Grace Lin, Kirby Bloom, Leo Wang

Understanding Success
Analyzing the signs of a Kickstarter project

Introduction

Kickstarter is a funding platform for creative projects. Everything from films, games, and music to art, design, and technology. Kickstarter is full of ambitious, innovative, and imaginative ideas that are brought to life through the direct support of others.

Everything on Kickstarter must be a project with a clear goal, like making an album, a book, or a work of art. The funding goal is the amount of money that a creator needs to complete their project. Backers are folks who pledge money to join creators in bringing projects to life.

Funding on Kickstarter is all-or-nothing. No one will be charged for a pledge towards a project unless it reaches its funding goal. This way, creators always have the budget they scoped out before moving forward.

We set out to investigate the data elements of Kickstarter projects in hopes of identifying attributes for creators to shoot for and those to stay away from.

Research Questions

In order to best characterize the data, we decided to break down the our investigation tasks into 3 main focus areas.

1. Categorical
   - Which categories are strongly correlated to success or failure?
   - Within the top categories, what is distribution of success across their sub-categories
   - Which categories attract the most backers?
   - Which categories require the largest funding amounts?

- What currencies are most used by projects?
- Do countries have any impact on category or project fate?
- What's the relationship between goals and pledged? Any correlation? Which category raise most above goal?

2. Timing
   - What is the ideal project timeline?
     - Does a longer timeline increase the fundraising amount?
     - What is the average timeline per category?
   - Is there an optimal month to launch or end a project?
   - Year to year comparisons
     - What categories are trending upwards / downwards over the years?

3. Numbers & Cents
   - What impact does the goal amount have on success / failure?
   - Do more backers equate to success?
   - What is the average amount pledged per backer?
     - Does the impact success?
     - Does it change per category
   - What percentage do failed projects typically run short?

The data

URL: [Kickstarter 2018 Data](#)

For this project, we are using the ks-projects-201801.csv file for our primary research. This dataset contains the features above for all projects launched from May of 2009 through February of 2017.

Published Features: ID , name , category , main_category , currency , deadline , goal , launched , pledged , state , backers , country , usd pledged, usd_pledged_real, usd_goal_real

Additional Data Sources

The url above contains an additional dataset with the same general information from 2016. We will be leveraging it to do a sanity check against the data accuracy for projects that overlap.

We will also be using some other external data sources to cross reference our findings. In particular, we would like to see how the trends we identify compare against the top 10 kickstarter projects.

We also used a list of country codes to translate the 2 letter country abbreviations into actual country names @ Country Codes

Initial Exploratory Analysis & Data Preparation

We analyzed a total of 378,661 projects which have an average of 106 backers contributing to around $9,683 in funding. Upon running describe against the data frame, we identified a few outliers that warrant some more investigation but all in all, the numerical series mostly seem to be fairly in line. Though there are some outliers, the distributions seem to be ok.

	ID	goal	pledged	backers	usd pledged	usd_pledged_real	usd_goal_real
count	378661.00	378661.00	378661.00	378661.00	374864.00	378661.00	378661.00
mean	1074731191.99	49080.79	9682.98	105.62	7036.73	9058.92	45454.40
std	619086204.32	1183391.26	95636.01	907.19	78639.75	90973.34	1152950.06
min	5971.00	0.01	0.00	0.00	0.00	0.00	0.01
25%	538263516.00	2000.00	30.00	2.00	16.98	31.00	2000.00
50%	1075275634.00	5200.00	620.00	12.00	394.72	624.33	5500.00
75%	1610148624.00	16000.00	4076.00	56.00	3034.09	4050.00	15500.00
max	2147476221.00	100000000.00	20338986.27	219382.00	20338986.27	20338986.27	166361390.71

After running some value counts against the rest of the data, the categories, currencies and state values where well curated and did not require attention.

We identify that there did look to be an encoding issue in country column as one country did not join with our country code list. There were 3797 rows coded for N,0" After looking up how many projects Kickstart has recorded for Norway, 786, we decided it was erroneous or coded for unknown.

Looking forward to our analysis, some convenience columns where required. In attempt to create a column for average duration, we noticed that the data columns came in as strings. These columns where converted and the duration column was created.   After running a describe on the newly created series, some outliers appeared. We were able to track it down to 5 rows that had a start date from January 1st, 1970. With such a small number and the projects all being unsuccessful, we left them in the data set.

```
count    378661
mean     33 days 21:58:10.533968
std      65 days 21:55:09.506515
min      0 days 00:07:17
25%      29 days 02:20:36
50%      29 days 16:32:32
75%      36 days 15:35:44
max      16738 days 23:00:00
```

One additional convenience column was added to track the average amount of dollars pledged per backer by simply dividing the the amount pledged in USD by the number of backers per project.

Similarly, when we compare 2016 dataset with 2018 dataset, we found the type of the pledged USD amount is string as well. When we converted the string to float, there were 623 observations (0.1% of total 323,750 observations) with characters in the USD amount column. We used 0 to replace those values.

Our findings

The following sections will walk through our exploration into the research questions defined above.

Categorical Overview

In terms of the number of projects by main category, Film & Video has the highest number of projects at 63,585, followed by Music and Publishing. Dance, Journalism and Crafts have the fewest projects.
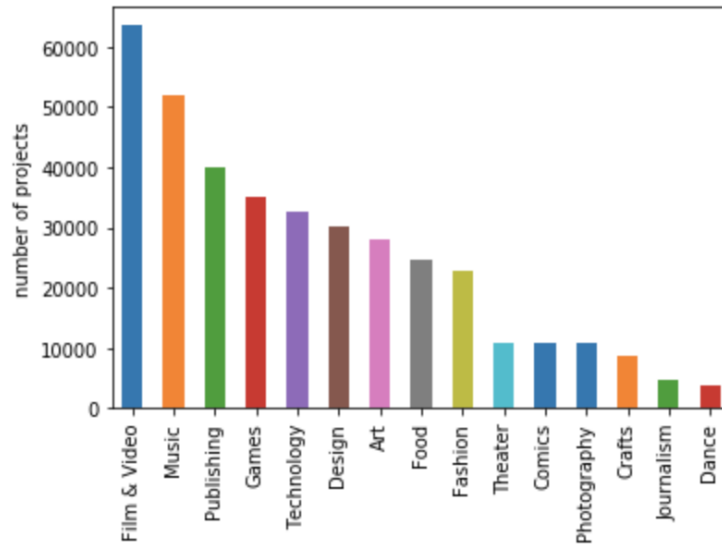
By currency, a dominant number of projects used USD as the currency.
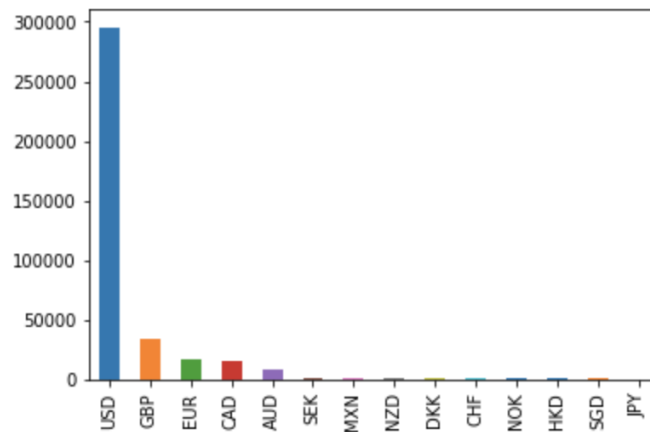

In terms of pledged dollar amount by main category, Design category raised the highest amount, followed by Games and Technology.
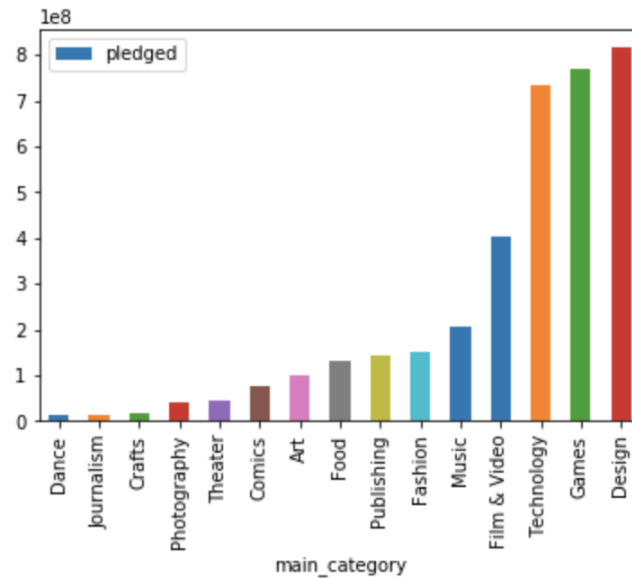
If we're looking at just the successful projects, the pledged amount for these successful projects by category is pretty much in the same order with Design, Games, and Technology as the top 3 categories.
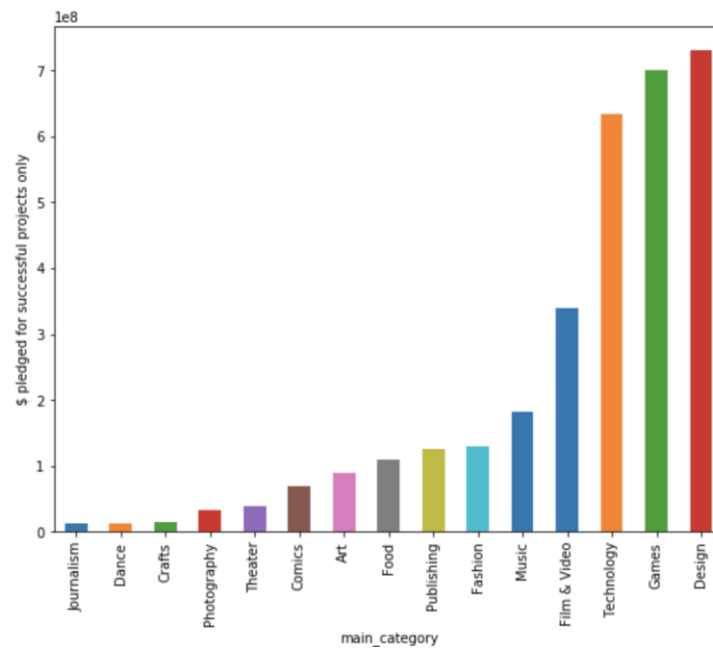
The ratio of pledged dollar amount to goal amount, measures how much a successful project raised over the goal. As shown below, successful projects in Design, Games and Technology category on average raised over 350% the goal. Projects in Dance and Theatre category perform relatively poorly, just raising over 100% the goal.
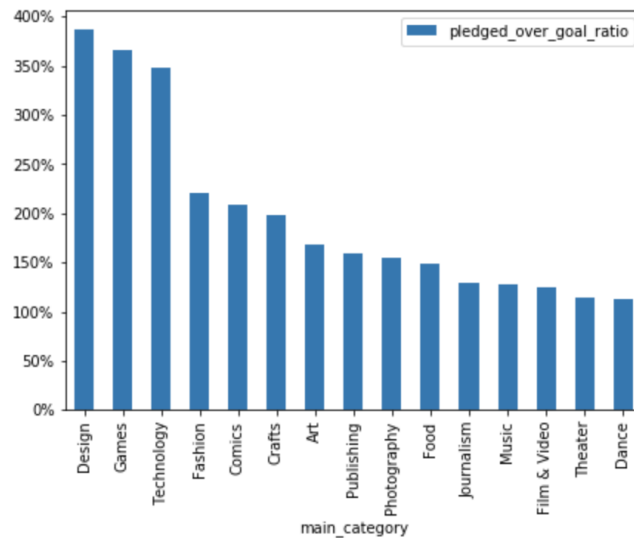
However, when it comes to Success Ratio, which is defined as the number of successful projects divided by the total number of projects per main category, we can see that Dance, Theatre and Comics have the highest success rate. It means that projects in these categories are more likely to meet their goals and become successful although they may raise just above the goal without exceeding too much.
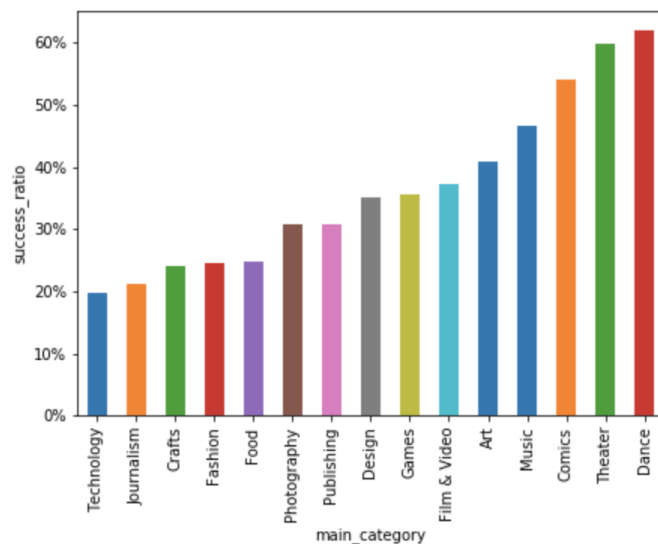
Plotting projects by category by state (sorted by failed), we found that Journalism, Food and Crafts have the highest failure rate. Technology projects have the 4th highest failure rate and actually are most likely to be canceled. Games and Design projects are also very likely to be canceled.
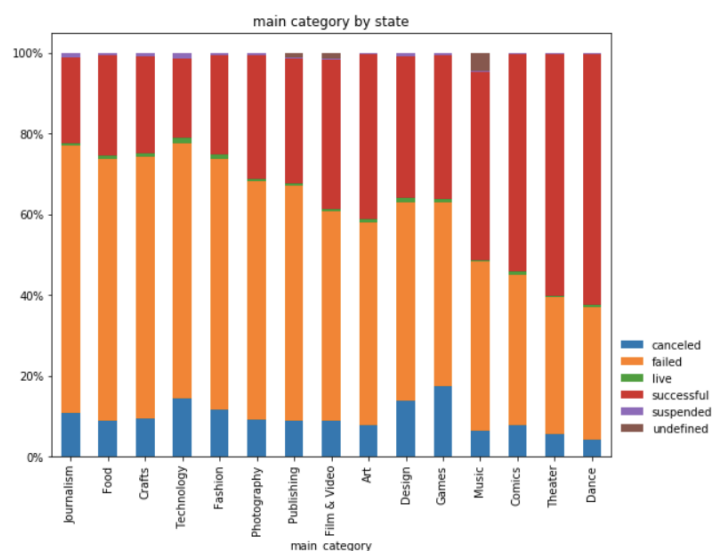

To understand what makes a project successful. We examined the correlations among various variables and found that the number of backers is highly correlated with pledged, usd_pledged and usd_pledged_real.

```
contrib.corr()
```

	id	goal	pledged	backers	usd_pledged	usd_pledged_real	usd_goal_real
id	1.000000	0.001679	0.000466	0.000654	-0.000633	-0.000025	0.001854
goal	0.001679	1.000000	0.007358	0.004012	0.005534	0.005104	0.942692
pledged	0.000466	0.007358	1.000000	0.717079	0.857370	0.952843	0.005024
backers	0.000654	0.004012	0.717079	1.000000	0.697426	0.752539	0.004517
usd_pledged	-0.000633	0.005534	0.857370	0.697426	1.000000	0.907743	0.006172
usd_pledged_real	-0.000025	0.005104	0.952843	0.752539	0.907743	1.000000	0.005596
usd_goal_real	0.001854	0.942692	0.005024	0.004517	0.006172	0.005596	1.000000

There is a positive relationship between the number of backers and the amount a project is able to raise, with the correlation coefficient around 70-75%. Below is a more intuitive illustration of that correlation.
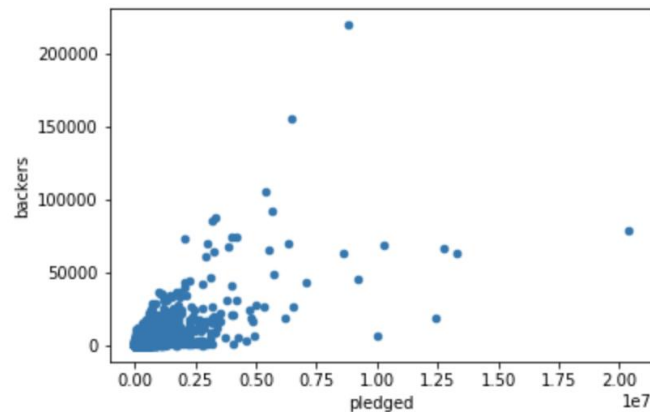

Triggered by the almost zero correlation between pledged and goal, so we plotted the correlation by state and found the successful projects have the highest correlation at 0.5.

| state		backers	goal
successful	backers	1.000000	0.430635
	goal	0.430635	1.000000
	id	0.001289	0.004960
	pledged	0.714339	0.500122
	usd_goal_real	0.487272	0.884641
	usd_pledged	0.694194	0.434677
	usd_pledged_real	0.749048	0.431234

Timing Overview

Through the previous analysis, we acknowledged that different categories had their own distinctive features. We started to wonder whether timing could impact the fundraising process.

We first explored the average fundraising duration among different categories. We defined duration as the time duration between project launch date and project deadline.

Duration for Successful Projects
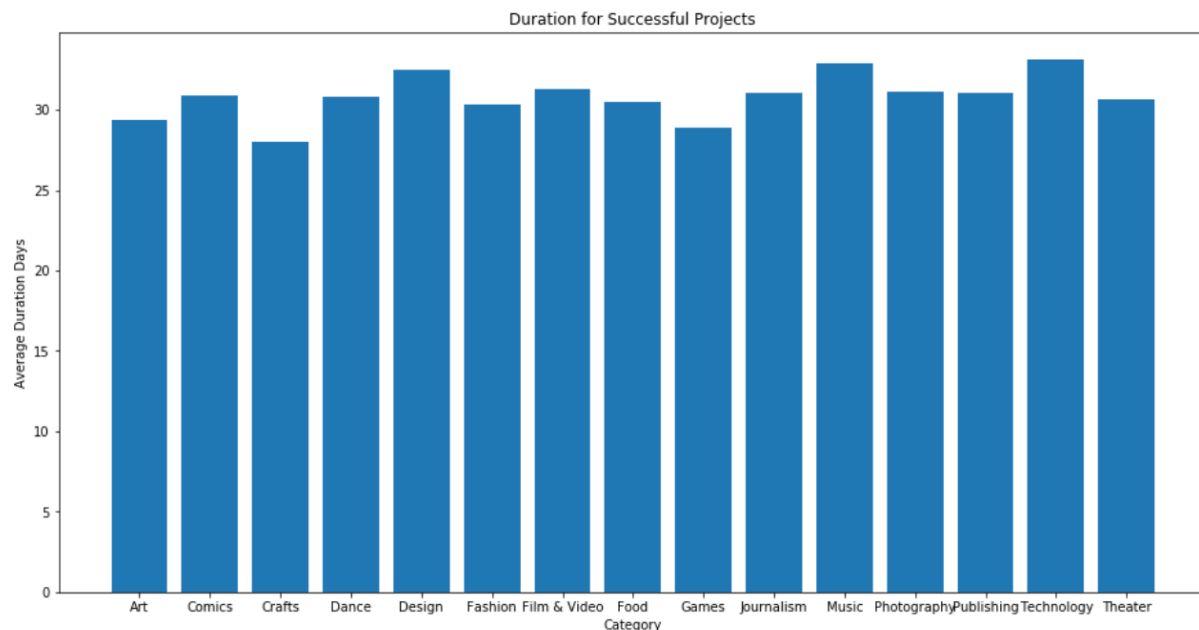
The above chart showed that the average duration was around 30 days for successful projects in all categories. And the following chart showed around 35 days for all failed projects.

Duration for Failed Projects
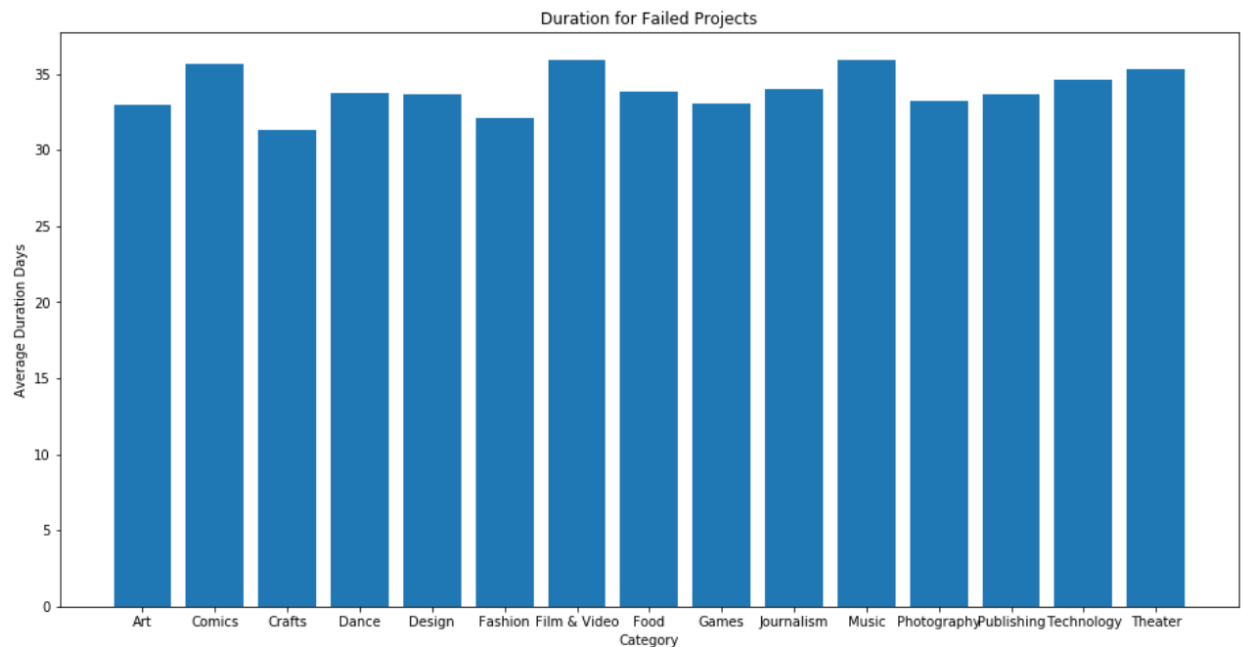
Since there was no much difference among categories in terms of timing, we were interested to know whether longer fundraising time would increase the total funding received. The following chart was a heat map for different features in this dataset:
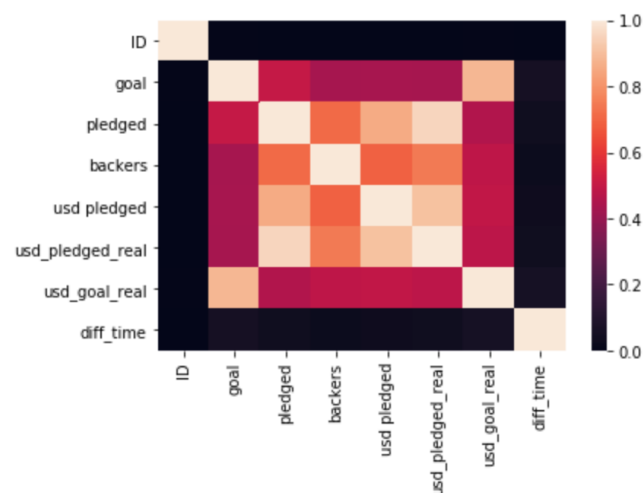
We further calculated the correlation between duration and usd_pledged amount, which was only 0.035. Both the heatmap and calculation indicated that there was no strong relationship between duration and raised amount.

What about project launch time and finish time? We grouped the time data by month

The relationship between launch month and pledged amount

The relationship between finish month and pledged amount
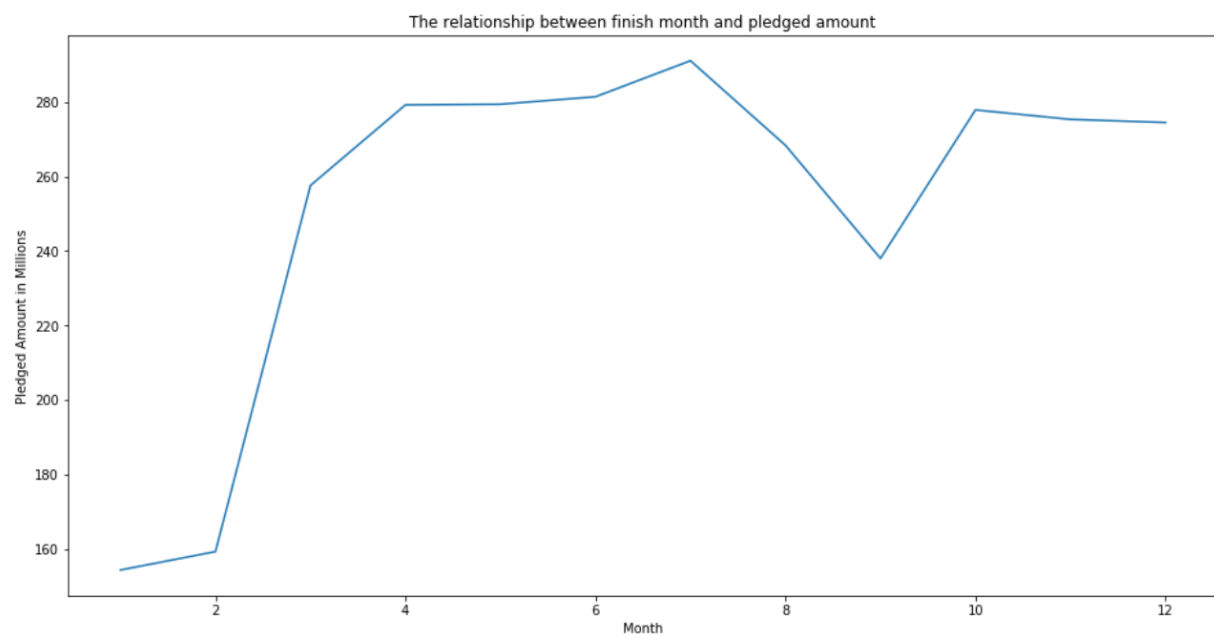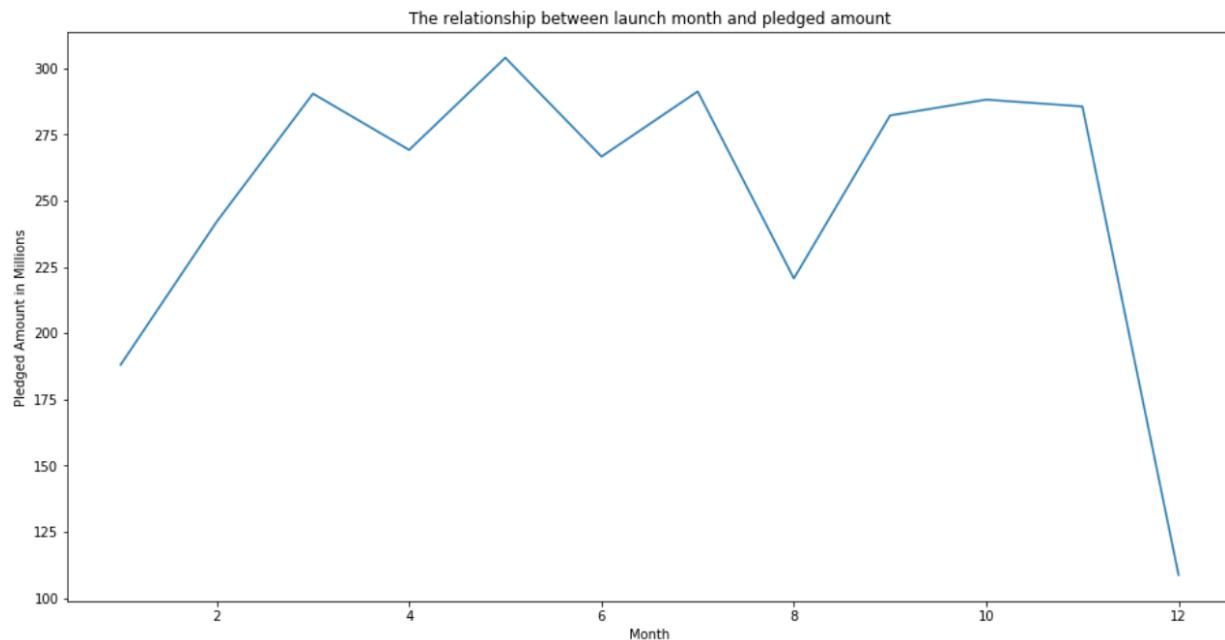
The best time to launch a project was probably from March to August and September to November. We should avoid winter holiday from December to January and summer vacation in August.

We further explored whether the trend was consistent from 2016 to 2018. We incorporated 2016 January Kickstarter projects in this analysis.

We grouped the dataset based on main category and calculate the difference between 2018 and 2016.

Difference in Category Percentage
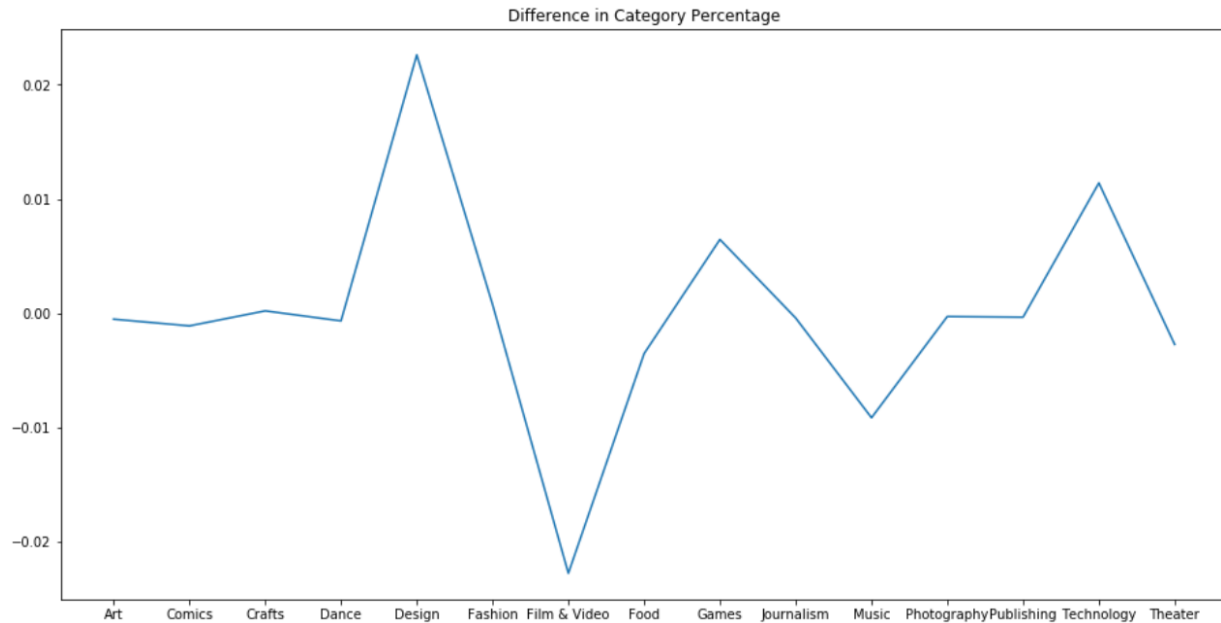
main_category	2018_usd_percentage	2016_usd_percentage	difference
Art	0.026424	0.026939	-0.000515
Comics	0.021902	0.023003	-0.001101
Crafts	0.003921	0.003709	0.000212
Dance	0.003999	0.004664	-0.000666
Design	0.218367	0.195768	0.022600
Fashion	0.037361	0.036595	0.000767
Film & Video	0.108527	0.131265	-0.022738
Food	0.034763	0.038285	-0.003522
Games	0.223530	0.217074	0.006456
Journalism	0.003447	0.003863	-0.000416
Music	0.058468	0.067604	-0.009136
Photography	0.011004	0.011287	-0.000282
Publishing	0.038221	0.038560	-0.000339
Technology	0.197213	0.185814	0.011399
Theater	0.012851	0.015570	-0.002719

The data showed that there was no significant change in category shares from 2016 to 2018.

Numbers & Cents Overview

After getting a better handle on the categorical and timing features of the projects , we took a look at a few other data points as they related to the financial terms and amount of support given for a project.

First we looked at the average goal requested and the fate of the project. This chart showed that the average goal for a successful project was less than half of those that fail. It was also interesting to see that the average amount of failed projects is lower than those canceled or suspended. This suggests that there may be a sweet spot for goal amounts.
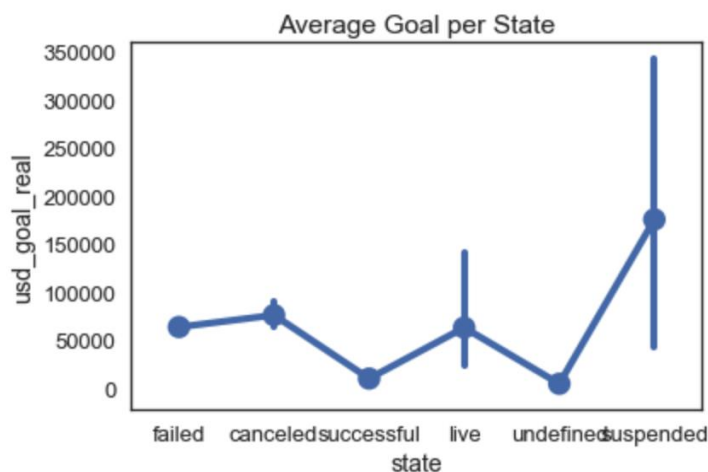
Next we took a peek at the average amount of backers and the fate of the project. The chart below showed support for what most would guess, the more backers a project had, the increased chance for success.
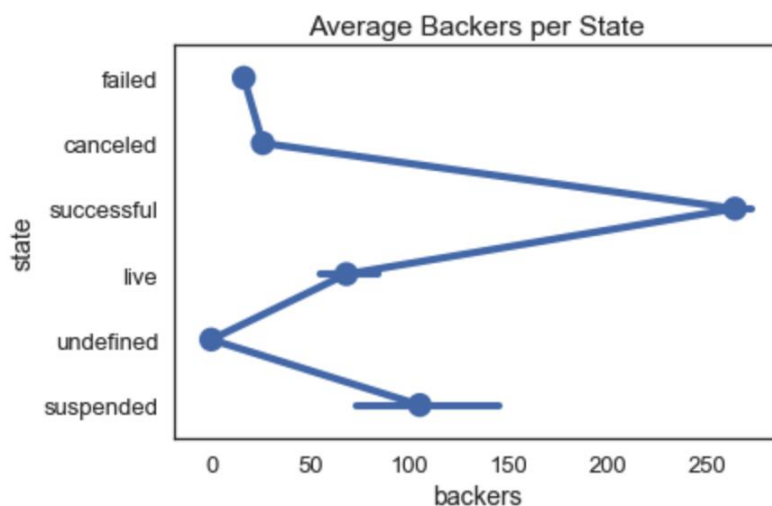
After looking at the previous chart for backers per state, the next question we asked was how much does a backer pledge, on average, for successful or unsuccessful project?

Successful	Not Successful
mean 91.12 std 125.53 min 0.78 25% 41.18 50% 63.31 75% 102.17	mean 64.78 std 158.41 min 0.45 25% 17.33 50% 36.17 75% 68.12

Interestingly enough, we found that the average amount given per backer for successful projects was also higher than those that were not successful. This suggests that backers are willing to commit more funds to quality projects. This could also be tied to a FOMO (fear of missing out) effect in which folks begin to pledge more as they see growing support.

Next we were curious to see if these trends held up across the individual categories. Please note the max dollar amount is almost half for the unsuccessful pledges versus successful.
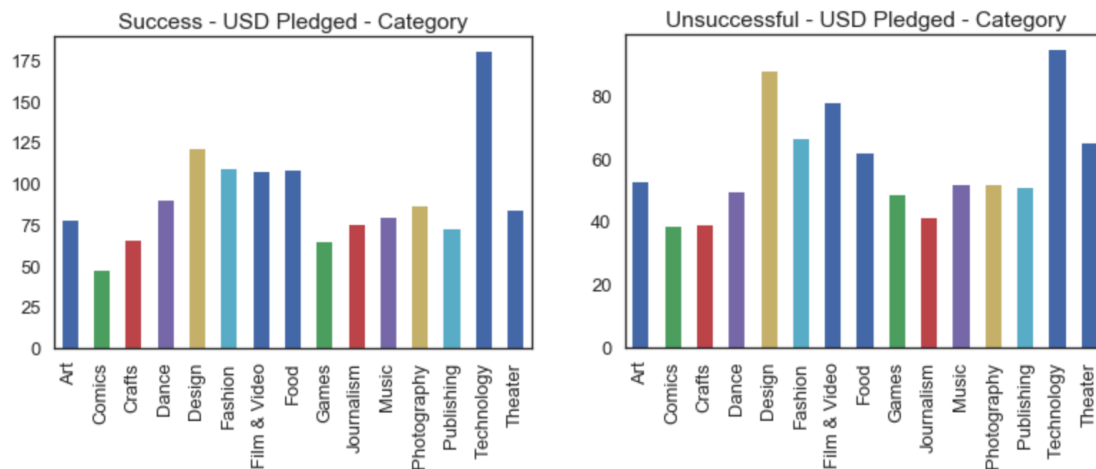
In general, the trend seems to be inline with no major category showing support differently than what the overall averages per back show.

Finally, we wanted to take a look at how short projects usually fall, when they fail. The list to the left shows that most projects come pretty close to reaching their goals which is bittersweet. However, leveraging the data from the category section above, it appears the harder they come, the harder they fall. Those categories that tend to pledge the most, also seem to be those who's gap is the biggest when they fail.

```
Games          1.23
Technology     0.56
Design         0.25
Film & Video   0.17
Comics         0.16
Fashion        0.16
Art            0.11
Theater        0.11
Crafts         0.10
Dance          0.10
Music          0.09
Publishing     0.08
Photography    0.08
Food           0.08
Journalism     0.05
```

Conclusion

In summary, for a Kickstarter project to be successful, being in one of Dance, Theatre or Comics industries would yield the best odds. Lower pledged goals would increase the success rate but does not necessarily increase the actual pledged dollar amounts. If the project can attract more backers, it could increase the success rate as well.

However, for a Kickstarter project to raise more funding, being in one of Design, Games, or Technology industries helps. This was also consistent when we cross referenced our findings with the top 10 kickstarter projects. These industries show to carry the most risk alongside tracking the biggest wins.

We also found that increased project duration does not necessarily increase the actual pledged amount. The average duration is around 30 to 35 days for both successful and failed projects and in general, we should try to avoid summer and winter vacation to launch the projects.

In the end, unicorns will prevail in the face of all the averages, but if you are looking for crowdfunding within some of the bumpers shown above, most projects can come pretty close to reaching their goals so it may be worth a shot!