

# Kickstarter: Exploratory Data Analysis

## Introduction

Kickstarter is an American public benefit corporation that maintains a global crowdfunding platform focused on creativity. According to the company, as of July 2020, more than 17M people have pledged more than \$4B to bring projects to life since 2009. How it works? - To raise funds, a 'creator' describes a project, sets a goal for the amount to raise, and appeals 'backers' to invest in the campaign. If the goal is reached, then creator receives all the funds pledged. Otherwise, the campaign will not be funded, and all the money is returned to the backers. Analysis on Kickstarter historic data would provide crucial insights for both creators and backers. This report tries to answer following questions:

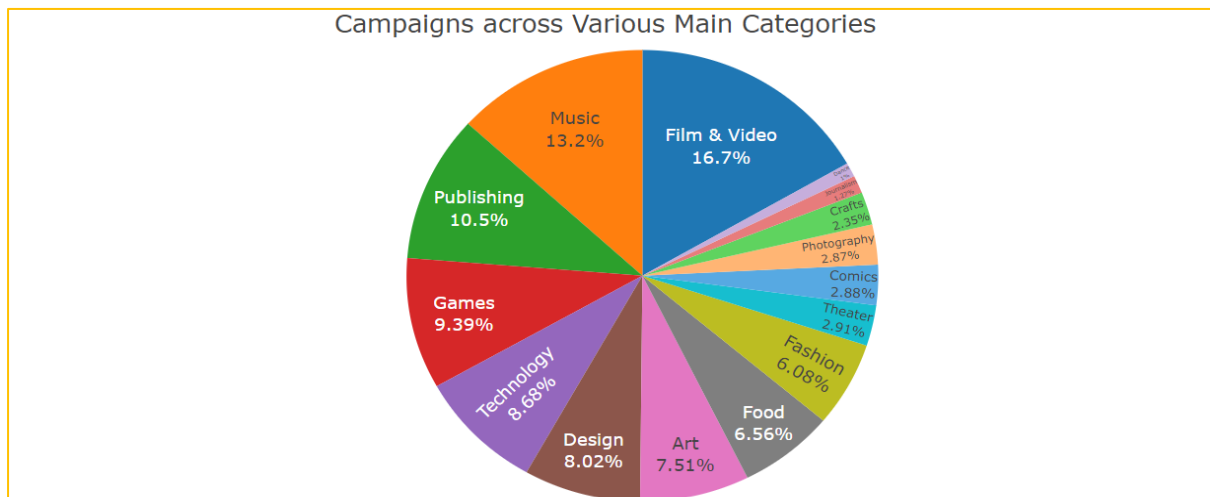
- Which main categories are more successful in reaching the goal?
- Which main categories attract more backers?
- Which main categories raise more money from successful campaigns?
- Are there any significant differences in the projects' success across countries?
- Does goal amount set by a fundraiser have any impact on the success of a project?
- Is there any noticeable trend over time?
- Does duration of the project influence the project success?

Answers to the above questions can guide 1) creators to target a section of audience by developing projects in the popular sectors and choose ideal parameters to maximize their chance to succeed, and 2) backers to invest in such successful projects.

## Data

The dataset contains information about 375k + projects with following major attributes:

- id: Project ID number
- name: Project name
- main\_category: Project main category (decided by project creators)



- category: Project subcategory within main category (decided by project creators)
- launched: When project was launched on Kickstarter
- deadline: Deadline by which fundraising goal should be met
- state: Whether project was successful (reached fundraising goal), failed (did not reach fundraising goal), or cancelled (cancelled early by the project creator)

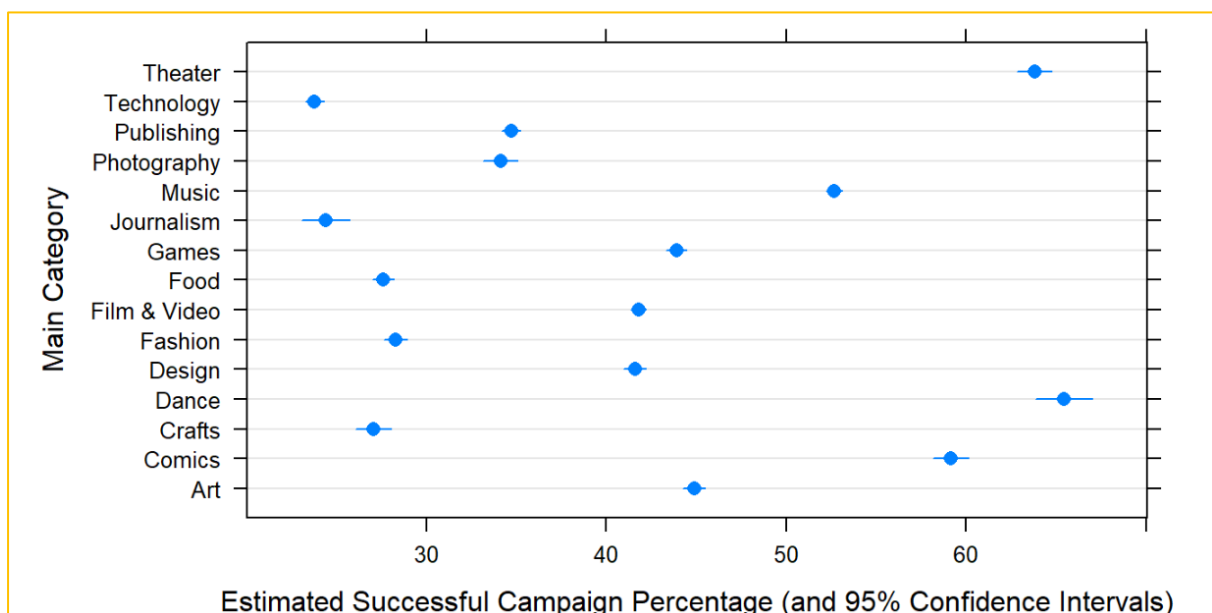
Var1	Freq
canceled	38777
failed	197718
live	2799
successful	133953
suspended	1846

- backers: Number of people who have contributed to the project by paying an amount of money
- country: Where the project is located
- usd\_goal\_real: Project fundraising goal in USD
- usd\_pledged\_real: Amount of money pledged by the deadline in USD

## Insights

### Which main categories are more successful in reaching the goal?

Around 50% Kickstarter campaigns belong to just 4 out of 15 main categories – namely Film & Video (16.7%), Music (13.2%), Publishing (10.5%) and Games (9.4%). Below dot-plot shows estimated success percentages with 95% confidence level for projects belonging to the 15 main categories. The dot-plot is constructed by conducting two-sided proportion-tests with the null hypothesis: Probabilities of success across main categories are the same (only successful and failed campaigns are considered).



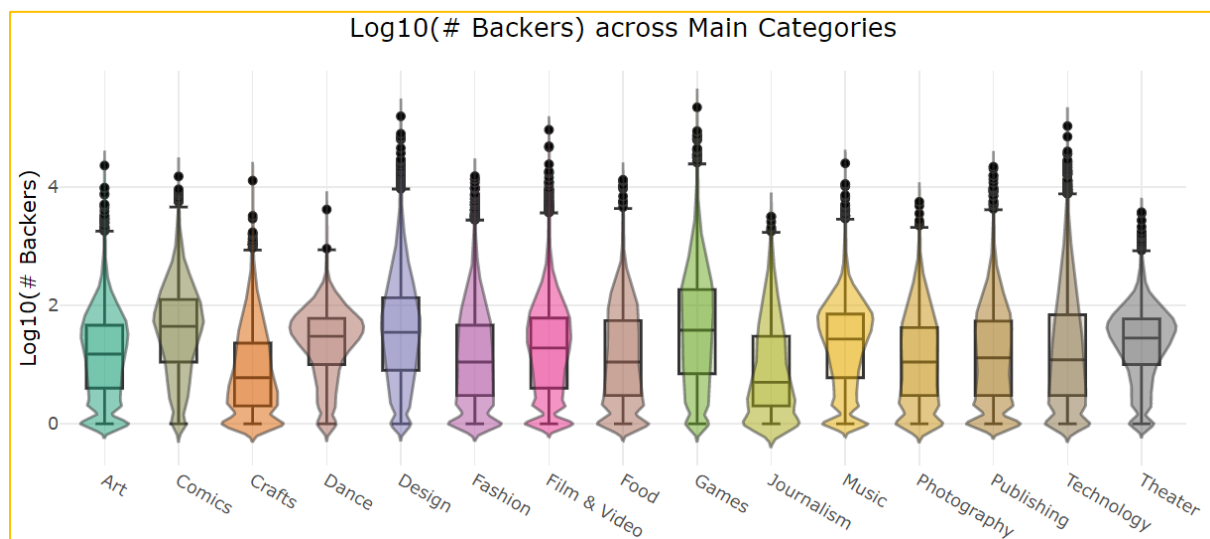
Dance (65.4%), Theatre (63.8%), Comics (59.1%), and Music (52.5%) stand out with highest successful percentages. However, these campaigns only account for approximately 6% of total campaigns. On the other hand, 3 out of 4 campaigns belonging to Technology and Journalism fail.

Conducting pairwise two-tailed proportion-tests have resulted in following set of main categories which have failed to reject the null hypothesis at 10% significance level:

Var1	Var2	value
Food	Crafts	1.0000000
Film & Video	Design	1.0000000
Technology	Journalism	1.0000000
Publishing	Photography	1.0000000
Food	Fashion	0.5855633
Theater	Dance	0.4915871
Fashion	Crafts	0.2895604
Games	Art	0.1597628

### Which main categories attract more backers?

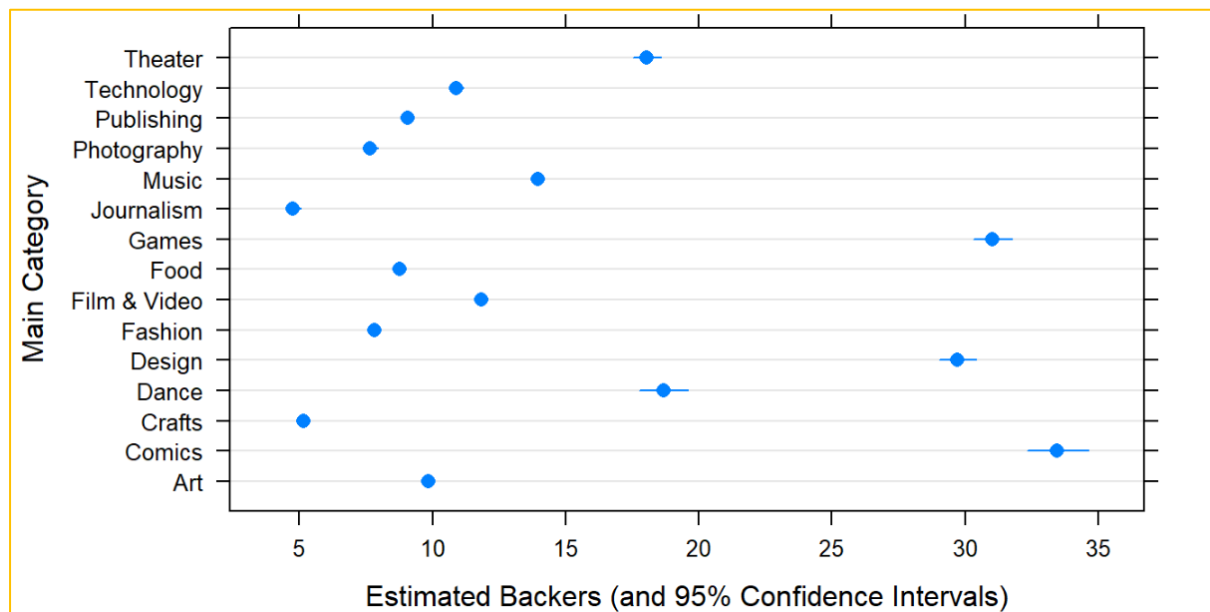
Generally, there are a greater number of projects with a lower number of backers and a lower number of projects with a greater number of backers. Hence the distribution is skewed to the higher side. By plotting logarithmic distribution/box plots of backers across main categories, couple of crucial insights can be found out.



Comics, Design, and Games have longer normal distributions which show that these projects have greater median/mean backers but also are highly uncertain to estimate comparatively. Dance, Music, and Theatre have fatter distributions on somewhat higher end, suggesting that these main categories have a greater number of projects with a decent number of backers and can be considered as safe bet.

Below dot-plot shows estimated number of backers (back transformed) with 95% confidence level for projects belonging to the 15 main categories. The dot-plot articulates formal two-tailed t-tests with

the null hypothesis: Estimated number of backers is same across main categories (all the project states are included).



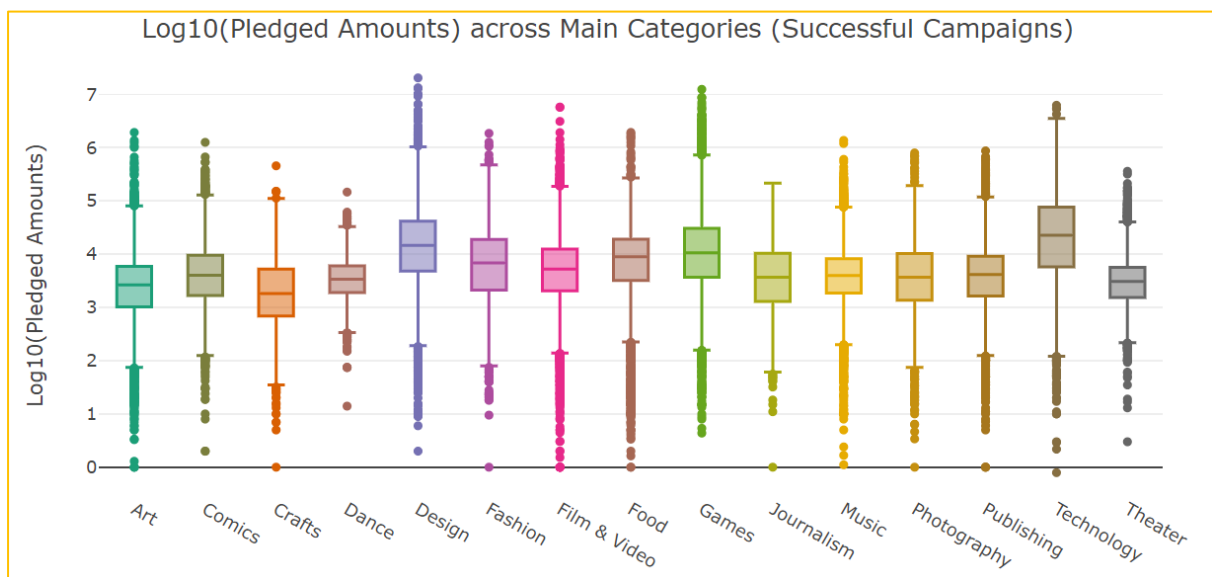
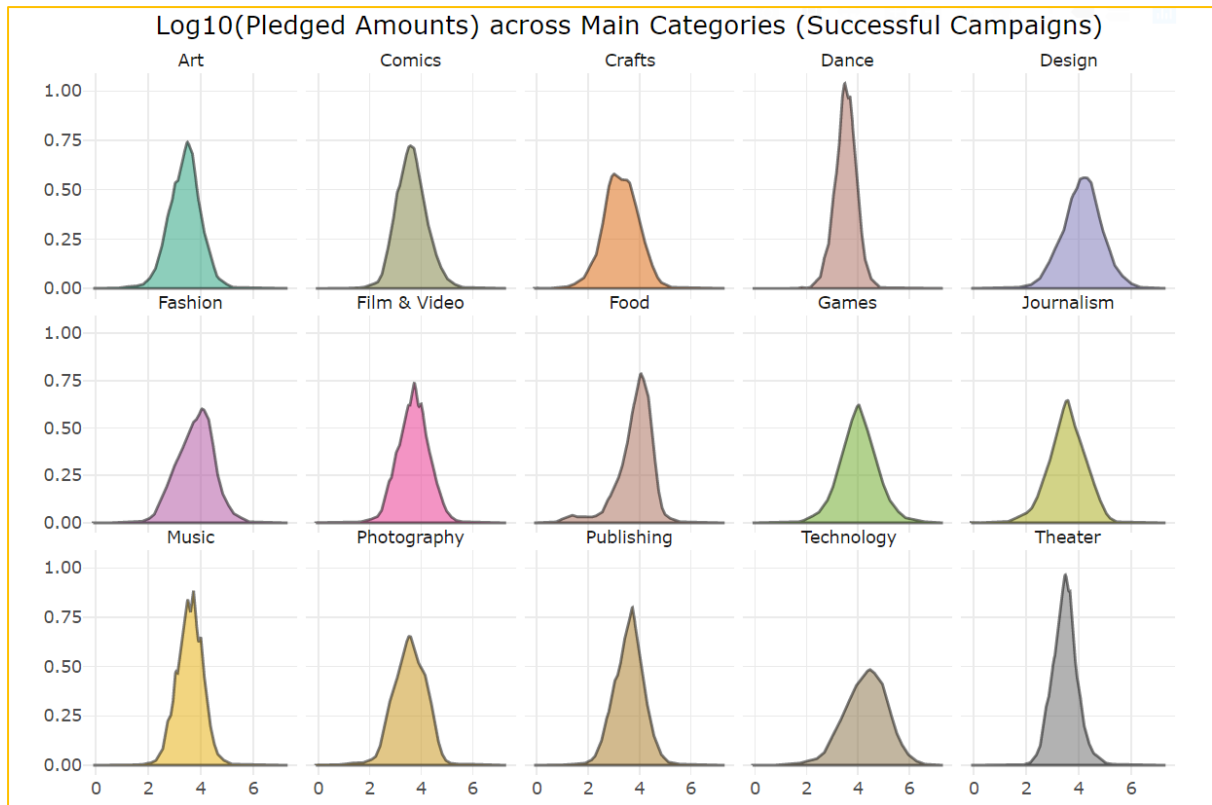
Coinciding with earlier reasoning, Comics (~33), Games (~31), and Design (~30) have greater number of estimated backers. Journalism and Crafts, on an average, attract very few backers.

Conducting pairwise Satterthwaite two-tailed t-tests have resulted in following set of main categories which have failed to reject the null hypothesis at 10% significance level:

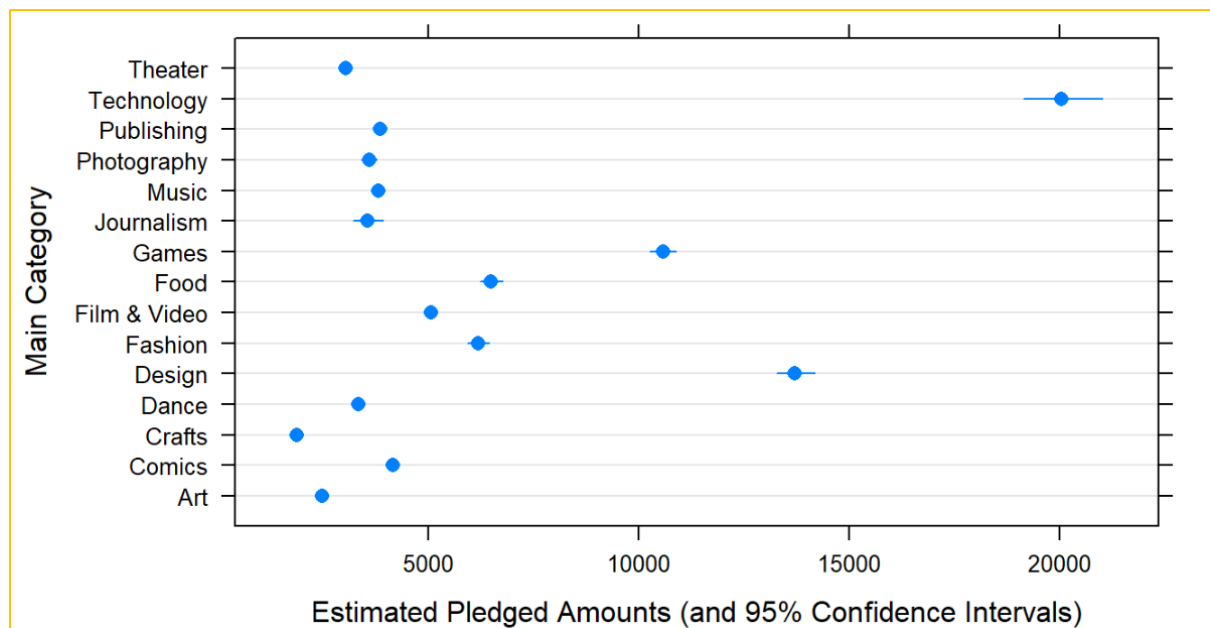
Var1	Var2	value
Theater	Dance	0.7160828
Photography	Fashion	0.7160828
Journalism	Crafts	0.2060869
Publishing	Food	0.2060869

### Which main categories raise more money from successful campaigns?

One of the main features of Kickstarter is that only when the pledges reach the funding goal value by deadline, campaigns are funded, or else money is returned. Hence it makes sense to only analyze pledged amounts of successful projects. The distribution and box plots of pledged amounts across main categories shown below suggest that the logarithmic transformation on right skewed data has satisfied the normality assumption fairly. Technology, Design, and Games seem to have greater median pledged amount as seen in the box plot.



Below dot-plot displays estimated pledged amounts (back transformed) with 95% confidence level for projects belonging to the 15 main categories. The dot-plot includes two-tailed t-test with the null hypothesis: Estimated value of pledged amounts is same across main categories (only successful projects are included).



It is surprising to see that some projects in Technology attract huge pledges (~\$20K) from a fewer number of backers (~11) comparatively. Design (~\$13.7K) and Games (~\$10.6K) come second and third in the pecking order. On the other hand, Comics attract lower pledges (~\$4.2K) from a greater number of backers (~33) on an average. Although Theatre, Dance, and Music have higher success rates with decent number of backers, they do not pull in huge bucks.

Conducting pairwise Satterthwaite two-tailed t-tests have resulted in following set of main categories which have failed to reject the null hypothesis at 5% significance level:

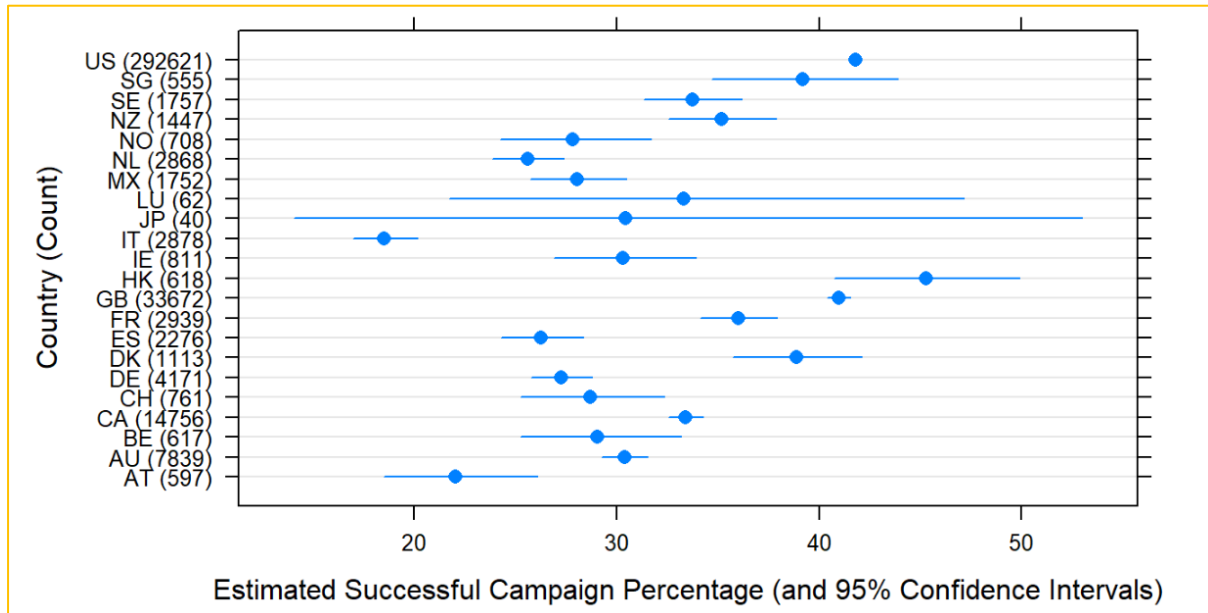
Var1	Var2	value
Photography	Journalism	0.8246424
Publishing	Music	0.7886746
Journalism	Dance	0.7809047
Music	Journalism	0.4947649
Photography	Dance	0.4382645
Food	Fashion	0.4382645
Publishing	Journalism	0.4382645
Photography	Music	0.2037243
Theater	Dance	0.0931892
Publishing	Photography	0.0931892

From the previous and present analyses, projects in Theater and Dance categories have similar estimated success rates, number of backers, and pledged amounts. This may be attributed to the intertwining nature of projects belonging to those categories.

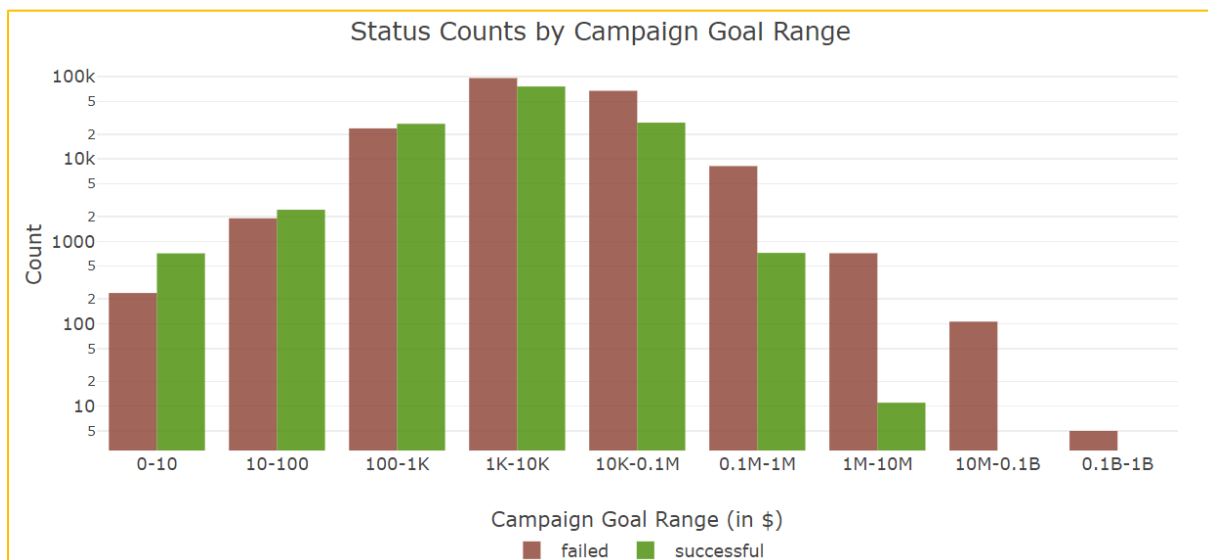
### Are there any significant differences in the projects' success across countries?

By examining the dot-plot which shows estimated success percentages of projects across countries, one can discover couple of striking points. The uncertainty in estimating can be linked with countries

having lower number of campaigns. This is evident in Japan (40) and Luxembourg (62) which have the lowest counts. Campaigns started in USA account for almost 78% of total and has success rate over 40%. Hong Kong (45%), Great Britain (41%), and Singapore (39%) have high success rates. The second point is that most of the success rates are significantly different and hence country effects should be considered in determining the success of projects.

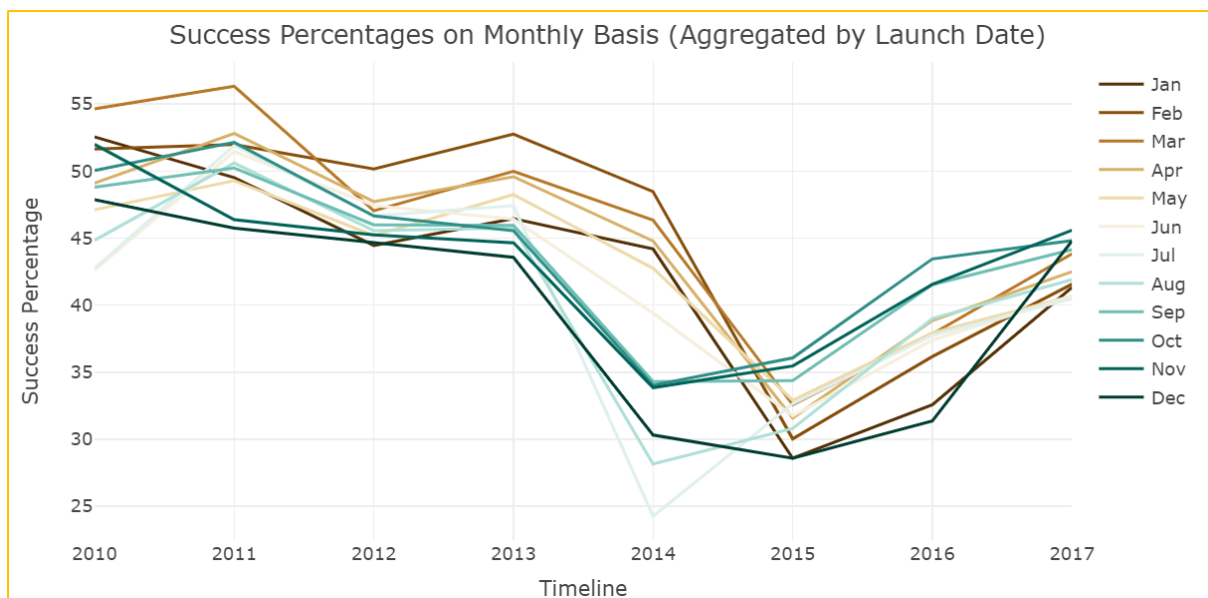
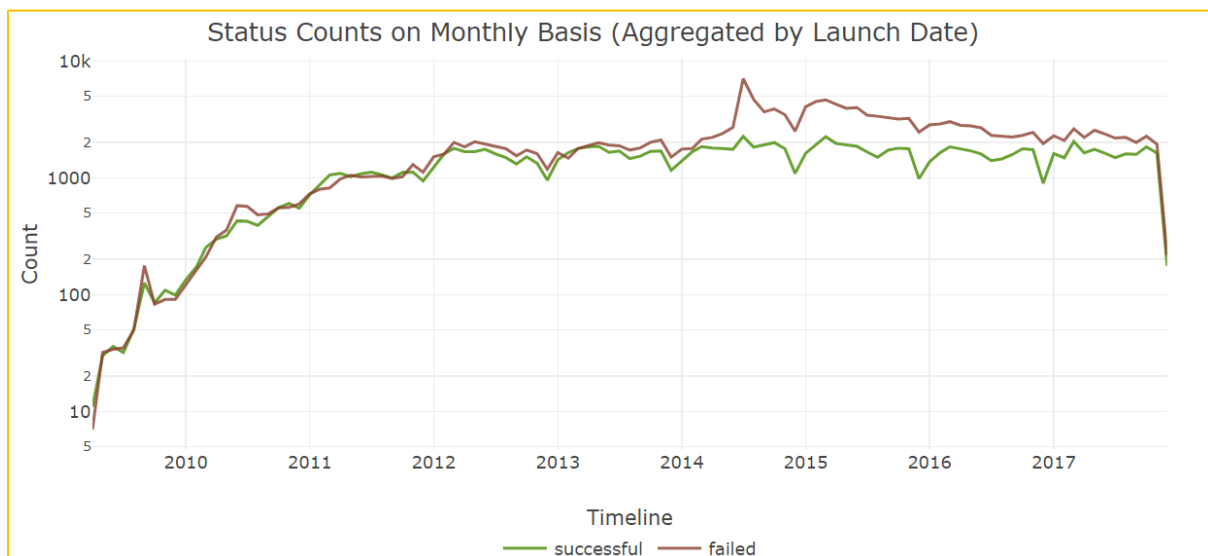


Does goal amount set by a fundraiser have any impact on the success of a project?



Kickstarter ventures are generally funded by raising small amounts of money from a large number of people. Most campaigns have goal value below \$1M. For a campaign creator, ideas that have realistic funding goal between \$1K to \$1M are generally suggestive. Only 11 out of 840 projects with funding goal over \$1M have succeeded. Therefore, the realistic odds of success decrease as the goal value of project increases.

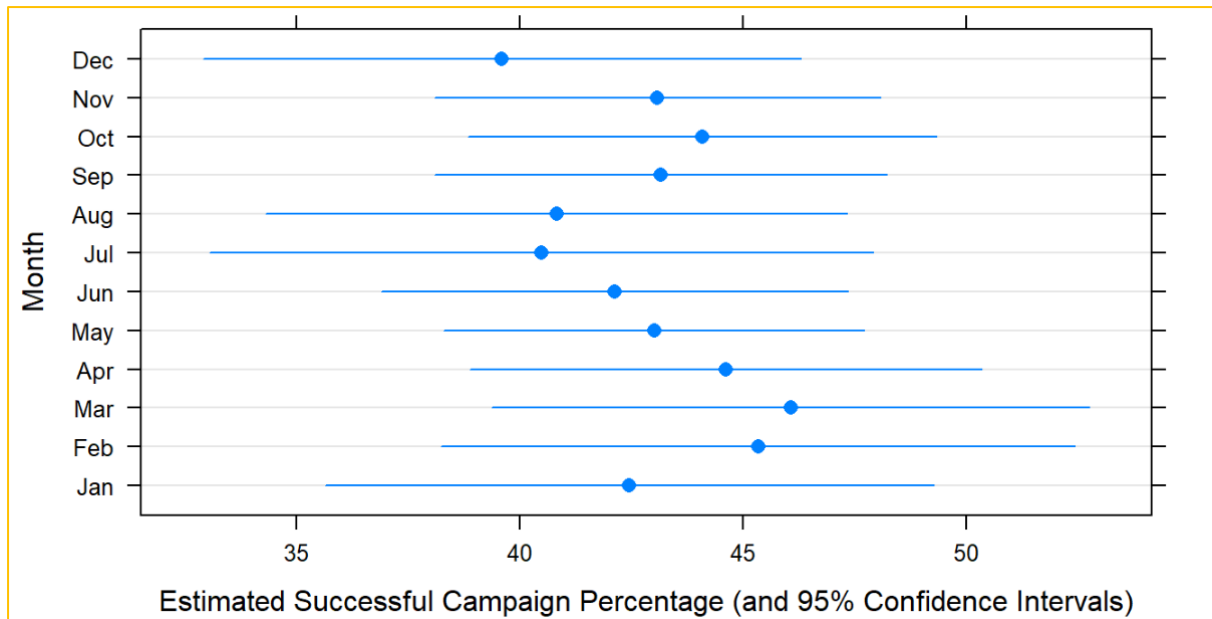
Is there any noticeable trend over time?



The first time-series plot detailing monthly project counts reveals that historically Decembers have fewer campaign launches since 2011. A few intriguing trends stand out by analyzing the second time-series plot which displays the success percentages on monthly basis from 2010 to 2017. Decembers and Julys have lower success rates for most years. Furthermore, campaigns launched in the first half of years from 2010 to 2014 have higher success rates compared to the ones launched in the second half. However, this trend reversed from 2015 onwards with higher success rates in second half. But it is imperative to check if these monthly trends are significant.

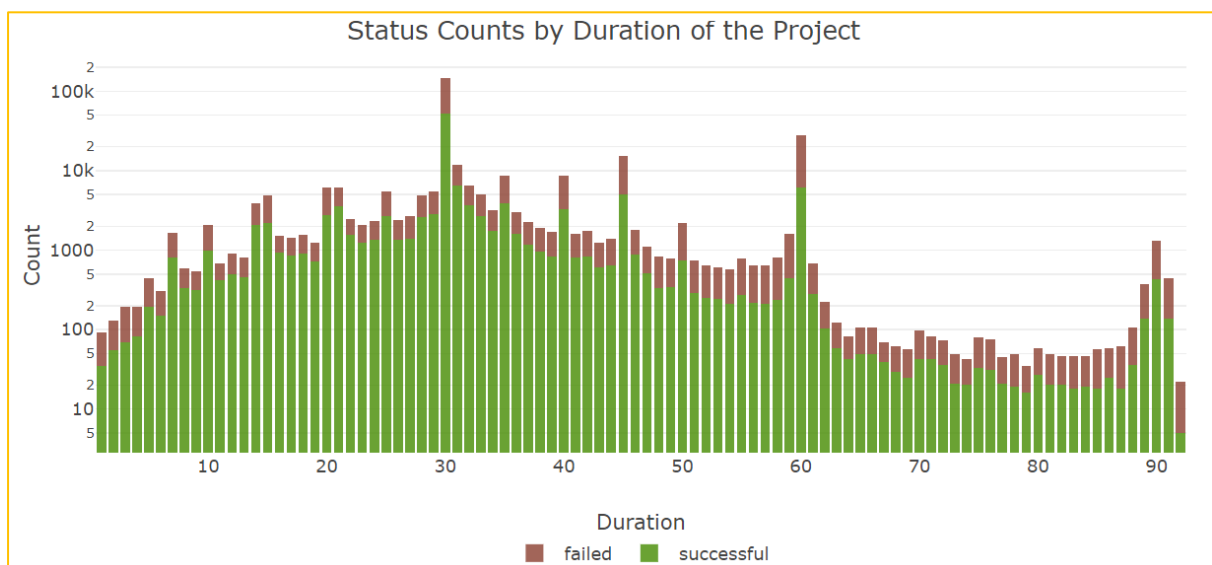
The below dot-plot is constructed by conducting a formal two-sided t-test and displays estimated success rates with 95% confidence level for projects launched in various months from 2010 to 2017. The dot-plot shows a clear pattern of bi-annual periodicity with January-June cycle and July-December cycle. The estimated success rate is highest in the month of March with 46% for the first half and is highest in the month of October with 44% for the second half. However, these rates are not significantly different at 5% significance level.



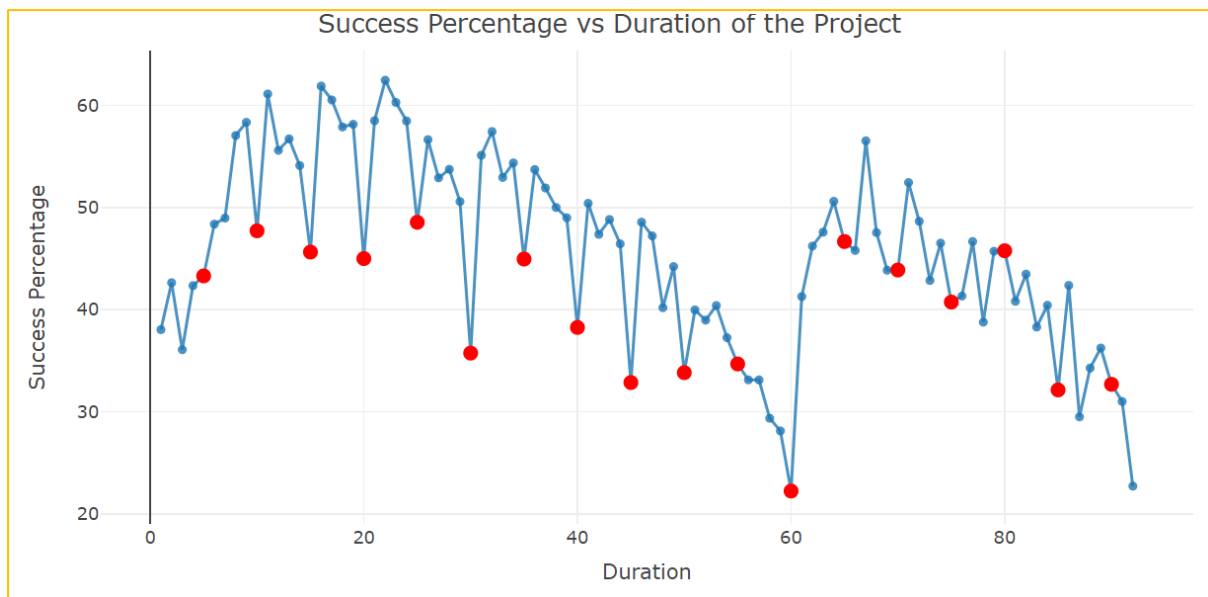


### Does duration of the project influence the project success?

Initially creators had an option to set funding duration for up to 3 months but later it was reduced to 60 days. This is also conveyed in the bar plot. Generally, creators tend to set project duration in multiples of 5 days.



The plot between success percentage and duration of project uncovers few crucial insights. Firstly, projects with lesser durations would not attract enough backers to support the cause and longer durations would generally backfire due to unknown reasons. Ideally, projects should have deadlines in the range of 2-4 weeks after they are created. Secondly, the success percentage does decrease for every multiple of 5 days which may be related to higher number of campaigns with such deadlines as explained by the former plot.



## Conclusion:

Kickstarter historic data has a lot of valuable insights to offer. As of January 2018, around 375K+ campaigns were hosted on Kickstarter out of which 36% were successful since its inception in 2009. Successful campaigns have brought in over \$3B from over 39M pledges. Pebble Time is the largest successful campaign which got funding over \$20M from over 78K pledges.

Generally, projects in entertainment categories like Music, Theater, Dance and Comics are most likely to succeed compared to Technology, Crafts and Journalism. Although, 3 out of 4 projects in Technology are bound to fail, these campaigns bring in huge bucks compared to projects in any other category. Music, Dance and Theater are safe bet categories which pull decent number of backers. Projects in Theatre and Dance have similar metrics due to their intertwined nature. Project success rates in most of the countries are significantly different and the American market shows dominance due to the platform's roots.

There are other crucial metrics, that a creator needs to take note of. Kickstarter being a crowdfunding platform, the realistic funding goal is between \$1K to \$1M. Since 2011, fewer campaigns are launched in the month of December. Starting a campaign during February-April cycle and September-November cycle has a slightly higher chance to succeed. However, this is not significantly different across all the months. Setting a suitable duration is also important for creators to request funds. It is generally suggestive to set deadlines in the range of 2-4 weeks after the project is created.

Kickstarter has provided a platform for many artists, designers, makers, musicians, and creative people at every level. The exploratory data analysis presented in this report can guide creators develop projects that could maximize their chance to succeed. Feature engineering based on these factors can help choose strong predictors to determine the project's success which is not covered in this report.

## References:

- *Kickstarter Wiki*. <https://en.wikipedia.org/wiki/Kickstarter>
- *How it works*. <https://www.kickstarter.com/how-it-works>
- *Kaggle Dataset*. <https://www.kaggle.com/kemical/kickstarter-projects?select=ks-projects-201801.csv>