NORTHEASTERN UNIVERSITY

College *of* Professional Studies

ALY6080 Integrated Experiential Learning

#### Final Group Project Proposal



|  |  |
| --- | --- |
| Professor: | Dr. Sharad Shandilya |
| Students: | Nisarg Bhatt  Avinash Tripathi  Barsha Rajbhandari  Duong Khanh Ha Nguyen |

June 25, 2020

**Project Overview**

The main objective of this project is to utilize the current RedCrow evaluation process to identify important features for short-term and long-term performance of early-stage healthcare startups getting listed on RedCrow’s platform. We will be using both business and clinical evaluations from RedCrow, and external datasets and research pertaining to crowdfunding as a whole and also specifically healthcare domain to conduct market research and provide insights to work towards the objective.

For short-term performance, we consider the “12 month milestone” as our target variable, and inspect which features are important for startups to be able to reach the goal. Furthermore, to support and provide guidance to reach this milestone, we analyzed the records of companies from the Securities and Exchange Commission(SEC) to provide guidance on what type and quantity of securities are recommended to startups that belong to different financial positions.

For long-term performance, we are inspecting important features that will help startups survive for at least 3 years in the market once they are founded. We looked at this component of startup performance as we came across multiple literature reviews that emphasized on the issue that one of the biggest challenges for startups is to survive the first couple of years. Moreover, as we inspected RedCrow’s evaluation process- we noticed features about 36 month exit and experience, which solidified the idea that this concept is relevant and already in the sponsor’s radar.

Lastly, we added a bonus component in our project where we analyzed the sentiments of comments in both business and clinical evaluations and segregated positive and negative comments so that the end-user(sponsor,startup) can efficiently look at negative comments without having to skim through the entire evaluation.

1

A significant portion of our project modeling is based on the sponsor's current evaluation methodology, and thereby our biggest assumption is that the evaluators have scored with due diligence.

**Literature Review**

Redcrow is an equity crowdfunding platform that helps bring the entrepreneurs and the potential investors under one umbrella. It leverages industry experts to vet startups, enabling a wide array of investors to help healthcare innovations grow. Along with that Redcrow also helps in running attractive funding campaigns through their online platform and keeps the investors engaged by sharing the policies of the startup, detailed deal terms of the raise, investment traction reporting as it relates to the current round of funding, and the ability to view comments and feedback from those who visit the site. (Redcrow, 2019)

The industry is changing from reward-based crowdfunding to investment-based crowdfunding and even more so equity crowdfunding. In reward-based crowdfunding, funders are promised a product or service in return for their contribution and depending on the project funders could get credits or access to other creative input. Investment-based crowdfunding comprehends lending and equity. Equity crowdfunding is appealing to potential lucrative startups since the possibility of monetary gain is very high. Moreover, since the funding contribution per investor is significantly larger, the crowdfunding sites are more driven at ensuring the credibility of the campaigns to maintain their own reputation. Likewise, in case of healthcare startups, there is an increasing trend of using innovative engineering and especially artificial intelligence in building healthcare products. Companies such as Buoy Health, PathAI, and Enlitic are based on increasing use of AI to check symptoms and algorithms to diagnose and treat illness (Daley,2020).

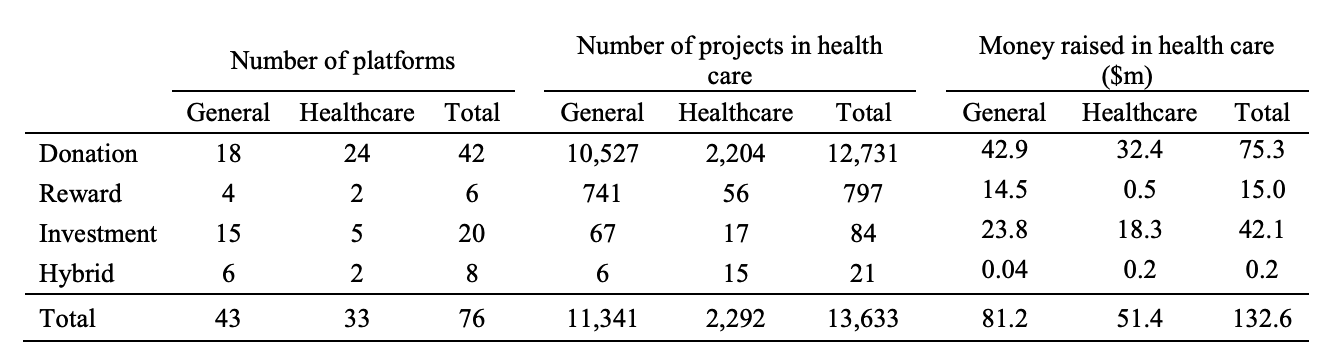
Zhang al. (2017) mentions that one of the reason behind the success of the startup is its ability to raise fund at an early stage and their major insight is that the attributes of completeness in profile and social engagement have positive correlation in crowdfunding success. he also mentions that company having an introductory video are 11.5 times more likely to generate funds than the ones who don’t. Also, In one of the research articles, it was mentioned that far too many startups succeed in achieving a regulatory milestone but then run out of funding before generating sufficient evidence to convince customers to buy their products. (Grand, P., Zavala, K.)

**Analysed Datasets**

We conducted exploratory analysis on the spreadsheets provided by Redcrow, 4800 health care startup research records from 2000-2014 on angel.com, 2019 fourth quarter equity crowdfunding data available in the Securities and Exchange Commission (SEC) website and crowdfunding dataset of indiegogo crowdfunding platform which contains more than 32000 entries, over the period of 2010- 2020.

**Domain Research**

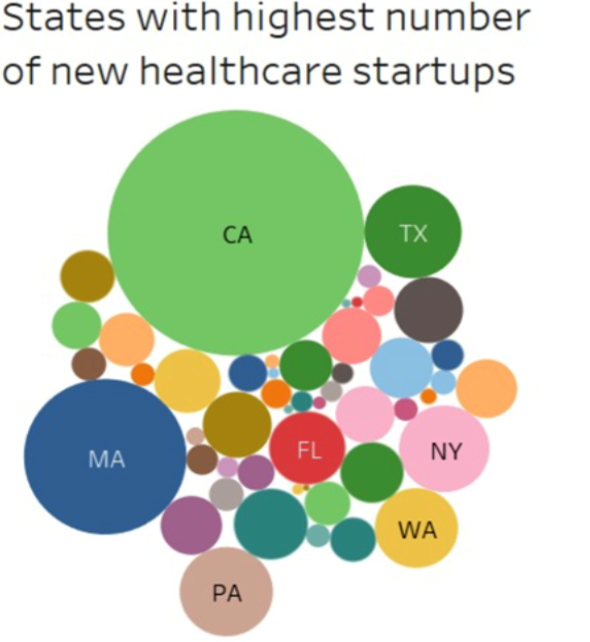
We looked at some past market research papers, in one of the research papers, which was based on 76 healthcare crowdfunding platforms. We noticed that there are different ways through which funds are generated, which are: (1) donations, (2) a future product, service, or some other reward, or (3) exchange for shares or debt securities of a company and lastly, (4) the hybrid form of funding which includes multiple options of funding given by the platforms.



This table shows the number of different crowdfunding platforms by type, the number of projects and the amount of money raised through different types of funding. The table gives information from the foundation of that platform till october 31,2017. As you can see that most of the funding is done in the form of donation and then comes the ones who are in the forms of investment. The detailed list of platforms which were considered for this study is mentioned under the reference(5) - appendix1 section of this document.

The research also talks about the the level of dispersion that is present in these startups even though the average number of successful healthcare projects per platform is 180, it is observed that there are only 4 platforms who has raised more than 500 successful projects and able to receive the funding (YouCaring, 10 Giveforward, Indiegogo, Kangu). while 38 platforms reported fewer than 100 successful health campaigns. Redcrow provides an investment type of crowdfunding and the investors mainly consist of people coming from doctoral, scientific or business backgrounds, who either have knowledge of the healthcare domain or know how the business works. If in the future, RedCrow wants to expand in any way then this could be a good opportunity to gain market visibility and help attract more investors to their platform by looking into different forms of funding options that are available.

**Exploratory Data Analysis**

We conducted EDA to understand the overall spread of healthcare equity crowdfunding in the United States. From the data obtained from angel.com, we see that a majority of the healthcare startups founded between 2000-2014, are based out of California and Massachusetts. If we want to look at a granular level, San Diego, San Francisco have the most startups in California and. Likewise, Cambridge and Boston have the most startups in Massachusetts. This information gives us a guideline of the market concentration of our domain.

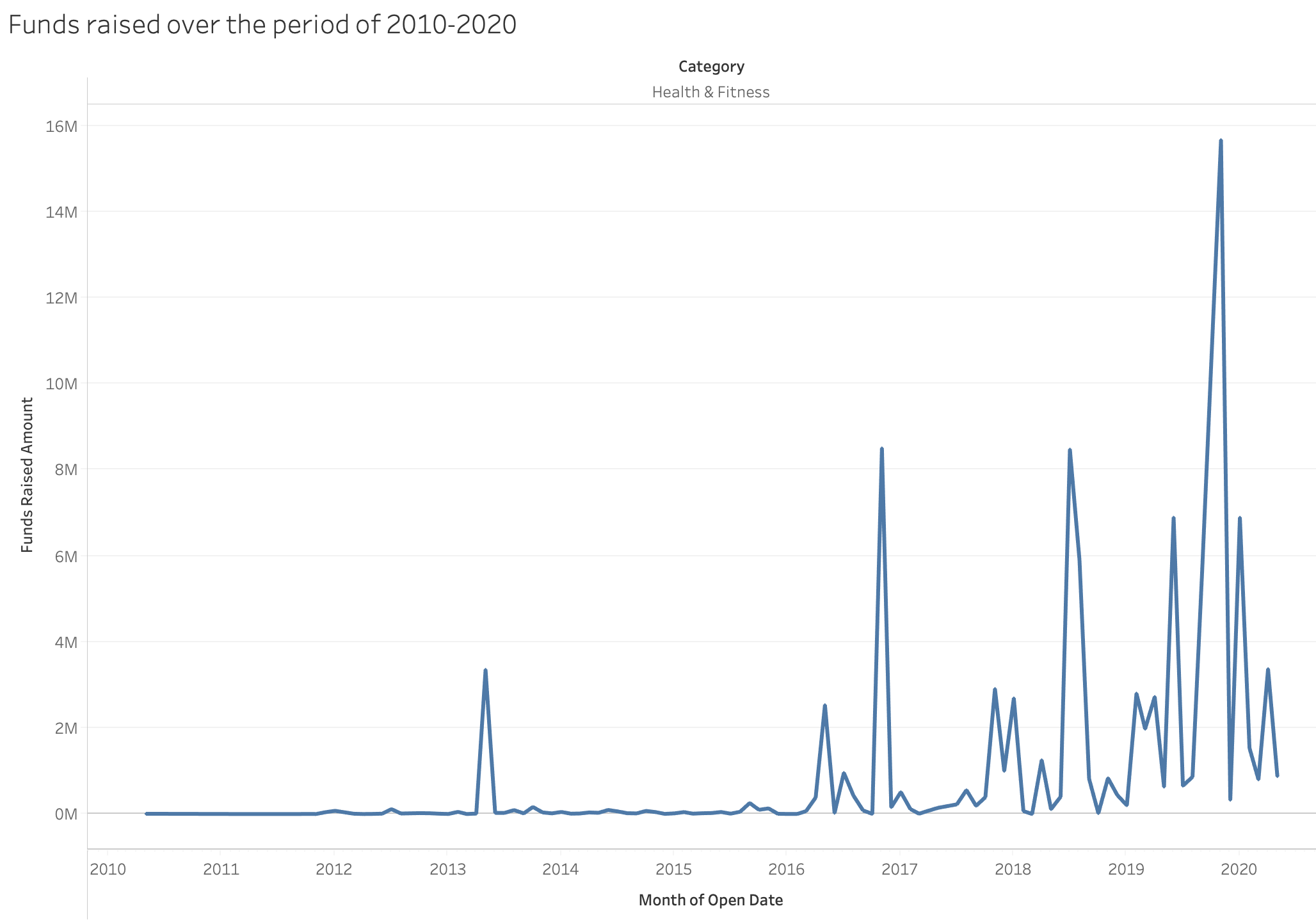
Next, we divided the overall healthcare startups into three categories, which are biotechnology, medical and health. An interesting finding was that about 66% of startups fell into the biotechnology category, followed by 19% medical and 15% health. However, the medical category startups are the ones that raise the highest average funding across the healthcare field. This information provides us an idea about the funding requirements of different types of healthcare startups.

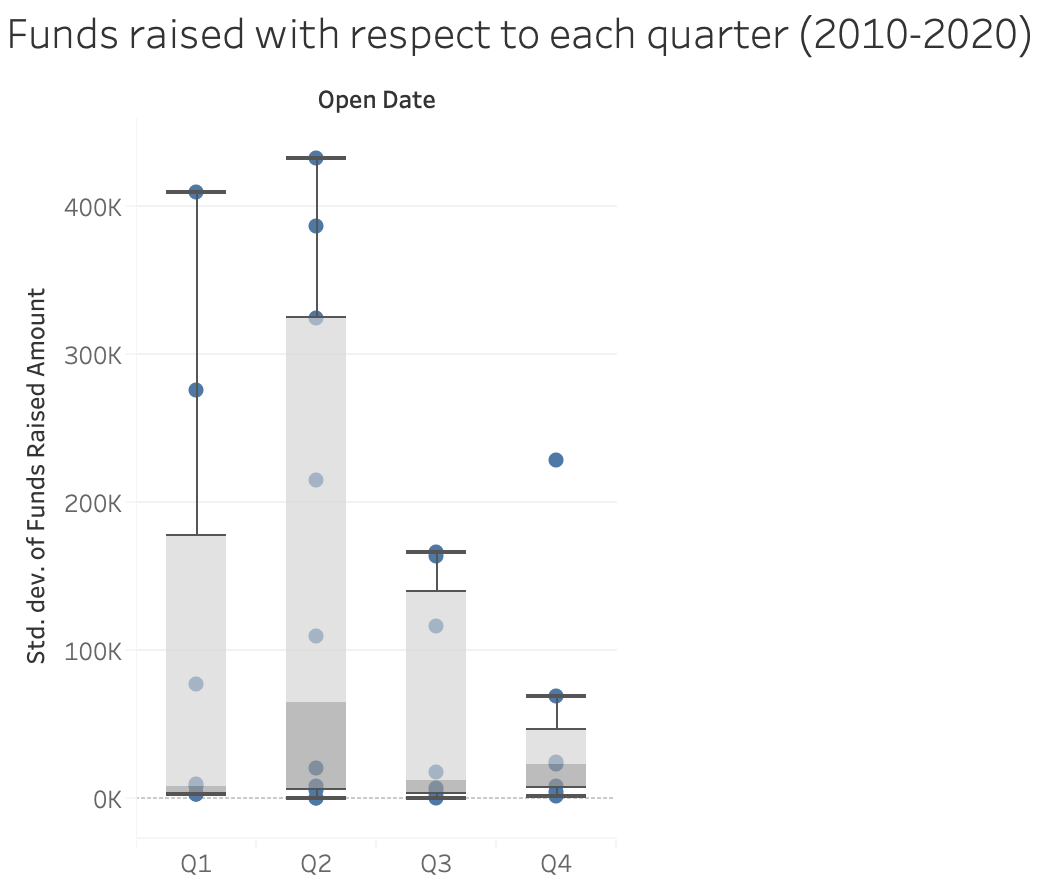
A screenshot of a cell phone

Description automatically generated

As a crowdfunding platform, Redcrow’s main objective is to look for as many investors as possible. For our project, we looked into a dataset from other crowdfunding platforms known as IndieGoGo. We found data through a web scraper robot which collects data for every month from IndieGoGo's official platform. This dataset contained more than 32000 entries and 25 features as a whole. As Redcrow is mainly focusing on the healthcare domain, hence, we filtered the found dataset for the category of startup’s which were related to health and fitness. The dataset had features such as open\_date, close\_date, funds\_raised, and funds\_raised\_percent.

Now, coming to the analysis, we firstly checked how well the IndieGoGo platform can generate funds over time. For this, we thought of implementing a time series plot that can help us understand the startup’s performance over every year starting from 2010 to 2020. We can see the total amount of total funds generated concerning every year in the below chart.





After removing some of the outliers, we came up with the above presentation, like the October month of the year 2019, which is surprisingly able to generate more than 20 million within a month even when we just consider the healthcare side. Removing that month gave us a better idea of trends and of how gradually Indiegogo was able to raise funds over the period. we can see that after 2016 they were able to attract more investors onto their platform.

We analyzed funding for every quarter as well and got the above chart. This chart shows which are potential quarters to generate more funding. As you can see that Q2 is the quarter which usually generates the maximum funding when compared with the other quarters and Q4 is the quarter of least fund generation.

**Analysis Methodologies**

1. **Using NLP for Sentiment Analysis**

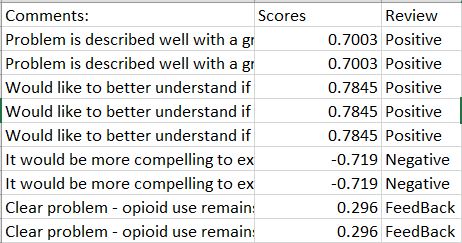
By using the nltk toolkit, we conducted NLP on the comments hoping to get some valuable insights that we could use for the data analysis of the startups. Using the Sentiment Analyzer, we were able to come up with the sentiment scores for each of the comments, this score was between the +1 and -1. Positive score indicated that the comment had positive sentiment and vice versa. These scores are also called the compound score, which basically weights the positive and the negative words in the comments and creates a generalized score. We have plotted these scores on a chart which represents the sentiment variation as shown below.

A screenshot of a cell phone

Description automatically generated

We can observe that there are only 2 comments for this current sample case which have negative sentiments. The reason for this approach is let's assume we have hundreds of comments for a given startup, so in order to save the time and rather than going through every comment in order to understand the context, in the initial phase we can only focus on the comments which have negative sentiments and try to take in considerations the suggestions mentioned.

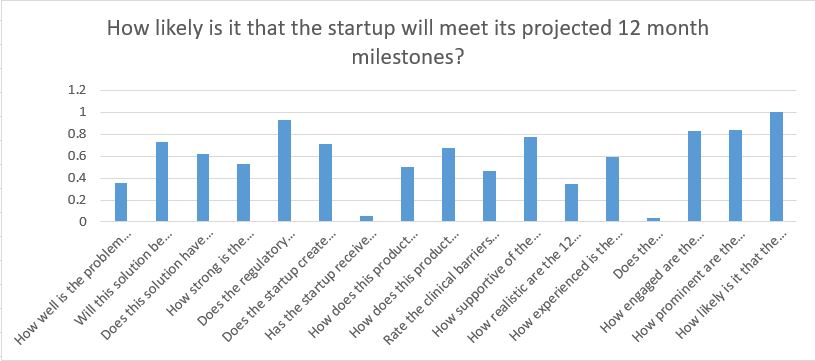
It becomes difficult to evaluate the score which has a range close to 0, as we cannot justify the score as positive or negative. So, we have planned to give the positive sentiment only to those comments which have scored greater than 0.4 and Negative sentiment to scores less than -0.4. Rest of the comments score, we have a Feedback system. These comments will be reviewed by the experts/investors in order to understand the details of the comments. This will help in better evaluation of the comments.



1. **Numerical weighted Matrix**

Apart from this, we have created a numerical weighted matrix. This basically shows the relation of each clinical evaluation feature with the dependent variable of how the startup would perform in the next 12 month period. This step evaluates the weight of each feature which would provide evidence to create the numerical scoring matrix.

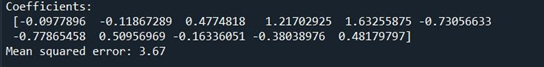




1. **Model Creation**

The final step was to create a model just using the numerical features with the idea that if we can come up with a model, we would have a prioritized list of features that affect our target variable. We would want our RedCrow team to focus on improving the features that are closely related to target variable.

We implemented simple linear regression and came up with the weighted feature matrix coefficients and the mean squared error. The results for the same are shown below.

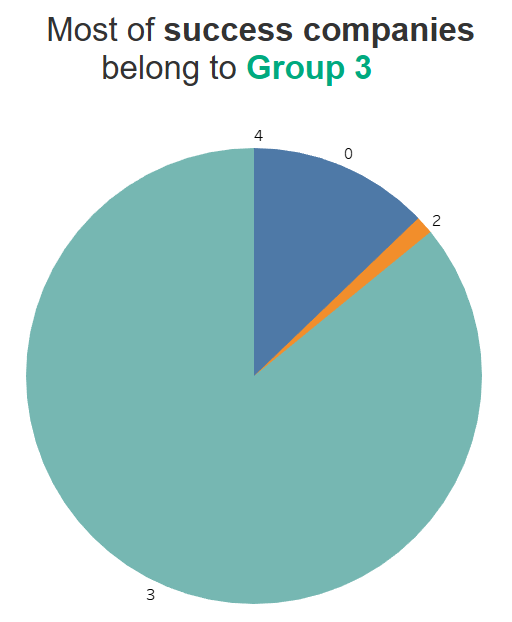


**Guidance for startups on crowdfunding platforms**

It is important that the startups get success in fund calling on crowdfunding platforms to obtain their 12-milestone target. In order to understand more about this type of fundraising and how startups from all fields get success on this platform, we are going to get insights from the dataset of Securities and Exchange Commission (SEC) that provides information about securities offered on crowdfunding platforms.

Since companies joining crowdfunding are from many different financial situations, we would like to cluster them into different groups using unsupervised machine learning methods since all of them are unlabeled. This activity would localize the data to make more accurate and effective recommendations for the companies in the similar situation.

Elbow algorithm has been used to find out the optimal number of groups for clustering. Then we apply K-means method to categorize startups into 5 groups (as the result of the Elbow method). At the same time, we also filter the crowdfunding offers that have closed successfully to see if they specifically belong to any financial group. The reason why we did not only filter successful companies to get insights is that the results of recommendations might be influenced by the strong financial background of startups, while these are only the minority in the crowdfunding community.

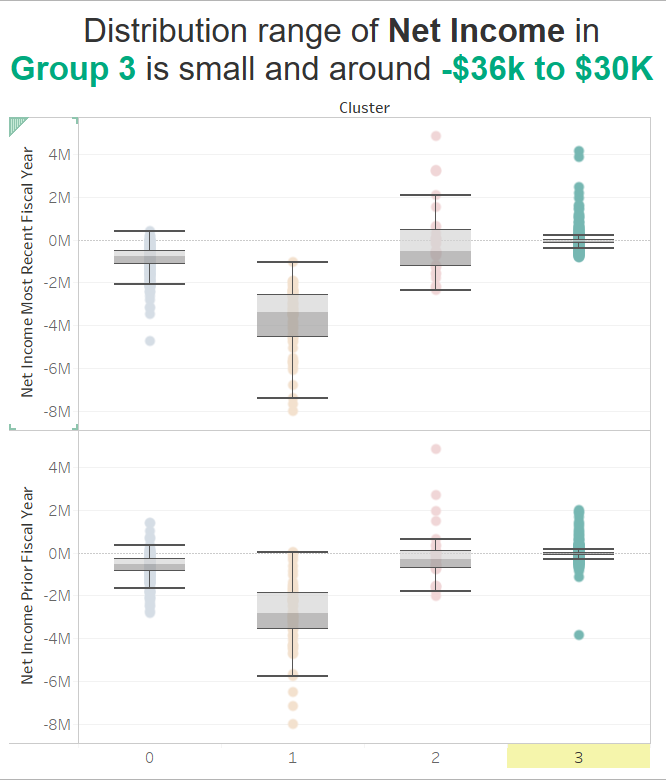
It turns out that Group 3 is not only the typical financial status on this platform but also has the largest number of successful fundraising. From that results, we would like to see how to describe group 3 in terms of finance compared to other groups.

Interestingly, this group does not have a strong financial background at all. Its distribution in Accounts Receivable and Cost of Goods sold in the prior and recent fiscal year is low and narrow compared to other groups. (Since Group 4 only has one member, we excluded it after the first graph).



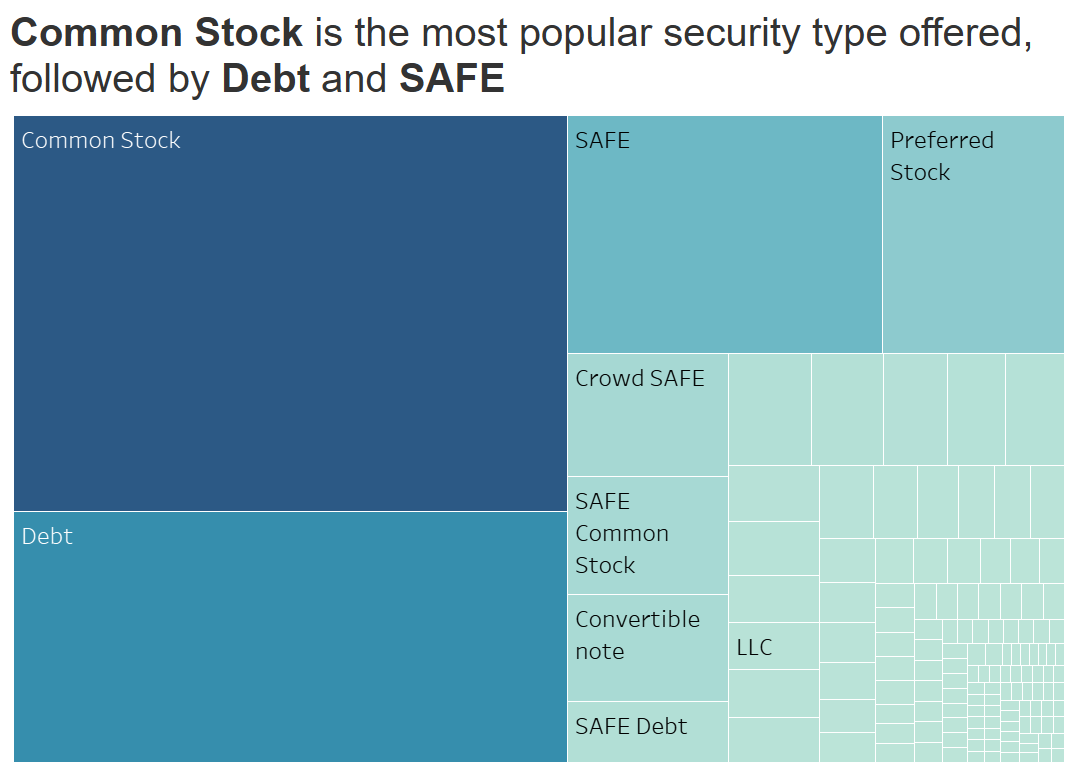
Group 3 distribution in debt presents pretty the same pattern. It can be seen that although the cash flow of this group is lower than others, its debt situation is not as worrying.



Finally, we want to look at net income in fiscal years of the 4 groups. Even when the net income of companies is not always positive, we can see that Group 3’s values concentrate in a narrower range of amounts compared to other groups.

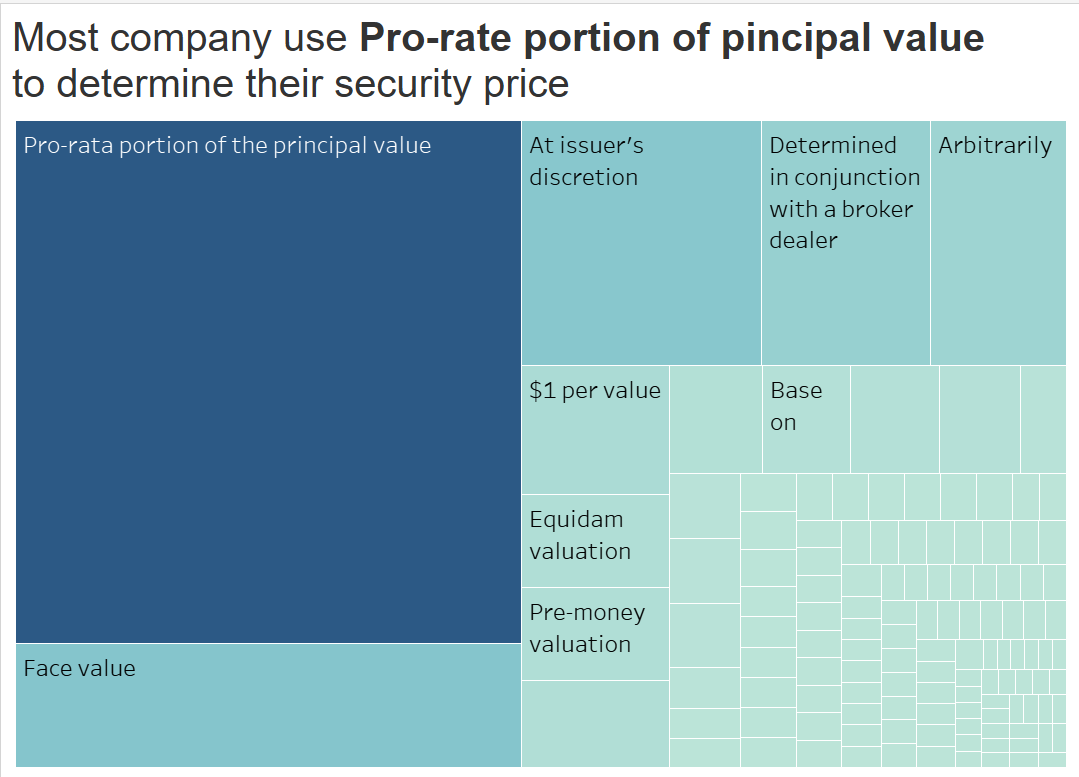
After investigating the financial characteristics of group 3, we would like to find out how companies in this group perform on crowdfunding platforms.

According to an empirical examination about the signaling equity crowdfunding (Ahlers G.K.C, 2015), it’s revealed that investors do have more concern about security offers and the level of detail of the financial projections as well as the human capital of the company. Therefore, we decided to take a look at disclosure information about securities offered on crowdfunding platforms to give Redcrow’s startups some recommendations.

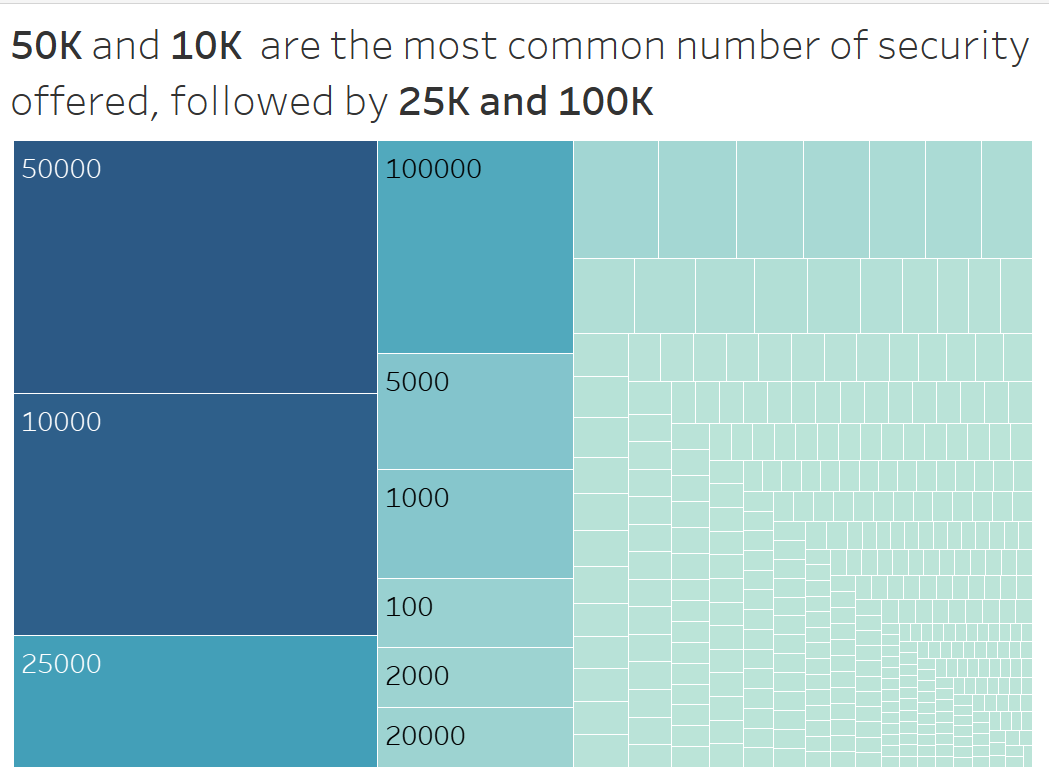


While common stock is the most popular normal security type, many companies chose SAFE (simple agreement for future equity) as their security to offer on crowdfunding platforms. This type of share is suitable for venture capitalists to invest quickly in hot startups without burdening startups with further negotiations.

It depends on the projects’ features that we can give recommendations for the Redcrow’s startups on the type of security they should offer on the platform. If it is a contemporary project that responds to certain social activities, we would recommend a SAFE type. Otherwise, common stock is preferred if the company is still immature and we suggest Redcrow’s startups use this type since Redcrow’s investors might want a long-term investment and share the company’s success for a long time.



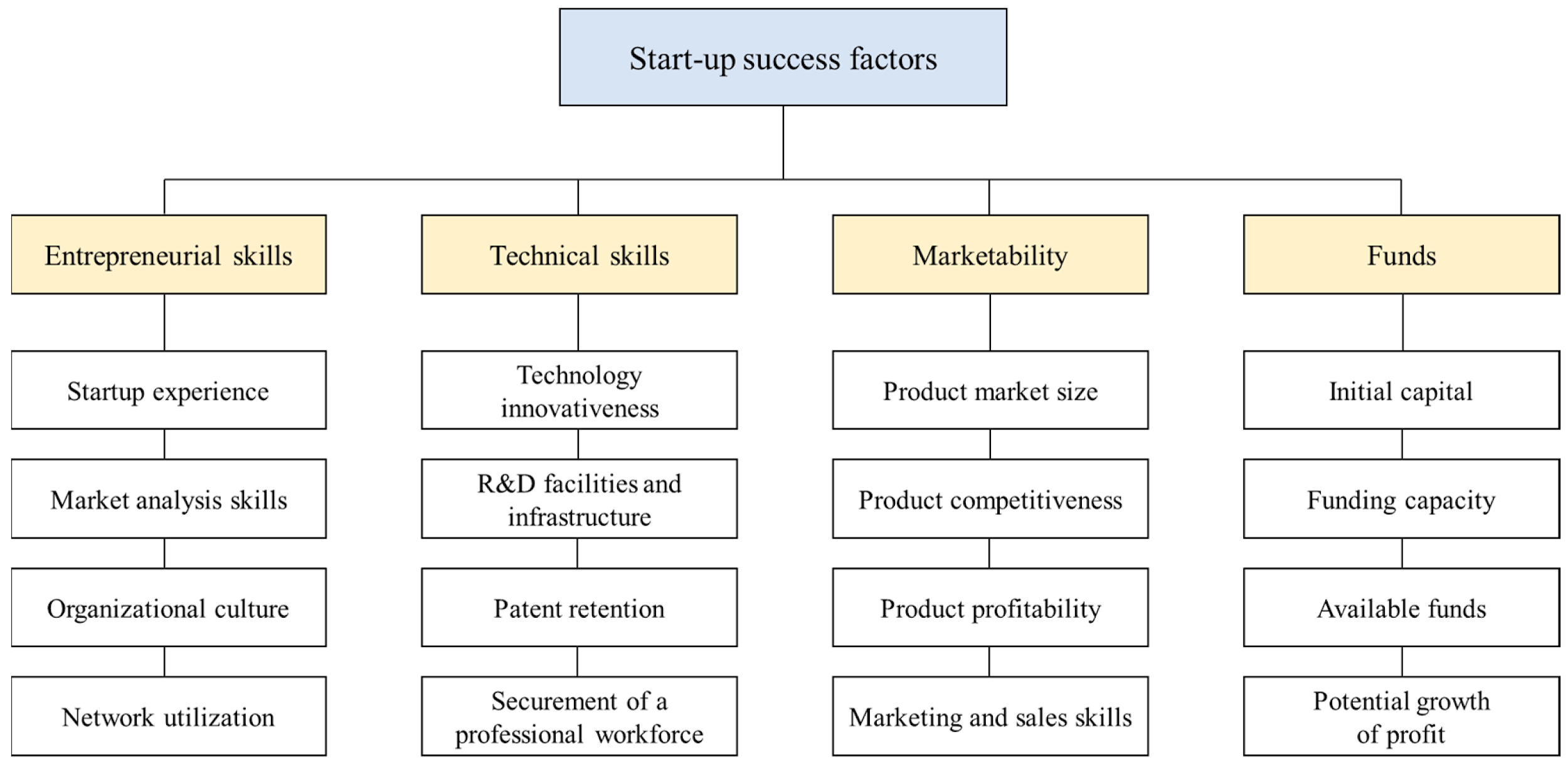
Still in the concerns of financial projections, we want to know what method is commonly used to determine the security price to suggest our startups. It reveals that nearly a fourth of the number of companies use pro-rated portion of the total principal to evaluate their security price. Face value and the issuer’s discretion are also two normally used methods. However, many startups have got help from the third party such as an intermediary, broker dealer, or from Equidam website. Therefore, apart from traditional methods of determining the security cost, we would recommend the startups define it on some market research or consult the third party to have a more competitive position in the market.



The number of securities offered also varies, but 10,000 or 50,000 is the most favorable number of shares opened. Although this could depend on the company’s scale and their plan to get funded on this platform, that could be suggestions for our healthcare startups not only to adjust their funding plan but also to help them to determine the appropriate share price.

**Long-term startup performance**

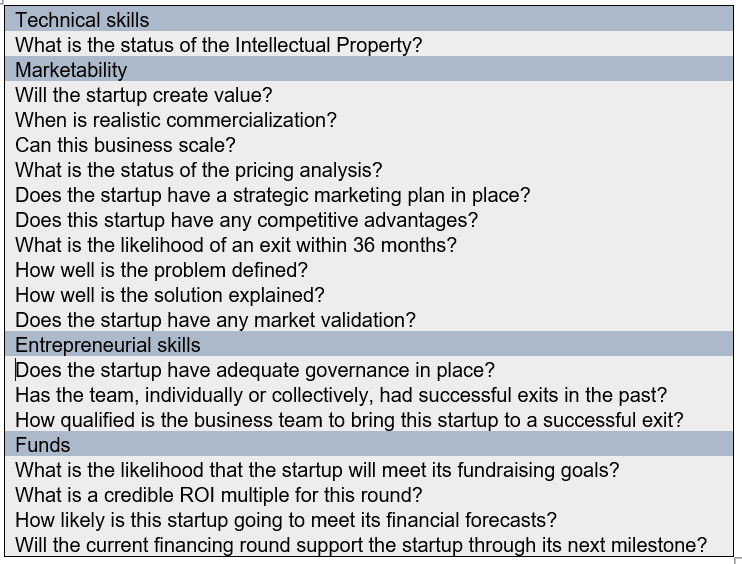
Now that we established the important features for startups to reach the 12month milestones along with guidelines in approaching the equity market given the financial position, our next step is to analyze the long term survival. This will help us identify strengths and weaknesses of startups in granular level providing a guide for startups to focus on different components. We found a research conducted in Korea on medical industry startups in the market that analyse the success features for survival. We used Analytic Hierarchical Process(AHP) study from the research, which is a method for ranking various factors and distinguishing relevant importance. The study derived the AHP method using pairwise comparison. The four categories are Technical Skills, Marketability, Entrepreneur skills, and Funds. We learnt that the employee expertise is one of the most important features, followed by the product competitiveness and marketability that help in generating profits. Likewise, the network utilization and experience of the entrepreneurs are also important factors, which aids in strong commercialization of the product followed by the funding capacity.



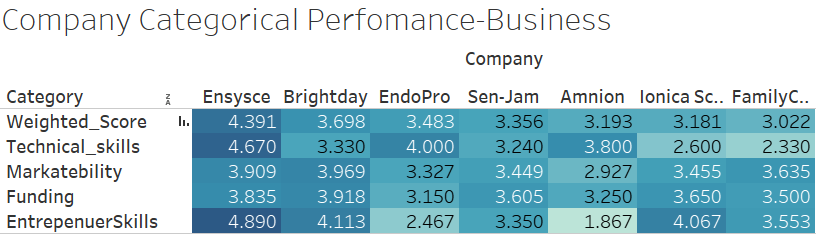
Each category has several subsections, and we took the subcategory guidelines to divide RedCrow’s current evaluation features. Most of these are covered by features in RedCrow’s evaluation. However, in order to get a better picture of the startups being listed, adding questions pertaining to current available funds, potential profit growth, tentative revenue or time frame needed to breakeven, entrepreneur network utilization (involved with relevant medical associations), what percent of the funding goal sought on RedCrow’s platform makes up for the total capital required by the startup, would be helpful in filling the gaps currently present in the sponsor’s evaluation methodology.

**Business evaluation**

The business evaluations were consolidated and a weighted\_score was calculated based on Technical skills:0.45, Marketability:0.30,Entrepreneurial skills: 0.15, Funding:0.10 which is derived from the pairwise comparison of 28 medical industry clusters (Lee et.al, 2019).



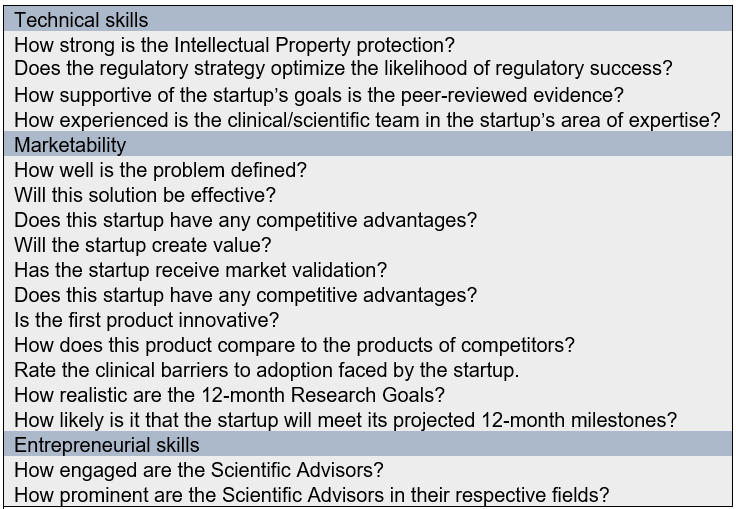
We modelled this calculation in R by referencing column names as well as categories, so if RedCrow has additional companies listed, there is little data preparation in consolidating score in the google spreadsheet, and then the model in R can directly calculate the scores.

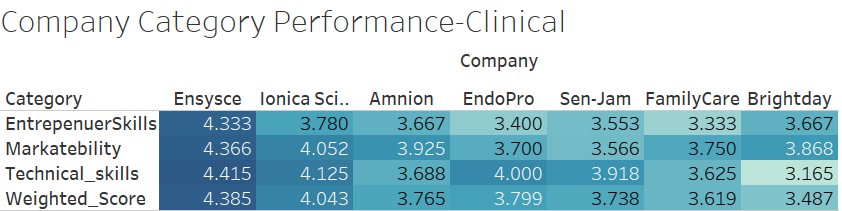


We have used R to aggregate the evaluator scores for business evaluation and then imported the dataframe into Tableau for easier visualization.I used a highlight table that enables us to see the final weighted score and the company categorical performance both vertically and horizontally. Dark blue indicates high rating, and as rating is lower (closer to 1), the color gets lighter. The insights that can be drawn are that Ensysce has the highest survival rate and scores the highest in two categories among its peers. FamilyCare has the lowest survival rate, and a factor behind this is its low score in the technical category which is weighted the highest. So, if the intellectual property status is improved-the rating for FamilyCare could increase a lot. A scope for improvement in this is adding more relevant questions in each category so that not one feature has too much weight on the overall score.

**Clinical Evaluation**

As for clinical evaluation, there were no features under Funding, so the weightage had to be changed, but I used the initial percentage to divide the remaining 10% in order to mimic the actual weight of the categories. Technical skills:0.495, Marketability:0.35,Entrepreneurial skills: 0.165, Funding:0.10.





Ensysce is again the company with the highest survival rate, and not only in total weighted score, but in every category as well. Brightday however slipped from the second position in business to last position, and mainly due to its low score in technical skills, which makes up for a huge percentile in the final score. This information is valuable for Brightday as it now knows to improve the clinical aspect more.

The end-users that benefit from this project are investors, RedCrow and startups. The investors get a better idea of likelihood of survival of a company that they are about to invest in, RedCrow could use this as a consulting aid to advise startups on which areas to focus depending on weakness categories and lastly the startups can enhance their strength and realize their weakness points for a better possibility of survival.

**Future scope**

A scope in the future could be conducting the analysis on healthcare startups founded after 2010 and cross checking whether the AHP weights coincide well in the United states market. Moreover, we could also conduct the analysis by subsetting the industry into health, biotech and medical industry since the customer base, required funding and barriers to entry are different for each, and will provide a more accurate customized model for each healthcare domain. Moreover, we can also use pairwise comparison within the subcategories so that we have sub category priority levels as well.

We can make use of feature selection through the scoring matrix to identify the most important features in the current feature set of redcrow’s evaluation process. We can also incorporate NLP in redcrow’s future evaluation process, and see how accurate the given comments are in identifying the strong and weak points of a startup. However, to bring the model into the actual existence Redcrow would require the combined effort of data analysis and the current evaluation team on board .

**Conclusion and recommendations**

For the entire project duration, our team tried to enhance the redcrow’s business model as much as possible. We analysed different sets of ideas and some worked, while some did not. Here, are some of the recommendations that we would like to put forward as a team:

1. By learning about the crowdfunding market for startups in many different areas, we can draw common characteristics about this platform. Although companies come from different financial backgrounds, the successful fundraising startups seem to have the similar financial features of controllable debt, and positive income. Furthermore, we will see strategies that companies often adopt to successfully raise capital on crowdfunding, and use that insight to give recommendations for startups of Redcrow on security information such as type of securities offered, number of opening offers, or even the method to determine security type.
2. Through the modeling part of data analytics, we can recommend the features which we are prioritizing when we are targeting its performance in the 12 month period. Some of these features include; ‘How prominent are the Scientific Advisors in their respective fields?’, ‘Does the regulatory strategy optimize the likelihood of regulatory success?’. ‘How engaged are the Scientific Advisors?’. The startup team members would be better able to manage their resources to the variables that are better impacting the performance, and also this would save a significant amount of time. We have also created a tool to monitor and prioritize the comments using sentiment analytics(NLP). Through this we would be able to monitor the sentiments of comments, we would be specifically targeting the comments with negative sentiment, by aligning resources on this side, we would be viewing the progress of the startup.
3. Using the AHP model, we recommend that startups focus on strengthening technical and marketability features to increase the likelihood of survival. Likewise, it is important to not overlook features under entrepreneurial skills and funding as these four components together cements a company’s foundation in order to battle the common startup survival challenge. Furthermore, we also recommend RedCrow to include more questions pertaining to current financial status of startups as well as entrepreneur networks to fill the gaps present in their evaluation process.

**References**

1. Data. (2016, May 1). Retrieved May 21, 2020, from<https://www.sec.gov/dera/data/crowdfunding-offerings-data-sets>
2. AngelList (2020). All companies. Retrieved from [https://angel.co/companies?markets[]=Healthcare](https://angel.co/companies?markets%5b%5d=Healthcare%5C)
3. Ahlers G.K.C., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in equity crowdfunding. Entrepreneurship Theory and Practice. Retrieved from <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2564121>
4. Lee,M., Park,S., Lee,K. (2019). What are the features of Successful Medical Device Start-Ups? Sustainability 2019, 11, 1948.
5. Bassani, G.,Marinelli, N. , Vismara S. (2019). Crowdfunding in healthcare. retrieved from <https://www.researchgate.net/publication/324666908_Crowdfunding_in_healthcare>
6. Indiegogo Datasets. (2020). Retrieved 17 June 2020, from <https://webrobots.io/indiegogo-dataset/>
7. Daley, S. (2019, July 4). Surgical Robots, New Medicines and Better Care: 32 Examples of AI inHealthcare. Retrieved from <https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare>