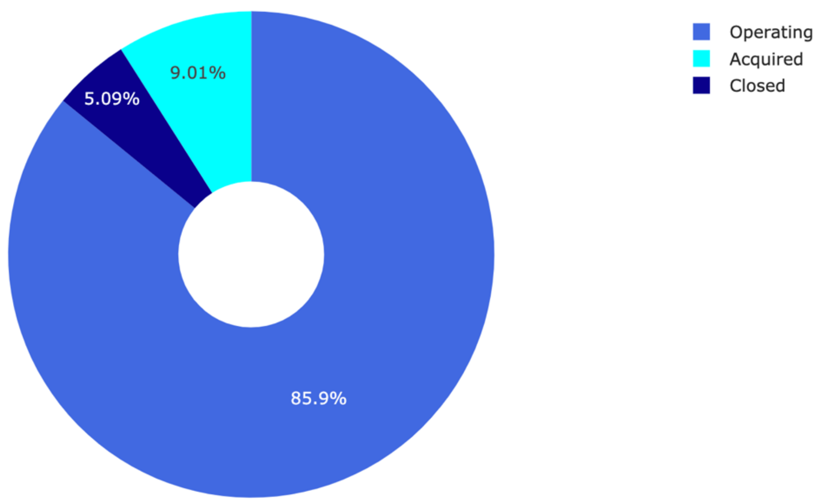
Appendix 2.1: Countries distribution prior to 2014



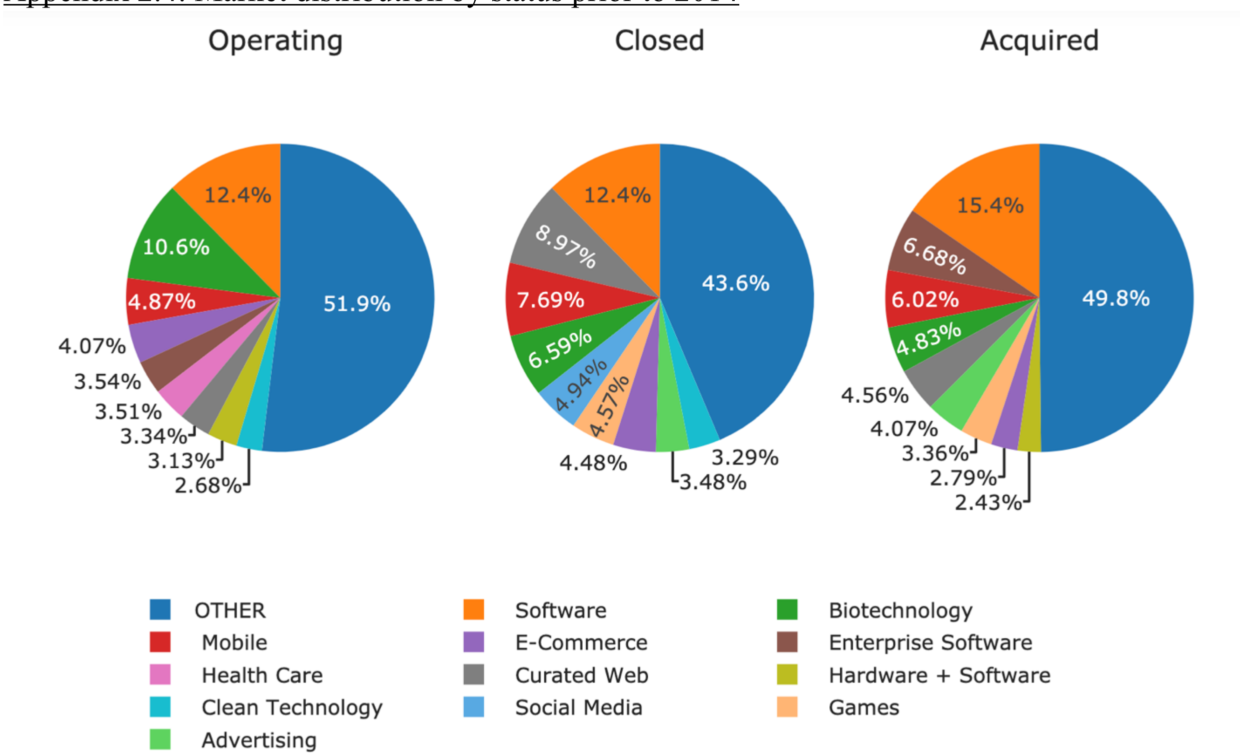
Appendix 2.2: Markets distribution prior to 2014



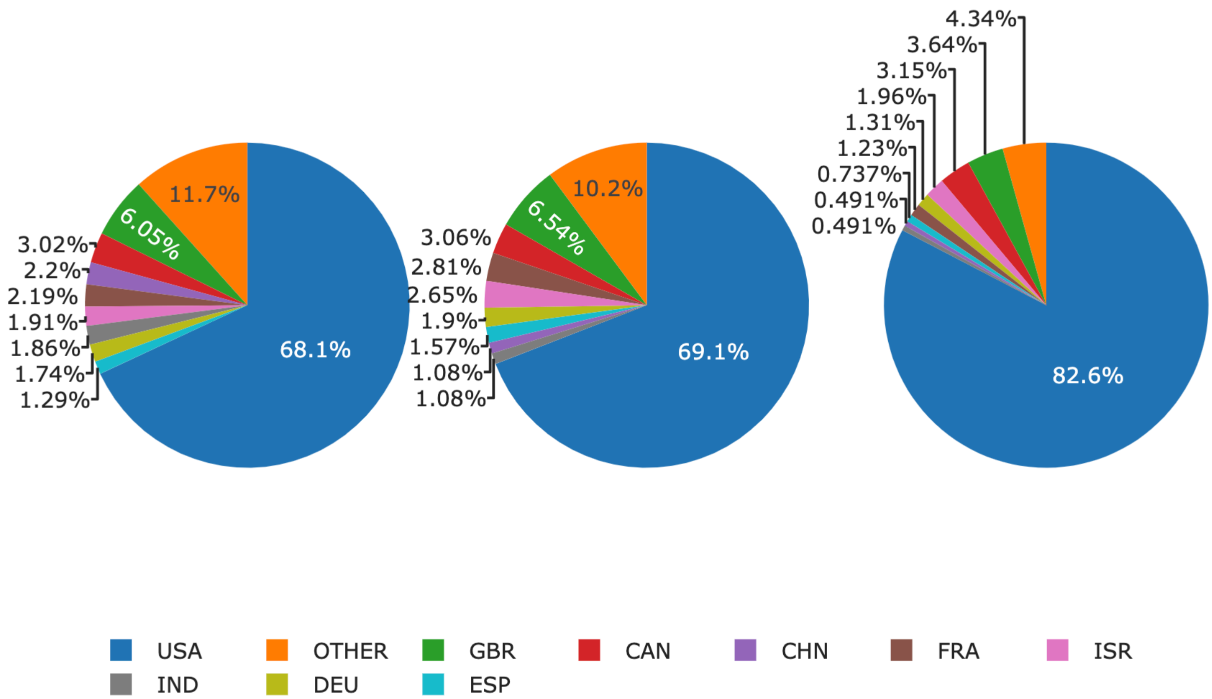
Appendix 2.3: Status distribution prior to 2014



Appendix 2.4: Market distribution by status prior to 2014



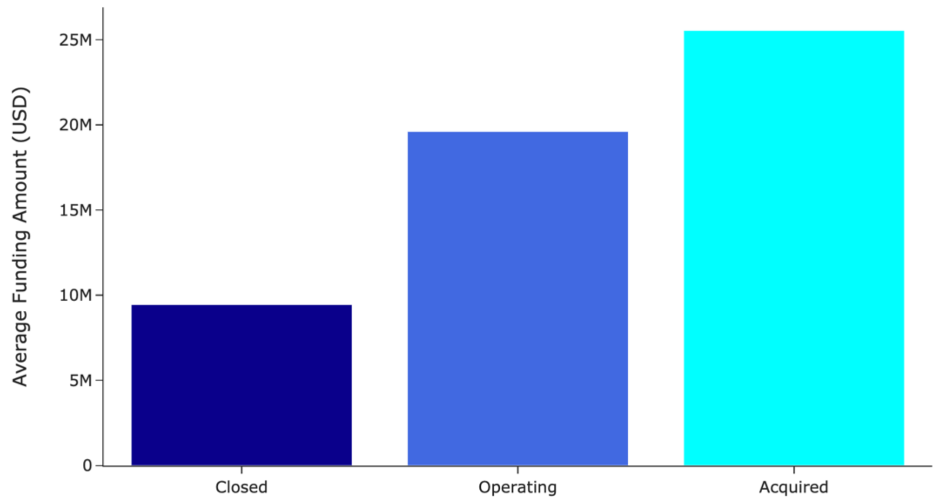
Appendix 2.5: Country distribution by status prior to 2014



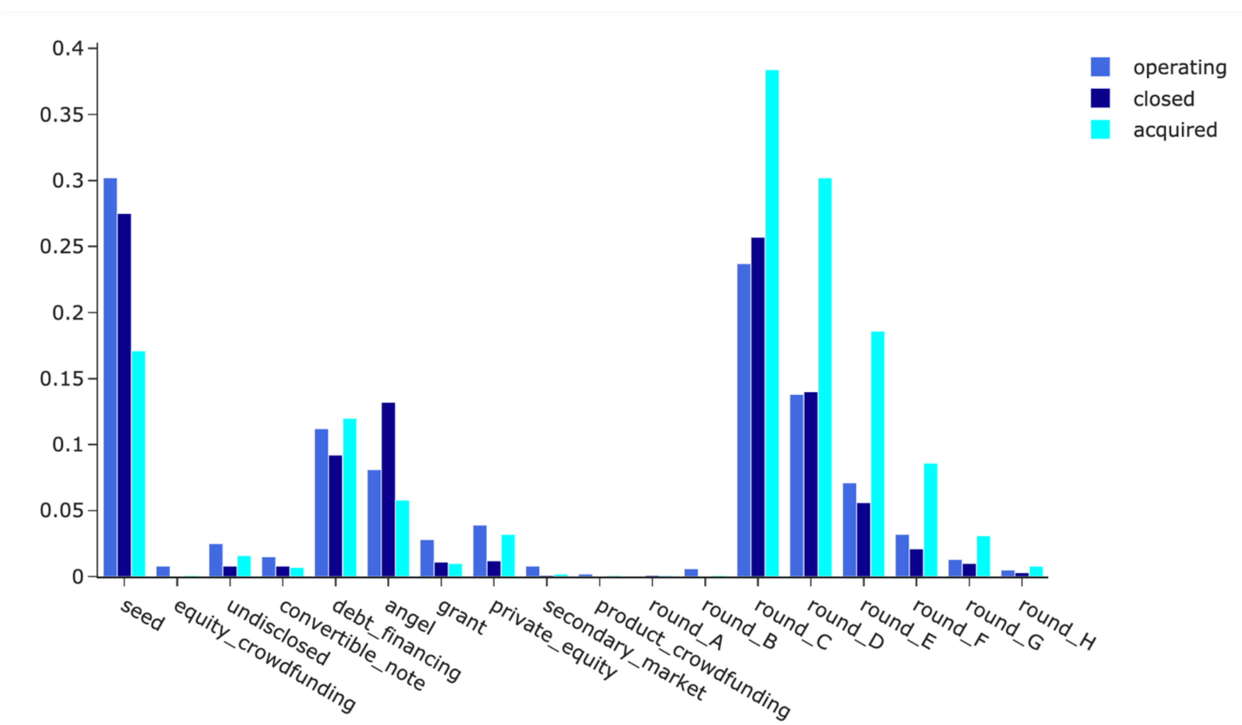
Appendix 2.6: Average acquisition rate by country prior to 2014

Country	Acquisition Rate
USA	12%
GBR	6%
CAN	10%
FRA	6%
CHN	2%
ISR	10%
DEU	7%
IND	3%
ESP	6%
SWE	5%

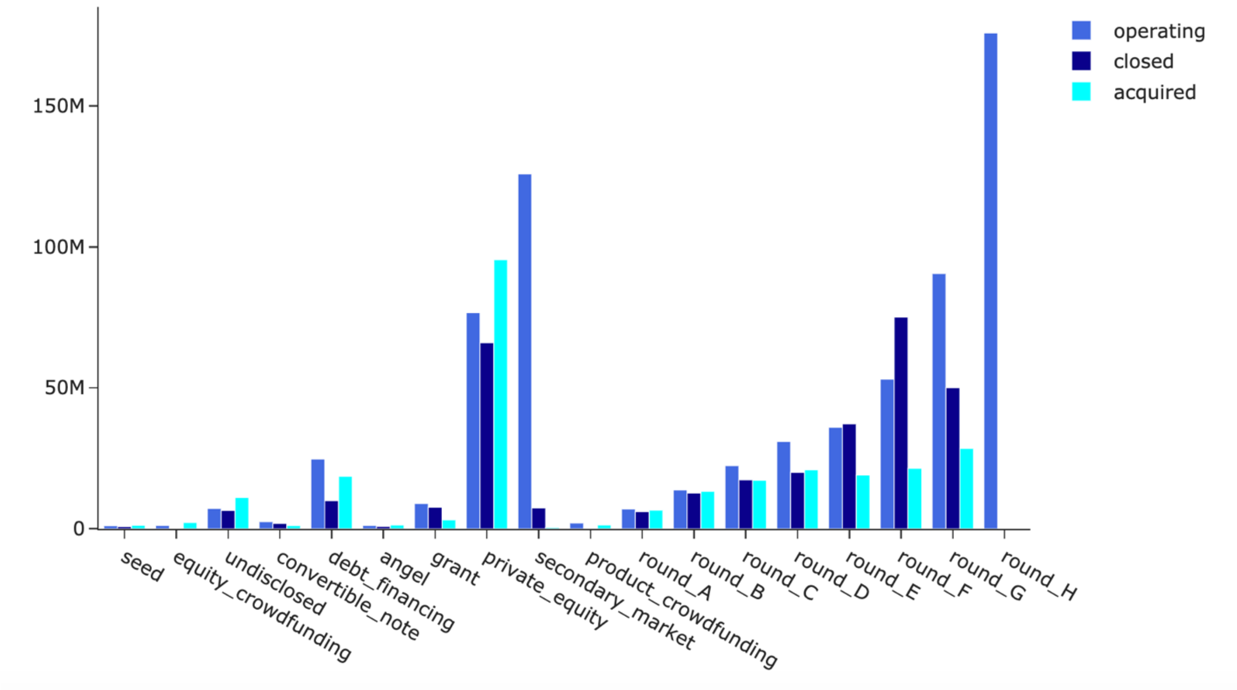
Appendix 2.7: Average funding raised by status prior to 2014



Appendix 2.8: Funding source distribution by status prior to 2014



Appendix 2.9: Average funding raised per funding source for each status prior to 2014



Appendix 2.10: Survival, acquisition and closing rates by funding source prior to 2014

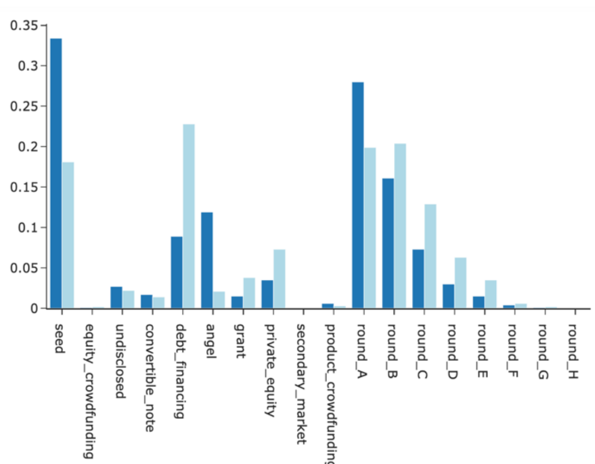
Funding Source	Survival Rate	Closing Rate	Acquisition Rate
seed	95%	5%	5%
equity_crowdfunding	100%	0%	2%
undisclosed	98%	2%	6%
convertible_note	97%	3%	4%
debt_financing	96%	4%	10%
angel	92%	8%	6%
grant	98%	2%	3%
private_equity	98%	2%	8%
secondary_market	94%	6%	11%
product_crowdfunding	100%	0%	1%
round_A	95%	5%	14%
round_B	95%	5%	18%
round_C	96%	4%	21%
round_D	97%	3%	21%
round_E	97%	3%	19%
round_F	96%	4%	15%
round_G	97%	3%	9%
round_H	100%	0%	0%

Appendix 2.11: Survival and acquisition rates by sector prior to 2014

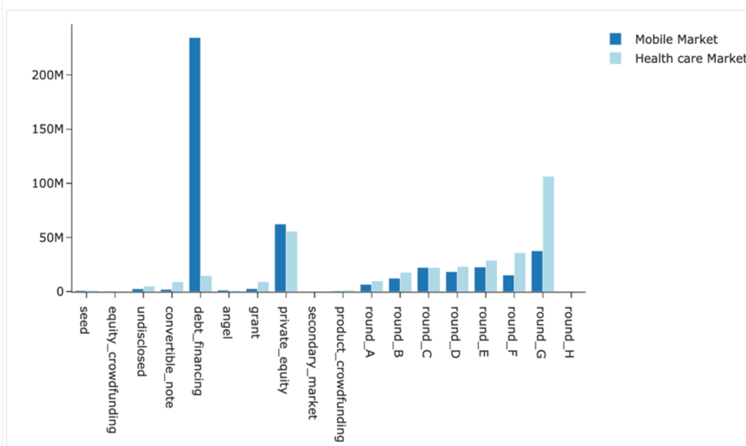
Market	Total Funding Secured	Survival Rate	Acquisition Rate
Software	$ 8,692,068	95%	12%
Biotechnology	$ 16,334,066	97%	5%
Mobile	$ 9,590,691	93%	12%
E-Commerce	$ 14,063,307	95%	7%
Enterprise Software	$ 14,234,193	97%	18%
Curated Web	$ 6,808,973	88%	12%
Health Care	$ 19,461,404	97%	6%
Hardware + Software	$ 9,885,858	95%	8%
Advertising	$ 14,612,191	94%	15%
Games	$ 8,685,024	92%	13%
Clean Technology	$ 25,663,745	94%	6%
Health and Wellness	$ 7,369,028	98%	3%
Finance	$ 10,218,464	96%	7%
Social Media	$ 5,403,797	88%	11%
Education	$ 5,050,545	98%	5%

Appendix 2.12: Funding source distribution for mobile and health care markets

Proportion per funding source

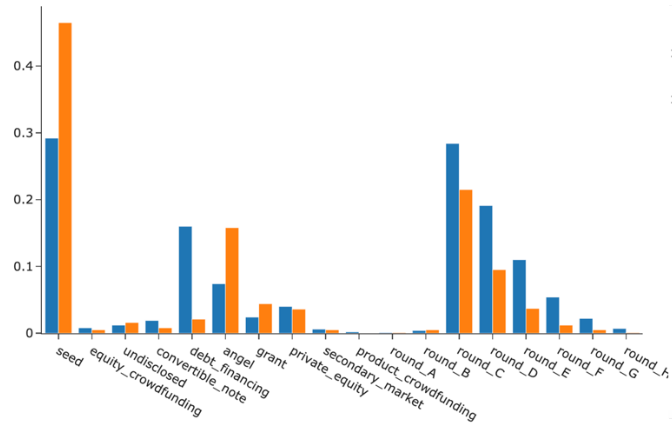


Average funding raised per source

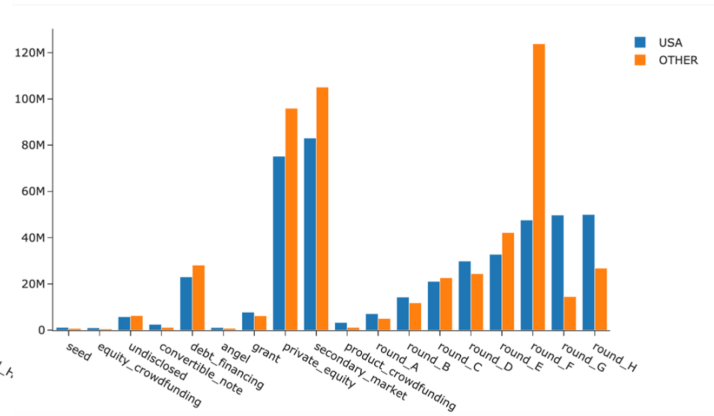


Appendix 2.13: Funding source distribution for USA and other countries

Proportion per funding source



Average funding raised per source

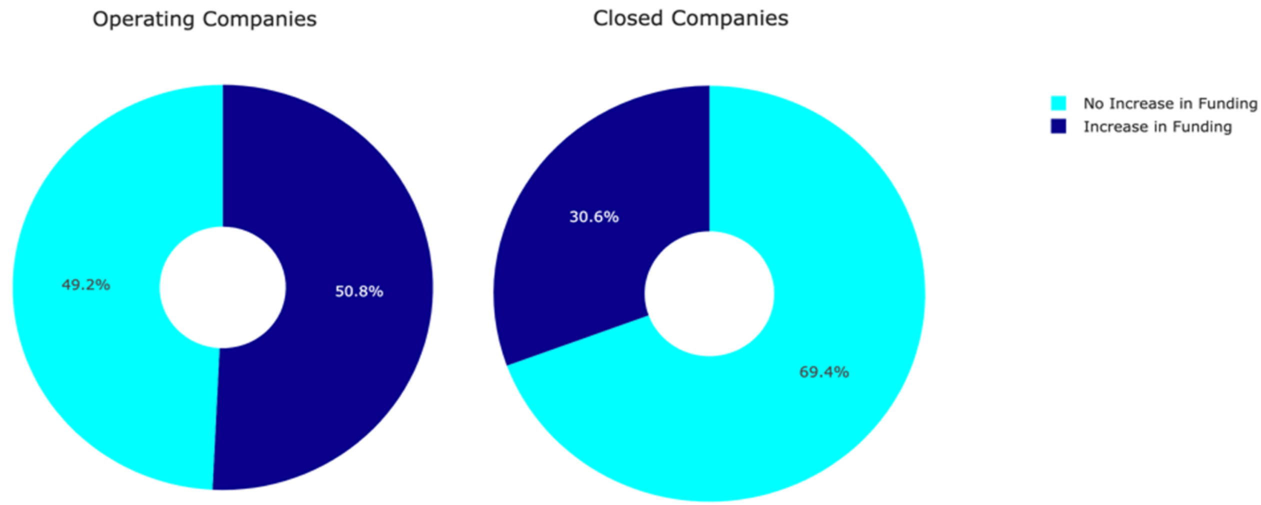


Appendix 2.14: Logistic regression results

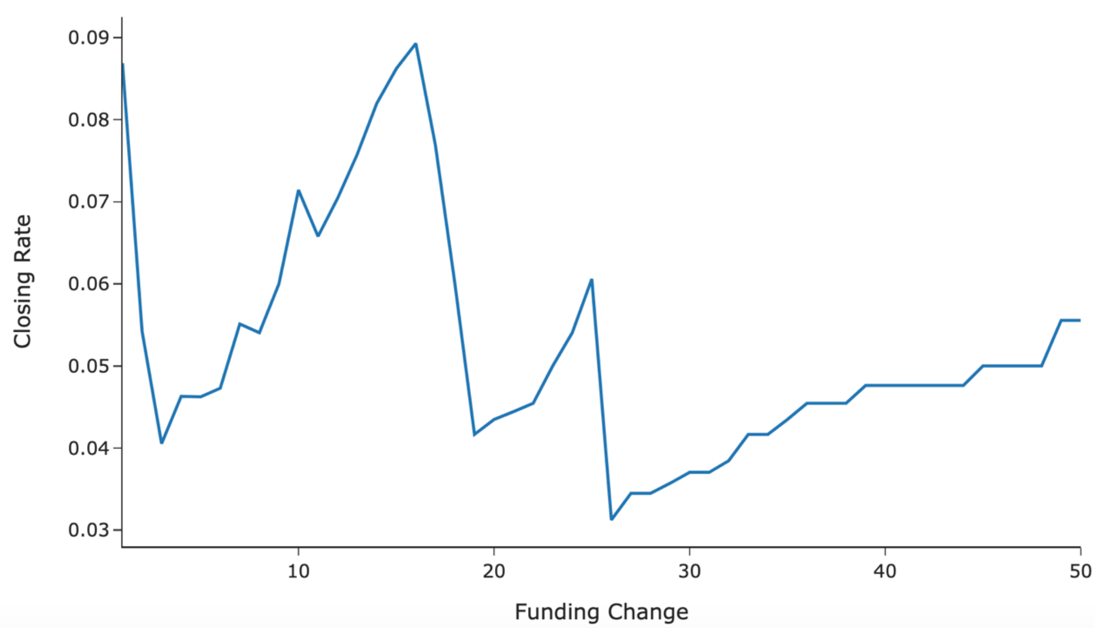
Target variable = acquired status (binary variable)

	Dependent variable:
	acquired
funding_total_usd	-0.000
	(0.000)
funding_rounds	0.051***
	(0.015)
seed	-0.00000***
	(0.00000)
equity_crowdfunding	-0.00000
	(0.00000)
undisclosed	0.000
	(0.000)
convertible_note	-0.00000**
	(0.00000)
debt_financing	-0.000
	(0.000)
angel	0.000
	(0.00000)
grant	-0.00000***
	(0.00000)
private_equity	0.000
	(0.000)
post_ipo_equity	0.000
	(0.000)
post_ipo_debt	0.000
	(0.000)
secondary_market	-0.00000
	(0.00000)
product_crowdfunding	-0.00000
	(0.00000)
round_A	0.000***
	(0.000)
round_B	0.000***
	(0.000)
round_C	0.000**
	(0.000)
round_D	0.000
	(0.000)
round_E	-0.000
	(0.000)
round_F	-0.000
	(0.000)
round_G	0.000
	(0.000)
round_H	-0.00000
	(0.00003)
company_life_duration	0.0001***
	(0.00001)
CAN	0.834***
	(0.159)
CHN	-0.898***
	(0.312)
DEU	0.662***
	(0.211)
ESP	0.431
	(0.264)
FRA	0.264
	(0.214)
GBR	0.378**
	(0.149)
IND	-0.517*
	(0.313)
ISR	0.954***
	(0.183)
SWE	0.072
	(0.359)
USA	1.065***
	(0.103)
Advertising	0.520***
	(0.119)
Biotechnology	-0.965***
	(0.104)
Curated Web	0.423***
	(0.111)
E-Commerce	-0.102
	(0.138)
Enterprise Software	0.577***
	(0.097)
Games	0.481***
	(0.129)
Hardware + Software	-0.408***
	(0.148)
Health Care	-0.764***
	(0.161)
Mobile	0.356***
	(0.098)
Software	0.186***
	(0.066)
Constant	-3.673***
	(0.107)
Observations	25,374
Log Likelihood	-7,522.551
Akaike Inf. Crit.	15,133.100
Note:	*p<0.1;* ***p<0.05;*** p<0.01

Appendix 3.1: Funding change per status (2014-2020)



Appendix 3.2: Closing rate in function of funding change level (2014-2020)



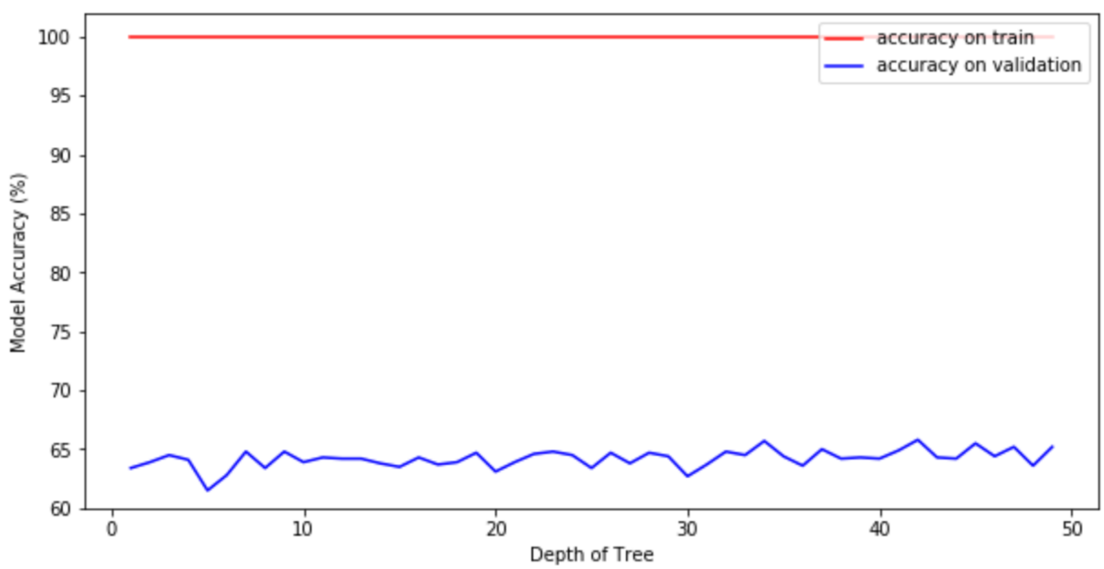
Appendix 3.3: Survival rate and funding change (2014-2020) in top 5 most frequent country

Funding Source	Survival Rate	Funding Change (2014 -2020)
USA	90%	10.78
GBR	94%	3.07
CAN	93%	2.99
FRA	93%	3.36
IND	95%	6.18

Appendix 3.4: Survival rate and funding change (2014-2020) in top 10 most frequent markets

Funding Source	Survival Rate	Funding Change (2014 -2020)
Software	93%	3.45
Biotechnology	93%	6.25
Mobile	96%	3.29
E-Commerce	91%	6.12
Curated Web	89%	6.17
Health Care	91%	3.26
Hardware + Software	97%	5.81
Enterprise Software	93%	3.25
Advertising	93%	2.04
Clean Technology	96%	2.79

Appendix 4.1: Decision tree performance on training and validation sets in function of tree depth



Appendix 4.2: Summary results of classification models and logistic regression

Column1	Nearest Neighbors	Decision Tree Classifier	Logistic Regression Model	Supper Vector Machines	Random Forest Classifier
Normalization required?	Yes	No	Yes	Yes	No
Accuracy training set	66.2%	100.0%	70.9%	72.2%	83.8%
Accuracy validation set	67.0%	65.4%	66.1%	64.8%	70.4%

Appendix 4.3: Final performance of random forest classifier on unseen test set

Acquired(1) - Non-acquired(0)	Precision	Recall	F1-score	Support
0	70%	76%	73%	447
1	74%	68%	71%	457

8. References

[1] Kupor, S., 2019. *Secrets Of Sand Hill Road*. Virgin Books, pp.36-38.

[2] Staniewski, M.W., 2012. Setting up a business and funding sources. Journal of Business Research, 1(3): 56-60.

[3] Kupor, S., 2019. *Secrets Of Sand Hill Road*. Virgin Books, pp.112-123.

[4] Anirudh Garg and Abhishek Krishna Shivam., Funding to growing start-ups. Research Journal of Social Sciences, 10(2): 22-31, 2017

[5] Ramadani, Veland, 2012. The Importance Of Angel Investors In Financing The Growth Of Small And Medium Sized Enterprises. International Journal of Academic Research in Business and Social Sciences. 02.

[6] Crunchbase | Knowledge Center. 2020. *Glossary Of Funding Types*. [online] Available at: <https://support.crunchbase.com/hc/en-us/articles/115010458467-Glossary-of-Funding-Types> [Accessed 11 August 2020].

[7] Strebulaev, I. and Gornall, W., 2015. *How Much Does Venture Capital Drive The U.S. Economy?*. [online] Stanford Graduate School of Business. Available at: <https://www.gsb.stanford.edu/insights/how-much-does-venture-capital-drive-us-economy> [Accessed 13 August 2020].

[8] Platt, E. and Edgecliffe-Johnson, A., 2020. *Wework: How The Ultimate Unicorn Lost Its Billions*. [online] Ft.com. Available at: <https://www.ft.com/content/7938752a-52a7-11ea-90ad-25e377c0ee1f> [Accessed 27 August 2020].

[9] Blankenship, R., 2020. *Biotech Startup Rankings - #1 Top Fastest Growing Biotech Startups*. [online] Angel Kings | We Design, Build & Launch America's Top Startups. Available at: <https://www.angelkings.com/biotech-startups> [Accessed 19 August 2020].

[10] Kuhn, M. and Johnson, K., 2013. *Applied Predictive Modeling*. 1st ed. Springer, p.67.

[11] Raschka, S., 2014. *About Feature Scaling And Normalization*. [online] Dr. Sebastian Raschka. Available at: <https://sebastianraschka.com/Articles/2014_about_feature_scaling.html> [Accessed 21 August 2020].

[12] Kerdprasop, N. 2011, Proceedings from 7th 10th 11th and 13th WSEAS international conferences. Discrete decision tree induction to avoid overfitting on categorical data.