W200 Python Fundamentals, Project 2

Section 2, Team 3: Sergio Ferro, Andres Kodaka, Marcelo Scatolin Queiroz

Introduction

Crowdfunding has become a viable way for entrepreneurs, artists, inventors, or developers to fund a project or idea for which conventional financing has traditionally been challenging.

Kickstarter is a successful online crowdfunding platform that lets creators submit their projects and have individuals pledge money to the ideas they would like to see come to fruition. The stated mission of the company is to help bring creative projects to life, focusing mainly on artistic endeavors over the years[1]. Since 2009, 14 million people have pledged $3.6 billion on the platform. The campaigns showcased on the site range from art and film-making to software and technology.

The format for getting a campaign funded on the site is simple. The owners or teams set up a campaign 'profile' to promote their project, they set their intended funding goal and funding target date (up to 60 days), and the project only gets funded if the stated pledge goal is met in the set amount of time. Pledges by creators are forbidden by the platform and are a valid reason for canceling a campaign[2].

Data and Project

Our data is sourced from a company named 'Web Robots' (https://webrobots.io). They created a bot that crawls the Kickstarter website for all active, ended, and cancelled campaigns that are currently searchable on the site. Once a month, the bot crawls the site and pulls relevant data on these campaigns such as campaign title, pledge goal, total pledges, funders, etc. In order to foster transparency, a campaign can never be removed from their database. However, campaigns may be removed without further notice in case intellectual property disputes are detected, the content is considered offensive, or violates other Kickstarter terms of use.

We'd like start this analysis making an effort to understand the platform. We want to explore the website's diversity in terms of countries, categories and subcategories.

Once we've started to understand the size and scope of Kickstarter as a platform, we will go into further detail about the differences between the categories on the site. We'd like to assess whether the company's purpose to give funding to artists has been accomplished. We'd like to know what features separate successful projects from unsuccessful ones. Even though most people may not have the ability to change the projects they choose to bring to the platform, an analysis of success and failure may help with expectations or find ways in which the project could be tweaked to reach a goal. Knowing that more expensive projects are least likely to get funding, creators could focus on efficiencies on their cost structure. Also, knowing that a good project is likely to get overfunded may give some margin to creators to adjust their goal properly.

---

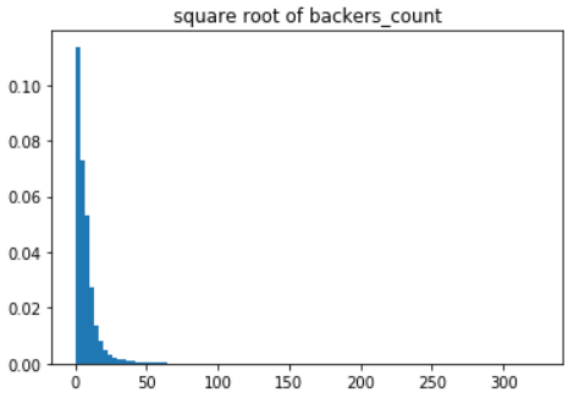
[1] https://www.kickstarter.com/help/faq/kickstarter+basics#faq_41846

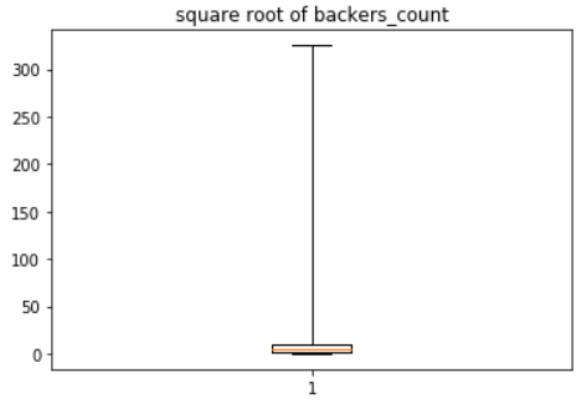
[2] https://www.kickstarter.com/help/faq/creator+questions#faq_41821

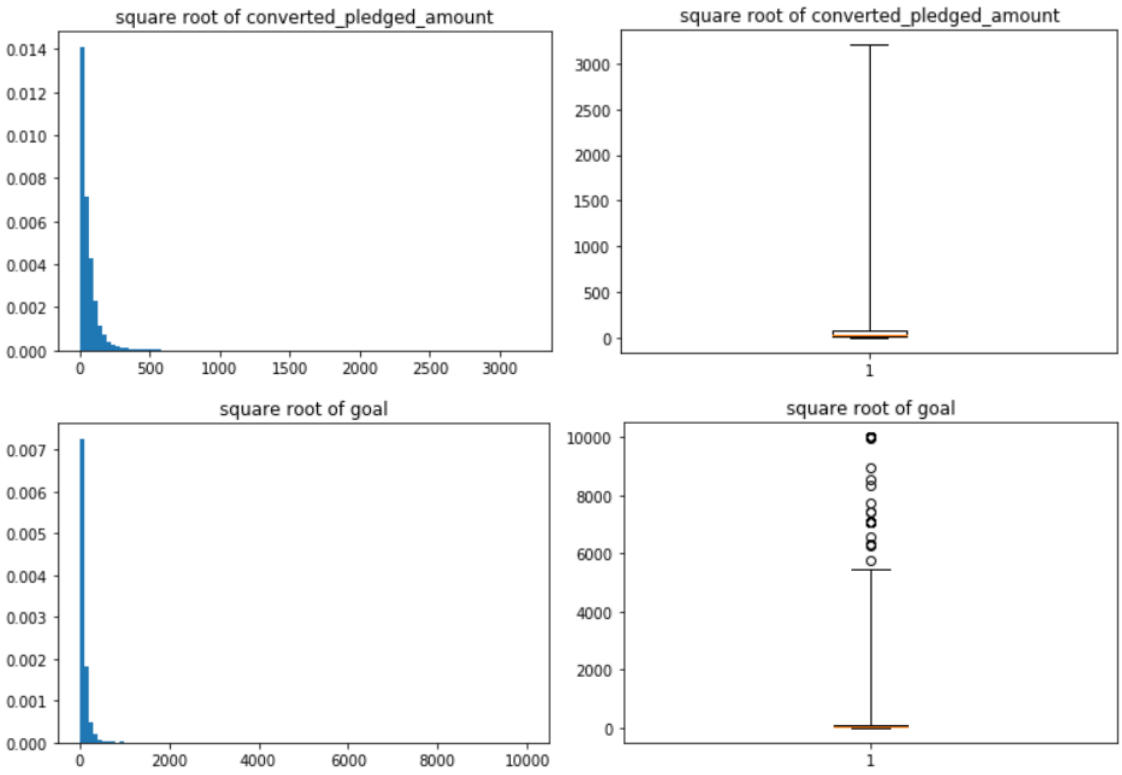
Thus, differences in the number of backers or the the amount pledged for successful campaigns might say something about the nature of backers.
All in all, we want to know whether the motivations of each of the stakeholders are being fulfilled. Has Kickstarter become a viable platform for art (i.e. their mission statement)? Are backers finding campaigns they like enough to support? Are creators able to improve their situation against traditional funding? How does a successful campaign look like?
Finally, we will find that some categories are inherently more expensive to back than others. We try to gauge whether this affects their success rate and whether backers are more discouraged by the attainability of a campaign's goal.

Data sanity checks

The dataset for February 2018 has 199,704 observations. The fields that the crawler obtains from Kickstarter amount to 38. A look at the data using the describe command shows that numeric variables (backers, USD pledged amount, goal, fx rates, and pledges in their original currencies) are complete and have no missing values. However, many of them show maximum values that are many orders of magnitude beyond the third quartile. For example, there was a goal of USD 100 million despite the fact that the median was only USD 5 thousand. A histogram of these raw series showed a cluster near zero and then no goals until the millions of dollars. To help with the visualization we took the square root of the series.







The distributions are what one would expect. Positively skewed, with most amounts well below the means and medians, and a few cases to the right of the distribution. The goal charts caught our attention because of the number of outliers, and because the magnitude is so much higher than the maximum of the pledged amounts. It doesn't make too much sense to ask for a goal so many times larger than the maximum historical amount pledged. In fact, there are 84 projects whose goal is higher than the maximum historical amount pledged. A quick check on the data showed that even though pledges had been turned to USD, goal did not. Thus, we created a new series using the currency rates on the database. After this correction, the number of projects that surpass the highest historical amount pledged drops to 79.

There are 13 projects that ask for USD 100 million and beyond. Most of them were publicity stunts or jokes, such as the "Multi-Trillion Dollar Dream at Las Vegas, NV" which will "employ millions of people and be home to millions more" or the "Let's prove the earth is flat" project which offered backers a first edition of the bible "signed by YHWH himself"[3].

Since all of the projects that surpassed the highest historical amount pledged failed, and seem to be jokes or publicity, we'll be dropping them from the database. Elsewhere, if

[3] https://www.kickstarter.com/projects/1882252849/lets-prove-the-earth-is-flat?ref=category_newest

we use an outlier filter it is only with the purpose of visualizing the core characteristics of the series, otherwise the orders of magnitude wouldn't allow for properly interpreting visualizations.

Below we have the most relevant columns that we use in our analysis:

Name	**Description**
backers_count	# of people who backed the campaign (integer)
blurb	String description of campaign (text)
category	Main and sub category campaign belongs to (text)
country	Country of origin of campaign (text)
created_at	Date/time of campaign creation (unix)
currency	Currency of campaign pledges (text)
deadline	Date/time campaign ends (unix)
fx_rate goal	Pledge goal of campaign, adjusted for FX rate (integer)
goal	Pledge goal of campaign in campaign currency (integer)
id	Unique identifier (integer)
launched_at	Date/time campaign launched (unix)
name	Name of campaign (text)
pledged	Total pledged (integer)
spotlight	Whether campaign was spotlighted by kickstarter (boolean)
staff_pick	Whether campaign was tagged as staff pick (boolean)
state	State of campaign ('live', 'successful', 'failed', 'canceled')
state_changed_at	Date/time state updated (unix)
static_usd_rate	FX rate (integer)
usd_pledged	Conversion in US dollars of the pledged column (integer)

All of the dates in the dataset are in unix format. In order for us to be able to use and manipulate them, we need to convert them to a usable date format within pandas.

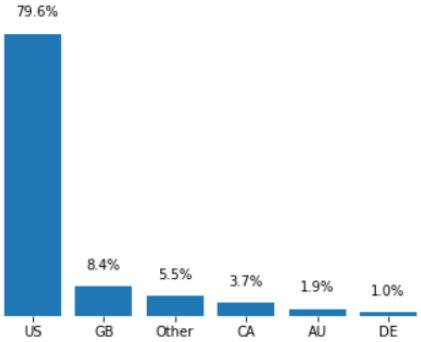
Additionally, in order to extract the top and sub categories from the ‘category’ field, we must substring them from the url that appears in the field. For this, we used the string extract function within the regular expressions library.

We also dropped duplicates from the database using the df.drop_dulicates() command, which erased 4,089 entries from the database.

Finally, an analysis of the categorical values shows missing values in blurb, which contains short descriptions of the projects. However, the column ‘name’ is just missing one value and thus, this information can be recovered from this column if needed. Sometimes, the location inside a country isn’t specified which explains why there are 1415 null entries in the location column. There are 168 starred projects with non null values, but when we checked them they were all False, which doesn’t help the analysis.

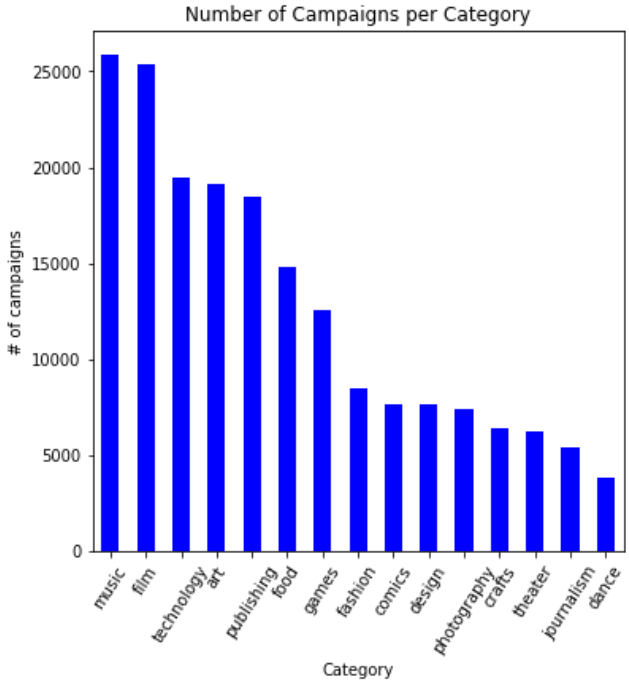
```
Data columns (total 38 columns):
backers_count               195536 non-null int64
blurb                       195526 non-null object
category                    195536 non-null object
converted_pledged_amount    195536 non-null int64
country                     195536 non-null object
created_at                  195536 non-null datetime64[ns]
creator                     195536 non-null object
currency                    195536 non-null object
currency_symbol             195536 non-null object
currency_trailing_code      195536 non-null bool
current_currency            195536 non-null object
deadline                    195536 non-null datetime64[ns]
disable_communication       195536 non-null bool
fx_rate                     195536 non-null float64
goal                        195536 non-null float64
id                          195536 non-null int64
is_starrable                195536 non-null bool
launched_at                 195536 non-null datetime64[ns]
name                        195535 non-null object
photo                       195536 non-null object
pledged                     195536 non-null float64
profile                     195536 non-null object
slug                        195536 non-null object
source_url                  195536 non-null object
spotlight                   195536 non-null bool
staff_pick                  195536 non-null bool
state                       195536 non-null object
state_changed_at            195536 non-null datetime64[ns]
static_usd_rate             195536 non-null float64
urls                        195536 non-null object
usd_pledged                 195536 non-null float64
usd_type                    195536 non-null object
location                    194186 non-null object
friends                     168 non-null object
is_backing                  168 non-null object
is_starred                  168 non-null object
permissions                 168 non-null object
usd_goal                    195536 non-null float64
```

How big and diverse is Kickstarter?



There are 22 participant countries, but 93.5% of them are concentrated in the US, Great Britain, Canada, Australia and Germany. That is, Kickstarter has mainly become a tool for artists in developed countries. Arguably, more could be done to increase diversification and bring funding to less developed economies, which may need more support for the arts.

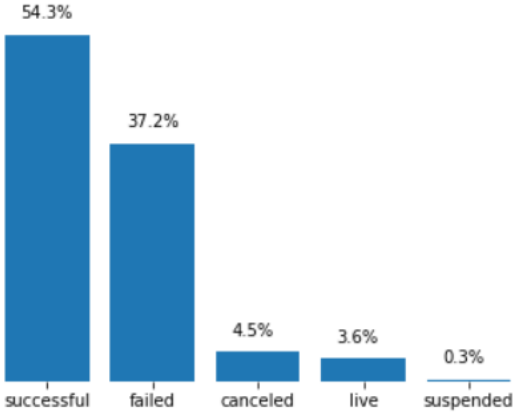
The site organizes its campaigns by categories. There are 15 ‘main’ or ‘top’ categories and 144 ‘subcategories’. Not all observations are organized within ‘subcategories’. For example, a campaign may fall under the ‘Art’ top-category and the ‘Mixed Media’ subcategory, while other ‘Art’ campaigns may have an ‘undefined’ subcategory.



Given the company’s artistic/creative mission, it is unsurprising that 75% of crowdfunded projects are art related. That is, Kickstarter as a stakeholder can be at

ease with the fact that they are fulfilling their mission statement, even though they may still consider expanding their reach towards developing markets.

Now we look into the proportions of successful and failed campaigns. As mentioned in the introduction, projects are only considered ‘successful’ is they end up getting fully funded at their agreed upon goal by the set amount of time. A project in our dataset could also be ‘failed’ (did not reach intended finding goal in time), ‘canceled’ (e.g. by the creator), ‘live’ (currently in its funding period), or ‘suspended’ (possibly for breaking Kickstarter’s policies, including adding creator money to the campaign).

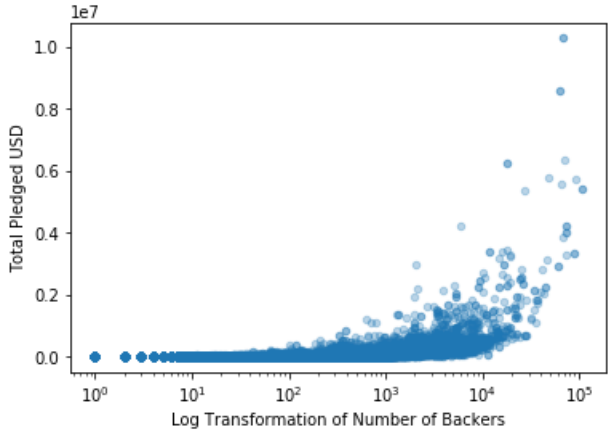


54.3% of the projects in the dataset were successful, while 41.7% failed or were canceled.

Next, we describe the comparison between the number of backers and the total amount pledged for *successful* campaigns. From the description below, we find that the range of the ‘Total Pledged’ variable is very wide, from a minimum of $1 to a max of more than 10 million.

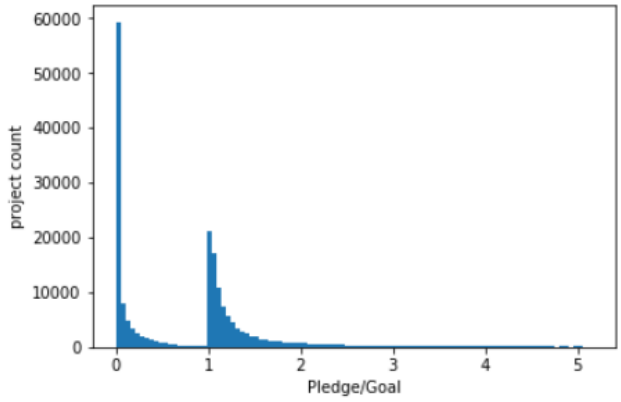
Statistical summary of pledged amount	
Mean	20,586.95
Standard Dev.	123,772.02
Minimum	1.00
1st Quartile	1,845.00
2nd Quartile	4,714.00
3rd Quartile	11,781.00
Max	10,266,845.00

A scatterplot of the amount pledged and the number of backers shows that the vast majority of total pledge amounts lie in the thousands of dollars and that most campaigns have under 10,000 backers.



The logarithmic scale also shows that the range of pledges increases with the number of backers. That is, there may be a positive relationship between goal and the number of backers, but the average pledge does not grow linearly. For some big projects, the average pledge might be higher, an analysis we drill down further in a following section of this document.

This concentration suggested some relationship between pledges and goals, with the hypotheses that once the goal was reached there were no more incentives to back the campaign. A closer look at the data suggests otherwise:



That is, most projects get exactly their goal or none at all, but once the goal is reached many projects keep receiving contributions.
This can be better visualized with the statistical summary of the proportion of pledges to goal in successful projects:

Statistical summary of pledged/goal rate	
Mean	3.65
Standard Dev.	259.98
Minimum	1.00
1st Quartile	1.05
2nd Quartile	1.16
3rd Quartile	1.58
Max	41,535.01

If backers were discouraged to contribute after the goal is reached, we should see most values concentrated around 1. However, backers tend to give a lot more to successful projects than what these ask, suggesting that most backers act like art patrons rather than venture capitalists. In other words, there are motivations for them to support projects they like beyond the point where they're viable.

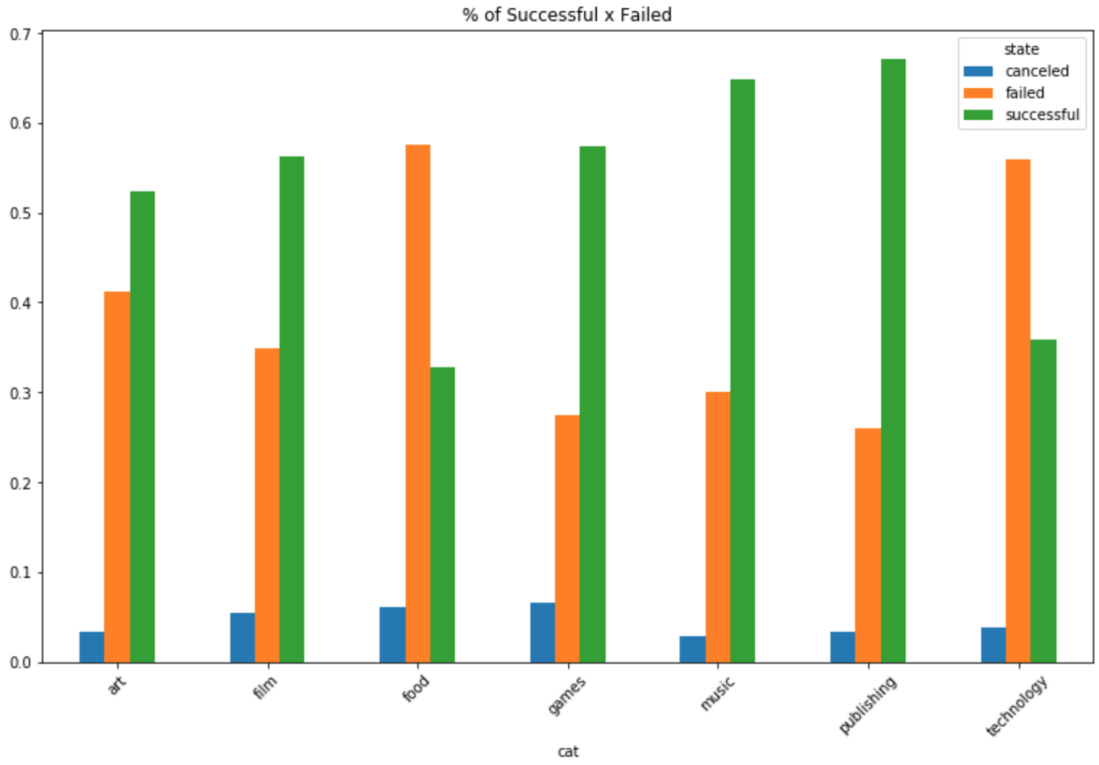
Analysis by state of the project

We analyzed successful vs failed projects, trying to understand if one category tends to gather more support among backers:

Top 5 categories in 'Successful'		
Category	**Count**	**% of Succ.**
Music	17112	16
Film	14734	14
Art	13088	12
Publishing	12828	12
Games	7850	7

Top 5 categories in 'Failed'		
Category	**Count**	**% of Categ.**
Technology	11344	15
Film	9127	12
Art	8838	12
Food	8781	12
Music	7923	11

Since music, film and art categories are among the larger categories by project count, it was expected that they might show a balance between fails and successes, but food and technology projects seems to be particularly underperforming. To analyze this idea deeper, we verified each of those categories success rates:



The chart shows relatively high rates of success for art related subjects such as art, film, music and publishing. Once again, this contributes to the narrative that the mission statement of Kickstarter is being fulfilled. This new funding source for traditionally ignored subjects such as creative projects spawned many new projects and many of them ended up being 'successful' by Kickstarter's definition. We can then presume that these creative projects were failing to get financed not because society didn't want them, but because of the risk averse nature of traditional means of funding.

Food and technology are two categories where failures occur more often than successes. To dig deeper into why, we decided to explore the technology category.

Analysis of Technology projects

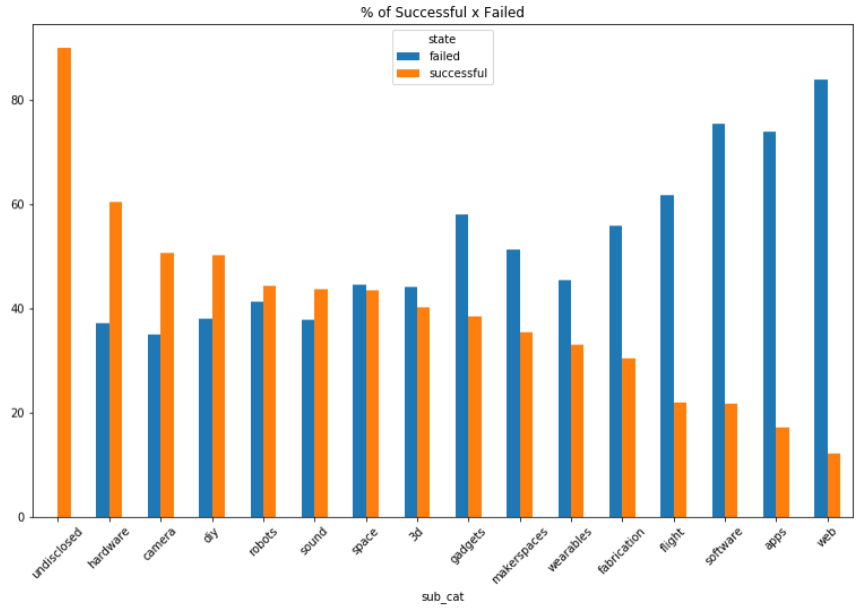
An interesting finding is that even though they make nearly 10% of Kickstarter projects, technology projects accrue most of the failures. By subcategory, technology doesn't seem to show a particular area that is struggling, with the proportion of failures mostly

even for the top five with a little preference for hardware and gadgets in the successful subcategories:

Top 5 Tech. subcategories in 'Successful'		
Subcategory	**Count**	**% of Succ.**
Hardware	1845	25
Gadgets	969	13
Software	571	8
Undisclosed	566	8
DIY	519	7

Top 5 Tech. subcategories in 'Failed'		
Subcategory	**Count**	**% of categ.**
Web	2089	18
Software	1979	17
Apps	1899	17
Gadgets	1459	13
Hardware	1135	10

The reason why hardware and gadgets might perform better is that they have an additional quality filter[4]. Creators are expected to submit pictures of a physical prototype, which signals the ability of the creator to get the campaign towards completion. The chart below follows rates of success and failure in the subcategory (as opposed to the tables, that looked at technology as a whole):



We can see that most successful projects are constrained by the rule that they must submit a prototype, such as hardware, cameras, DIY, robots and gadgets. We believe this creates an additional quality check and boosts confidence in backers, which suggests that a similar approach to software could yield better results. Thus, if you're a software developer, it would be good to provide a sample of what your program or

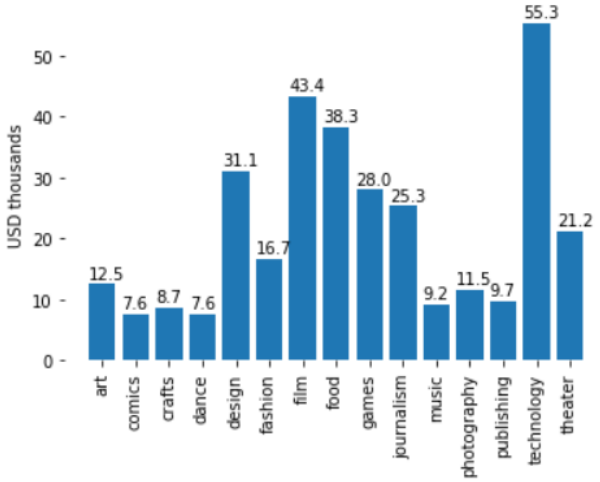
[4] https://www.kickstarter.com/rules/prototypes?ref=rules

website may do, links to your portfolio of projects, and other quality checks to increase the confidence of backers.

Finally, ‘undisclosed’ is the category with the most success overall. We ran a check on the projects inside this category and found that these were mostly misplaced. They were not catalogued inside other categories since they were related to sports gadgets or biology gadgets (for example, a tracker for reptiles in the city of Perth lead by a PhD student). Once again, the tangible prototype proxies as a quality check and boosts rates of success and confidence among backers.

High goals vs successful pledges

After some discussion, it was suggested that technology projects were likely to be the most expensive to make and may have a harder time reaching their goals. This section explores that concern.



The above bar chart shows that, of the campaigns that were no longer ‘live’, technology has the highest average campaign goal. This is also evident when comparing the statistical summary of technology projects and all other combined:

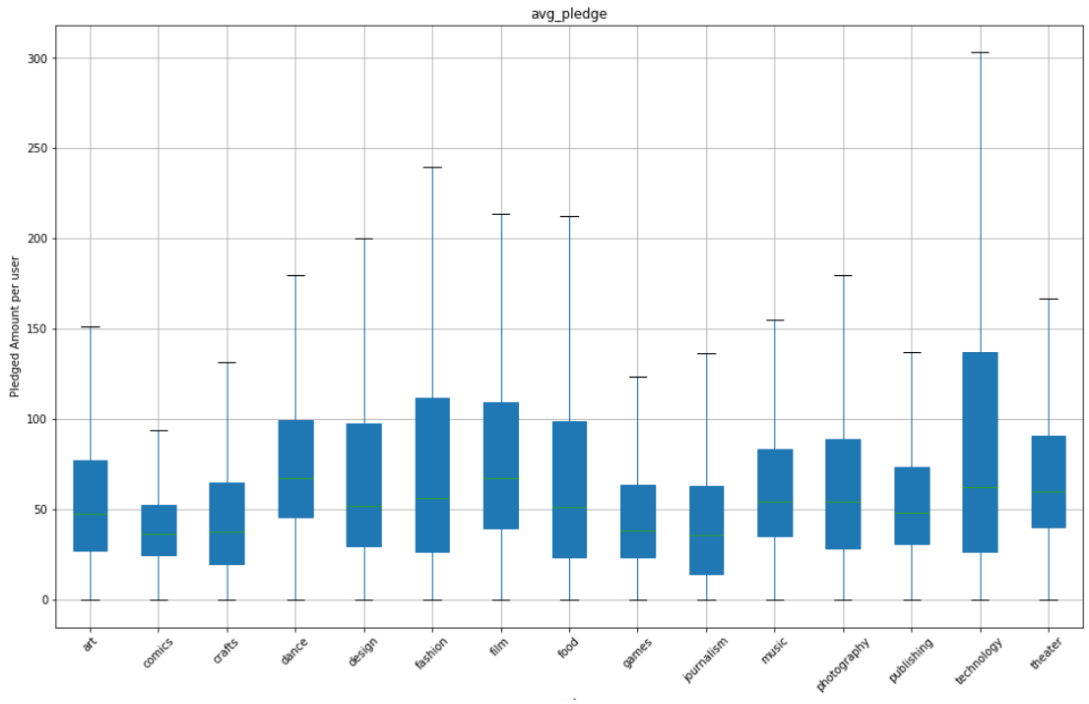
Statistic summary - Project Goals Technology	
Mean	69,479.28
Standard Dev.	827,241.18
Minimum	1.00

1st Quartile	5,000.00
2nd Quartile	15,000.00
3rd Quartile	50,000.00
Maximum	100,000,000.00

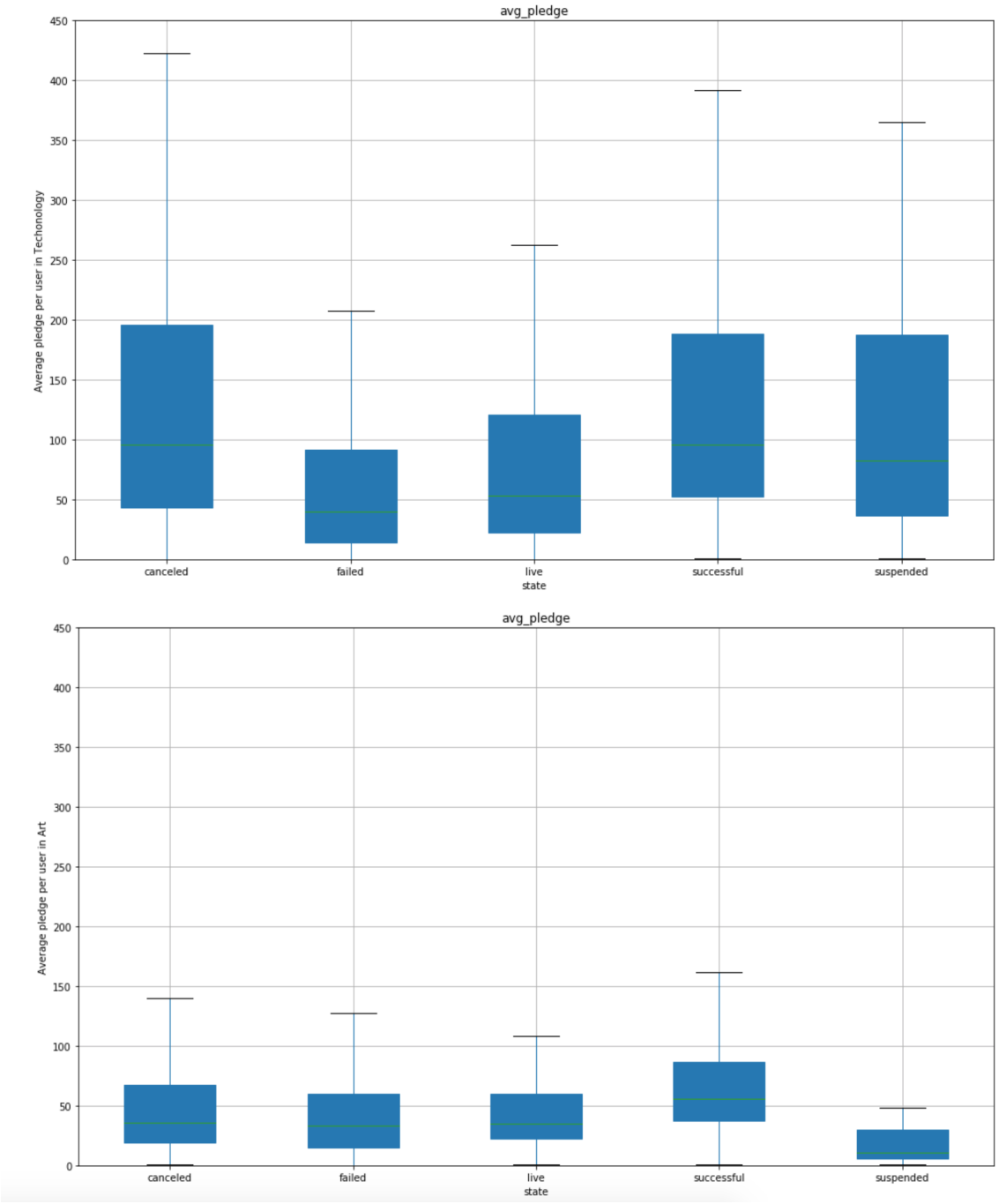
Mean	23,249.48
Standard Dev.	235,129.89
Minimum	0.01
1st Quartile	1,500.00
2nd Quartile	4,500.00
3rd Quartile	10,000.00
Maximum	42,900,000.00

Statistic summary - Project Goals
All others aggregated

We can see the technology projects have goals, on average, almost 3 times the rest of campaigns. The distribution is more skewed to higher amounts as noted when comparing the quartiles. With such higher goals, it is more likely to have higher pledges per backer. We performed an analysis checking the average pledge per backer grouped by category in a box plot chart:



As we can see, not only does the average pledge per backer in technology have one of the highest medians, it also has the most disperse pledge amounts. When comparing data from technology projects and a more successful category, like art, we see that, indeed, projects with higher goals may be less successful despite having higher average pledges.



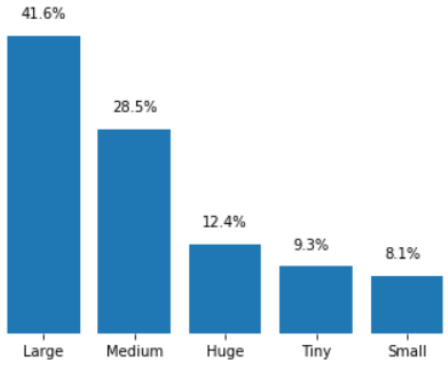
The last boxplots above may seem counterintuitive. Although the “lower goals drives success” idea spreads across all categories, the average pledged amount per user in technology is substantially higher for successful projects. This suggests that the quality of the project can bring people interested enough to spend more.

Overall, the results reinforce the idea that the size of the campaign’s goal affects the probability of success. However, if a high-goal project of good quality includes an

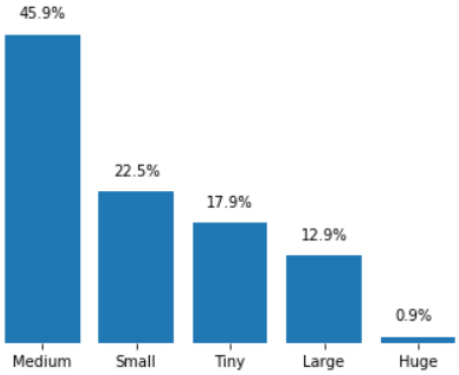
attractive reward such as a gadget or something tangible, the project can be successful with probability between 50 and 60%, similar to other top categories like Art and Music.

Pledged Amounts by Categories

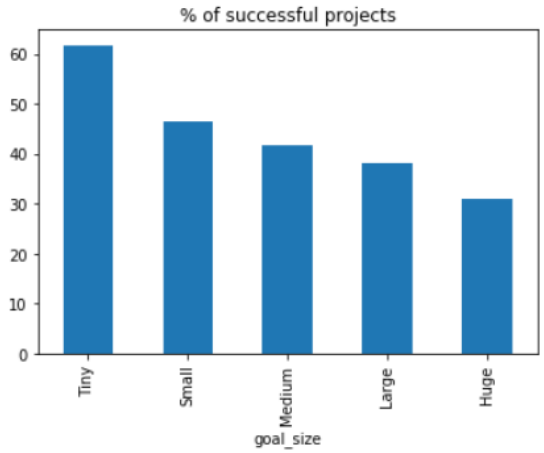
For the following analysis, we divided the amounts pledged into five bins: Tiny (<1k), Small (1k to 2.5k), Medium (2.5k to 10k), large (10k to 50k) and huge (>50k). For technology, more than half of the projects ask for more than 10 thousand dollars (large), with 12.4% of them asking for more than 50 thousand dollars (huge).



Let's see the results for the most successful category: Music



For the overall database, success looks negatively correlated with goal size:



To make a sense on how goal size and success correlate, we made a pivot table with the project categories, the size of the goals, and the average success proportion:

goal_size cat	Tiny	Small	Medium	Large	Huge
art	62.8%	60.6%	49.1%	30.9%	14.4%
comics	81.9%	74.7%	68.3%	59.1%	31.0%
crafts	61.7%	43.0%	36.2%	26.7%	7.9%
dance	84.4%	87.2%	82.8%	65.0%	10.5%
design	62.9%	58.1%	66.9%	65.0%	50.2%
fashion	51.8%	39.2%	37.8%	39.8%	10.9%
film	65.5%	67.6%	61.9%	47.7%	26.2%
food	40.5%	44.3%	40.5%	30.0%	7.3%
games	67.4%	61.7%	58.2%	52.6%	44.1%
journalism	38.8%	32.9%	27.4%	25.4%	9.7%
music	70.3%	71.7%	66.1%	46.9%	15.3%
photography	59.9%	54.0%	46.7%	47.7%	11.7%
publishing	75.7%	71.7%	67.1%	53.2%	22.8%
technology	56.2%	47.0%	38.5%	33.8%	30.1%
theater	79.1%	81.7%	72.5%	45.2%	13.6%

The results from this table are quite telling. As the project goal increases, successful rates fall overall. Nonetheless, technology seems to have lower than average rates of success across all sizes except huge.

A closer look of successful rates within the technology category shows the following:

goal_size	Tiny	Small	Medium	Large	Huge
sub_cat					
3d	53.3%	34.9%	37.3%	44.4%	43.4%
apps	35.8%	26.9%	21.0%	14.3%	3.1%
camera	75.0%	56.5%	43.8%	52.1%	57.7%
diy	75.1%	57.4%	49.1%	42.7%	23.3%
fabrication	63.2%	42.9%	29.9%	27.4%	37.8%
flight	40.5%	21.6%	23.6%	22.8%	15.0%
gadgets	61.8%	56.5%	46.0%	37.5%	27.1%
hardware	83.3%	79.0%	67.3%	57.0%	54.9%
makerspaces	68.2%	35.0%	46.4%	32.7%	5.0%
robots	58.1%	56.3%	47.8%	42.2%	38.7%
software	47.4%	36.3%	28.9%	13.0%	5.5%
sound	61.5%	47.7%	45.8%	46.1%	42.3%
space	76.7%	50.0%	47.7%	41.2%	18.9%
undisclosed	98.6%	94.4%	90.3%	89.4%	86.8%
wearables	59.2%	34.6%	29.9%	35.9%	35.2%
web	30.1%	15.2%	13.2%	10.4%	1.5%

Most of the success in huge (>50k) is due to projects that require prototypes, confirming our hypothesis that some quality check can boost confidence for backers, even for ambitious goals. Even then, there is not a single subcategory in which the success in huge surpasses that of tiny.

Concluding remarks

Given our findings, we have come to 4 key conclusions which tell a story about Kickstarter as a whole and about what projects are most successful on the site.

First, Kickstarter opened a new space for people with no access to traditional funding to obtain resources for their projects. The results reflect a bias in traditional funding against artists, which in response, flocked to Kickstarter and make up three quarters of the available projects (90% if you count food and games).

Second, one might expect that once the campaign goal is reached people would stop contributing to the campaign, but this should be seen in context. Art projects are not looking to be viable in a business sense, and thus, most of the backers have a more altruistic approach, pledging more than the project's goal.

Third, technological projects fail the most even though they make up a minority of the platforms' projects. One part of the explanation is that technology has more access to traditional funding channels, which means that many projects might use Kickstarter only after being rejected by banks or venture capital. Also, many projects could be listed on indiegogo, which is a more tech focused platform. Overall, this study suggests that if a project has a high goal, it must offer high rewards and provide proof of feasibility, in order to improve chances of success.

Finally, while it's true that many creators don't have a lot of flexibility in terms of the project they are attempting to fund, there are a few suggestions that may improve the rate of success. If the project is artistic, make sure you are making it as efficient as possible. Could you get materials at lower prices? Are there any features that could be added if the amount pledged suprasses the goal? If people really like your project they will likely pledge more than the stated goal, and you have an average margin of 16% to reduce your goal and manage to complete it with overfunding.

The case study shows that as goal amounts increase, quality standards matter much more, as shown for hardware projects in technology. So, if your goal is larger than 10k you should focus on showing your experience and ability to make the project a reality.

For technology, further work on viability and quality checks could be made, since these tend to be more expensive. In fact, we believe banks and venture capital may be interested in the results of this study, as they could provide know-how for ailing technology projects, especially in terms of improving feasibility.