KICKSTARTER PROJECT

Abstract

We will analyze the project dataset and predict the success rate of these projects, and various inferences we make from tableau.

Tobias Sundheimer; Ashisha Konnur

BAN 4500

Contents

Abstract

Crowdfunding is an evolving phenomenon and currently characterized by high dynamics as this research shows. While it has its roots in the analogue world where Beethoven or Mozart have financed their concerts through public subscription, the global digitalization process allows this ecosystem of innovation and finance to flourish. Increasing numbers of projects of different natures are trying to acquire money from the internet public through Crowdfunding. At the same time, the number of Crowdfunding platforms is massively growing and has doubled between 2009-2018 (Figure 11), providing various options for fundraisers to take their ideas forward.

Keywords: Crowdfunding, Kickstarter, Visualization, KNIME Model

I. Introduction

Due to the short amount of time that crowd funding has been an effective capital raising strategy, there has not been an extensive amount of qualitative research done on this topic. On the discussion forum that was linked to this dataset, most of the other work that had been done on this was about cleaning the dataset itself, or descriptive background work. While this information is helpful for our data preparation, it does not give any insights into what the current consensus on this topic is. If anyone has been publishing more in-depth findings on this topic, they are published somewhere else.

Expanding our search beyond the discussion about the data source itself, we find a handful of more in-depth findings. There are many different articles that have been written about it, but only a handful come from more reputable sources. One study from Bentley University [1] was conducted and found that goal amount, description, and some other selected statistics influenced the successfulness but does not provide more in-depth analysis. A study conducted in 2017 by a Stanford alum found metrics that would measure the effective marketing strategy (such as title length, number of images on campaign) as well as characteristics about the campaign organizer (such as number of campaigns worked in the past)[2]. This information lines up with the main study that has been conducted on this topic, which was a study that was performed in 2014 by Ethan Mollick. This study found that things like the perception of the campaign runner and the product were important, instead of the product itself being the main driving force of success.[3] What this study fails to discuss is the characteristics of the people who make up the campaign. There are significant differences in items that have a large group of people donating a small amount of

[1] Bentley University. “Shh … We Scraped Kickstarter’s Data.” *Medium*, 29 May 2015, medium.com/@bentleyu/shh-we-scraped-kickstarter-s-data-to-find-out-the-secrets-to-engineering-the-perfect-campaign-3b7b537b6c34.

[2] Benavides, Nicholas. “What Makes a Successful Kickstarter Campaign?” *Medium*, 23 May 2018, towardsdatascience.com/what-makes-a-successful-kickstarter-campaign-ad36fb3eaf69.

[3] Mollick, Ethan. "The dynamics of crowdfunding: An exploratory study." The Journal of Business Venturing (2014): 1-16.

money, compared to a campaign that is being funded by a small group of very wealthy people. Other groups such as Bentley University have done some work on this topic, but with only descriptive statistics having been performed. While we will aim to recreate some of these, this will not be the true focus of our project. There is clearly some information on this topic, but there is much more nuanced information that we may be able to find in this dataset.

Another angle that we want to take this is in the differentiation between barely successful and extremely successful campaigns. There are some campaigns that may be successful but were only just barely able to make the goal amount. At the other extreme, campaigns may make significantly more than they originally set out to make, sometimes multiple times over. In addition to this, there are also examples of campaigns that were able to hit their goal extremely fast, whereas others drag on for the entire duration of their campaign. In our baseline dataset, all these different scenarios are classified as equally successful. We want to see if there is anything systemically different between the types of campaigns that are "extremely" successful and those that are not. As a company, Kickstarter should be aiming for not only the most successes possible, but the highest number of extreme successes.

Overall, our group is trying to build a more nuanced predictive model on what crowdfunding campaigns may succeed, and to what degree they will succeed. Since Kickstarter only makes money once a project has passed its goal – and has been deemed a success – it is already in their best interests to only have successful campaigns. But if there were a way to determine which campaigns would be extremely successful – and could theoretically be even more successful with the right publicity on their end – Kickstarter would find this information extremely valuable. We could also provide a service to a prospective user of a crowdfunding campaign, on ways that they can optimize their campaign itself to give themselves the best chance of being successful.

II. Research Questions

- Do certain socio-economic factors play a role in determining whether a crowdfunding campaign will be successful?
- Does the country this campaign is in help determine anything?
- Time of year have an effect.

Successful vs super successful

- Any significant predictors for super successful? Will need to create a variable for this (ratio
- Given that the company managed to raise pledges than goal amount, why do they still fail (search for examples and find out more through research)

III. Method and Analysis

Tableau

We used tableau to diagnostic and descriptive analytics.

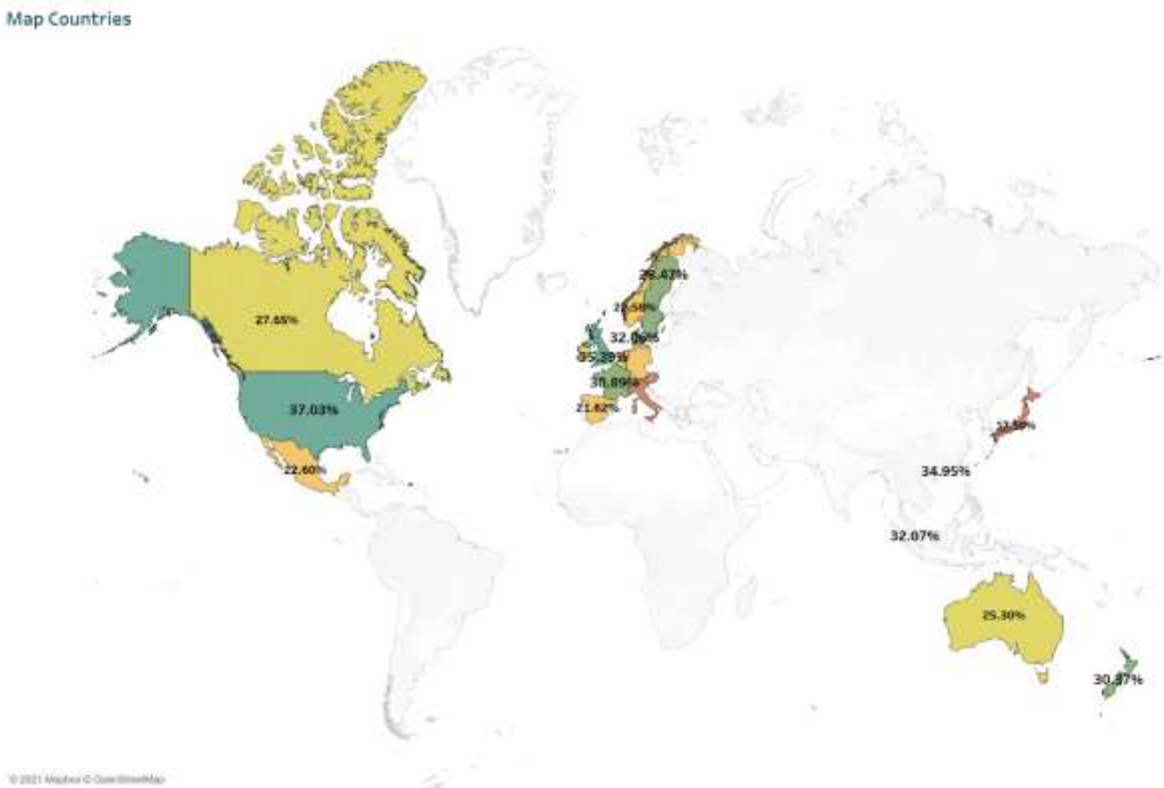

Figure 1-Map Region showing the Success rate and pledge amount per Country.

We wanted to first see how the data is spread out across the continent and relate it with the success rate and pledge amount. And from Figure 1-Map Region showing the Success rate and pledge amount per Country.Figure 1 we notice that America region receives over 3 billion pledges which is the most for any region and we will focus our project in this area.

It is not surprising to note that most of the Kickstarter projects come from USA since Kickstarter is from US. But we also noticed that outside of America, the projects had much higher success rate and required higher amount of funding which we think is to attract attention of the "Venture Sharks".

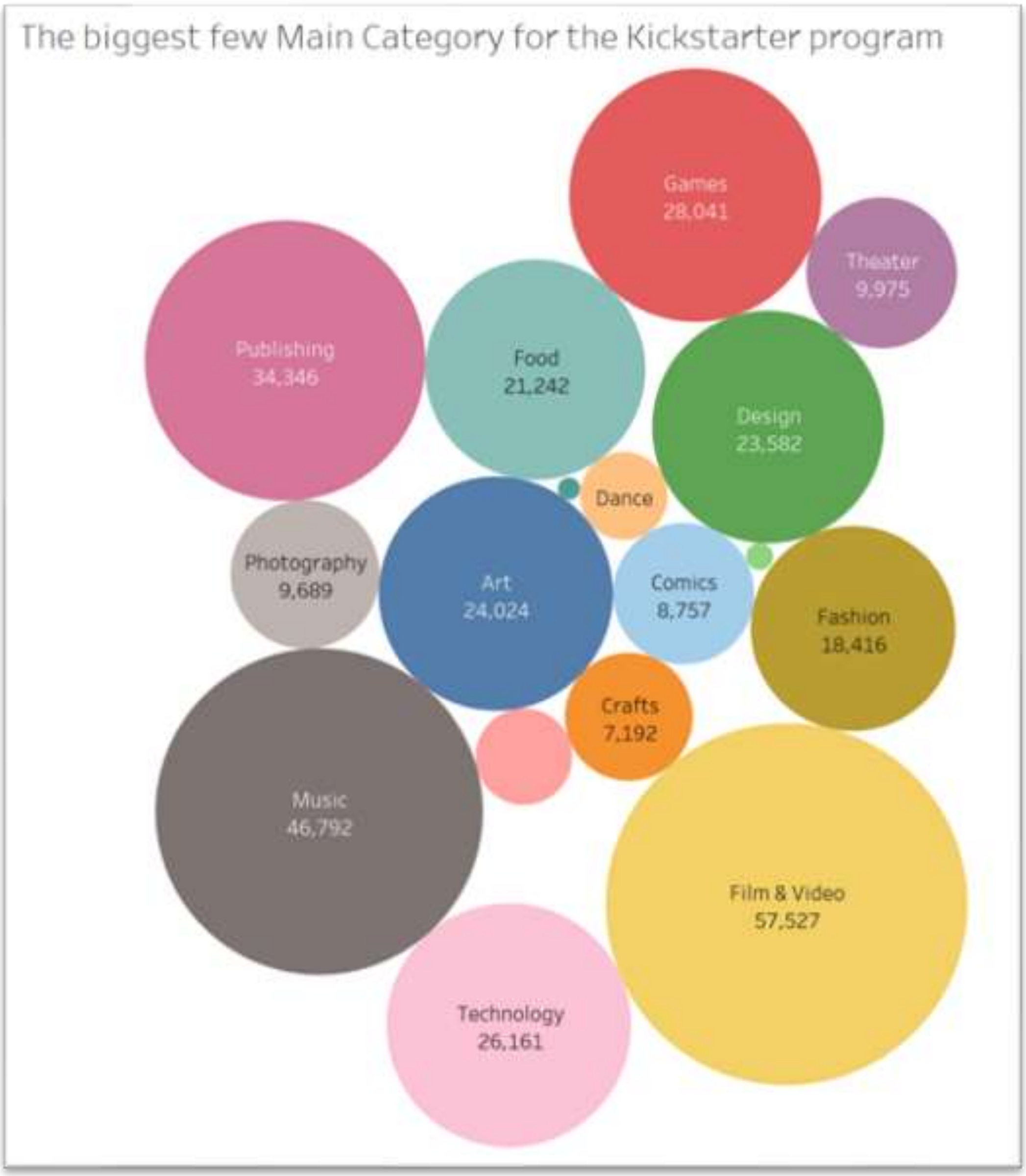

Figure 2 - The top Main Category of our data.

We can infer from Figure 2 that most of the projects that were launched on the Kickstarter were of Film& Video, Music, Publishing, Games and more. But when we analyzed how do these category fair with their project success from Figure 3 and find that Dance and Theatre categories which are not very popular in the Kickstarter program, tend to have higher success rate that their peers. We think that media category has better chance of having higher success and it is a niche area.

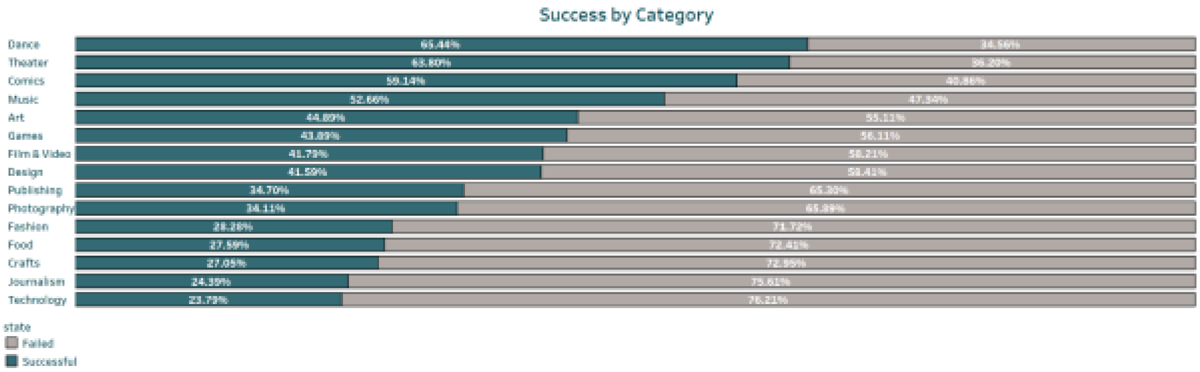

Figure 3-Success rate for each category

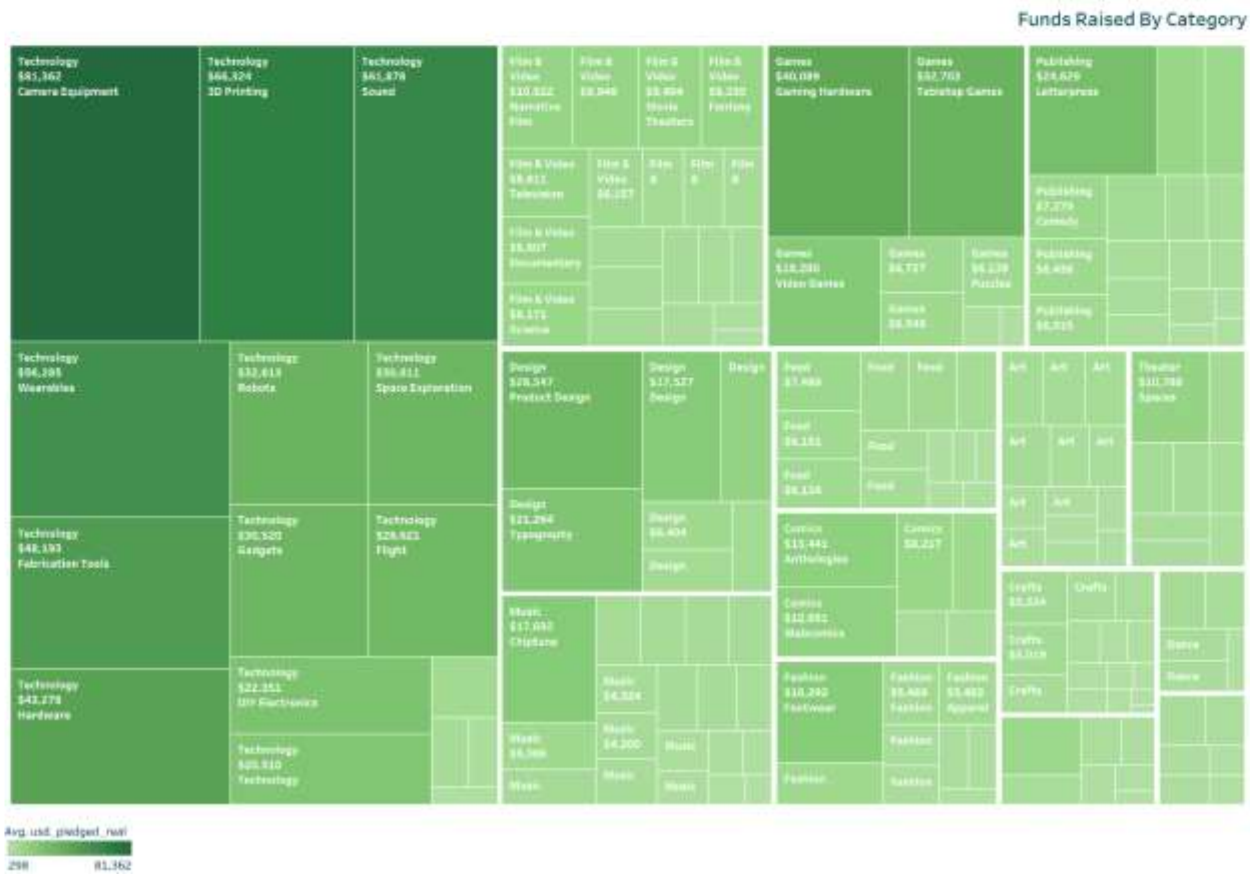

Figure 4-Avg. Pledges received by Main Category.

So, Technology receives the most funding on average for the Kickstarter program as seen in Figure 4. This may not come as a surprise to people who are familiar with the some of the projects that came to mainstream from these pet projects. Among the top technology projects are Camera Equipment, 3D printing and Sound.

From our dataset, what percentage of projects are successful or failure?

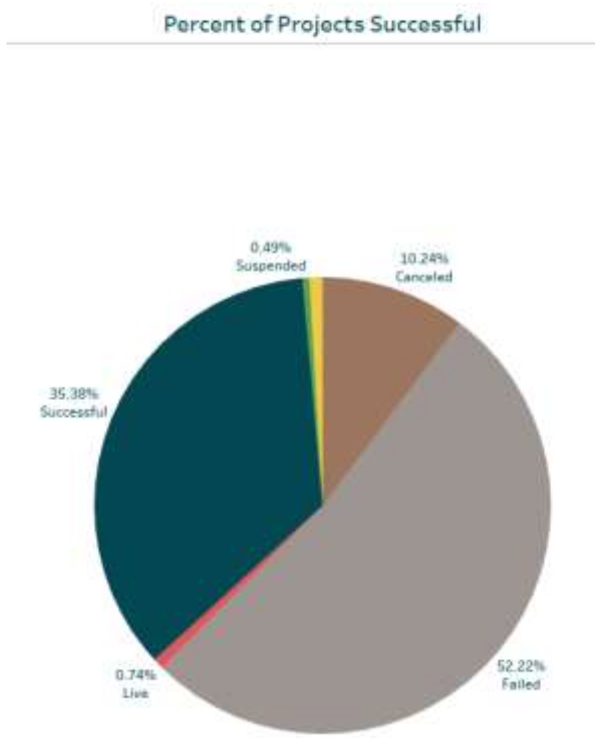

Figure 5-Demograph of projects success and failure

50% of projects fail (Figure 5). And those can be a disappointing number for someone whose dream is already dashed when they want to become entrepreneur. And 30% of all the projects in this data are successful.

From Figure 6 We also explore which of the projects in the entire dataset and total pledge amount received. We wanted to see how much these projects receive from their target goal amount. And certain projects like the Pebble Time received raved results and much more than their goal amount. We also reached to a conclusion where we charted **success rate** (Number of successful projects / Number of projects for each goal bucket) over project **goal** and found a trend whereby smaller project goal had higher success rates. Larger projects (>$100,000) had a success rate of 18%.

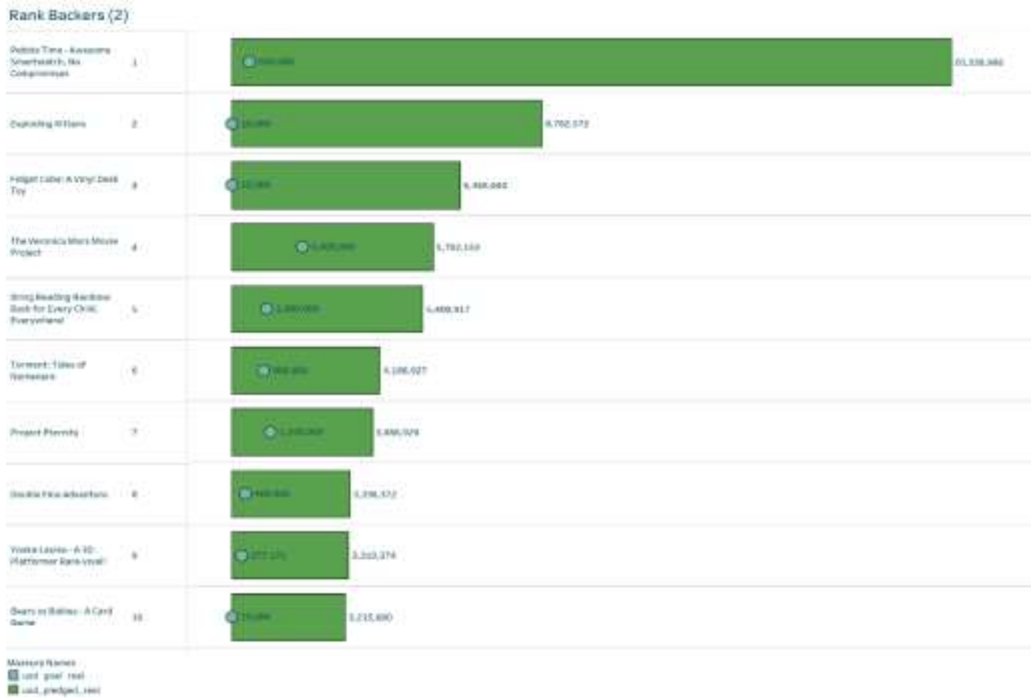

Figure 6-Backers who contribute to the pledges and projects.

From Figure 7 we need to estimate how many days on average to the projects from category take and the average goal amount. And for sure, Technology projects took the longest time to reach their goal and Crafts took the least time.

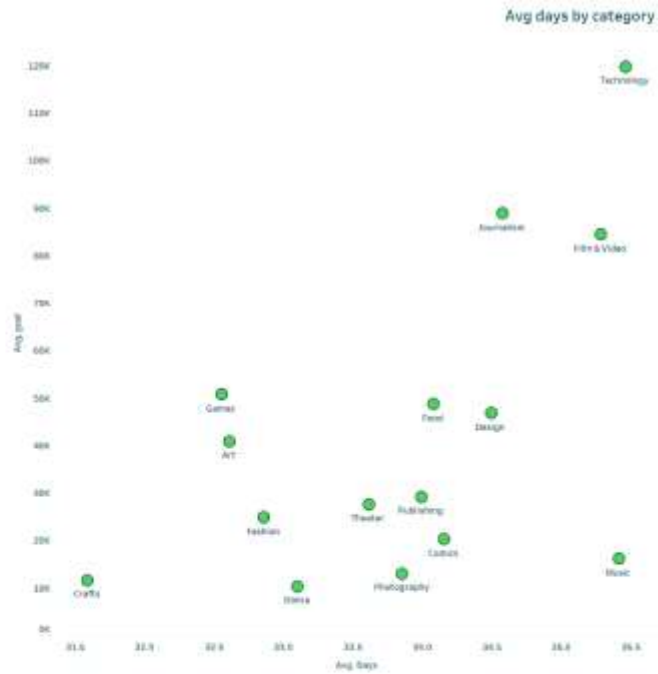

Figure 7-Average Days to success by category

KNIME

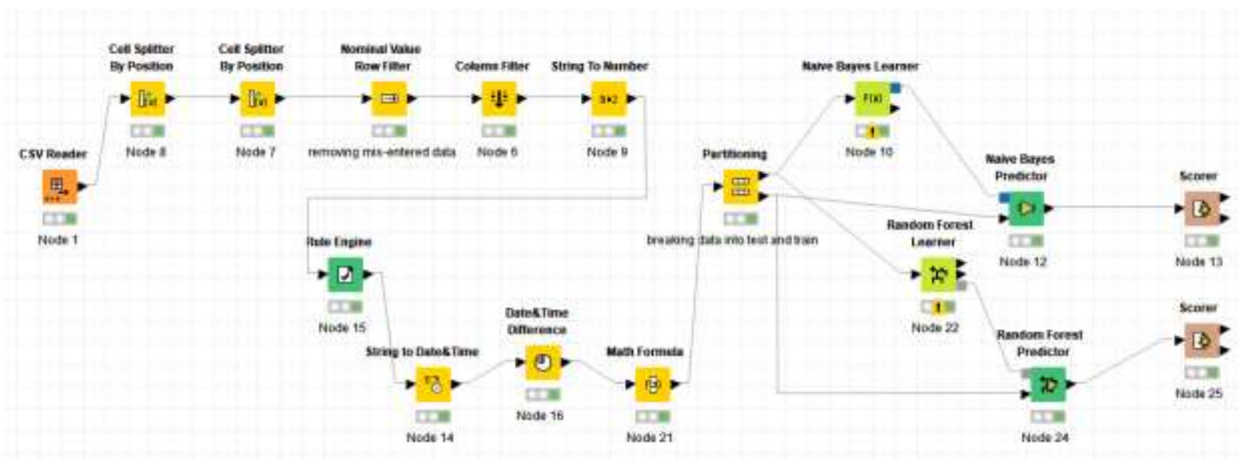

Figure 8-KNIME modelling

We started by splitting the launched date and deadline date and finding the difference between those dates will give us the time taken for different projects that launched seek for pledges to achieve their goal. With the 'partitioning' node, we separate the data in 70% and 30% train-test data for the predictive analytics.

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Cohen'...
failed	42534	35698	10882	7822	0.845	0.544	0.845	0.234	0.662	?	?
canceled	73	613	86685	9565	0.008	0.106	0.008	0.993	0.014	?	?
successful	5057	514	62341	29024	0.148	0.908	0.148	0.992	0.255	?	?
live	36	100	95518	1282	0.027	0.265	0.027	0.999	0.05	?	?
undefined	1092	11219	84625	0	1	0.089	1	0.883	0.163	?	?
suspended	0	0	96485	451	0	?	0	1	?	?	?
Overall	?	?	?	?	?	?	?	?	?	0.503	0.111

Figure 9-Naive Bayes model score

Since we are applying machine learning prediction methods to a classification problem that has more than 2 different classes, we cannot calculate over all precision, Recall, or other performance metrics, only accuracy and the Cohen's Kappa value. When we use the Naïve Bayes classifier, the model we get has an accuracy of 0.503 and a Cohen's Kappa value of 0.111. since the current baseline accuracy of randomly guessing the right outcome of a given campaign is 1/6, being right half of the time is already a noticeable improvement. However, having a Cohen's Kappa value of 0.111 implies that the model may still be right by random chance, and that we should try other models as well.

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Cohen'...
failed	50037	10715	35865	319	0.994	0.824	0.994	0.77	0.901	?	?
canceled	32	99	87199	9606	0.003	0.244	0.003	0.999	0.007	?	?
successful	34036	733	62122	45	0.999	0.979	0.999	0.988	0.989	?	?
live	80	40	95578	1238	0.061	0.667	0.061	1	0.111	?	?
undefined	1092	71	95773	0	1	0.939	1	0.999	0.969	?	?
suspended	0	1	96484	451	0	0	0	1	NaN	?	?
Overall	?	?	?	?	?	?	?	?	?	0.88	0.781

Figure 10-Random forest model

From the Random Forest predictor model, we get a much more impactful model. While implementing a Random Forest predictor is much more computationally intensive, the results speak for themselves. With an accuracy of 0.88, we are now right almost 90% of the time, which is a vast increase from the original 1/6 that purely guessing would give us, and almost double what the Naïve Bayes classifier gives us. Furthermore, having a Cohen's Kappa value of 0.781 implies that our model is robust and producing accurate results, instead of stumbling into a correct response via luck, like what was implied in our Naïve Bayes model.

IV. Conclusion

The success rate of the projects depends inversely on the project goal. The bigger a project, the less likely will it get funded (see Figure 7)- for larger projects it is usually the Film & Media, Technology that need enough momentum to raise the money and become successful. The failed rate of projects increased sharply early 2011 peaked the failure rate in 2015 and then drop sharply in 2016 ~ 2017 (see Figure 11) to be just above the success rate. This could be due to popularity and influx of projects.

The strongest relationship was probably between the category of a project and the success rate. Design projects, for example, are incredibly successful. My guess is, that they are rather small and hit a certain established fan base. Journalism projects on the other are not very successful, with a success ratio of 2.74. My guess is, that it is more difficult to establish a fan base and publishing magazine to people who know what they want to see.

We have established that project size does plays a role for the success rate. However, it’s influence is also different for different categories. The success rate for projects from some categories (e.g., Music) drops faster for larger goals than from other categories (e.g., Technology) (graph not shown here).

With KNIME, we were able to achieve a 90% accuracy which is very good to future project users to strive for, and usefulness of our model in their Kickstarter projects.

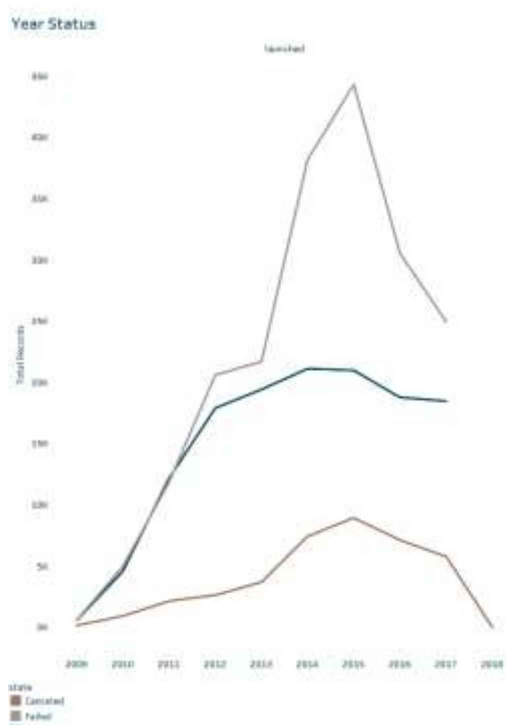

Figure 11-Year Status

References

Mollick, Ethan. "The dynamics of crowdfunding: An exploratory study." *The Journal of Business Venturing* (2014): 1-16.