Alex Arnold
SMU Data Science Bootcamp
October 4, 2020

Excel Challenge: Kickstarter

The massively successful crowdfunding service Kickstarter has launched more than 300,000 projects but only one-third of these projects have been successful enough to make it through the funding process with a positive outcome. Therefore, what does it take for a Kickstarter project to get funded and become successful?

After organizing and analyzing a sample database of 4,000 past Kickstarter projects, I will identify trends in order to provide a recommendation that will most likely result in a successful campaign. I will analyze possible limitations to the data and any future work that could help support a project's further success.

To begin working with this data set, I altered the raw data to aid in determining trends amongst the Kickstarter campaigns. I applied conditional formatting to easily identify whether a campaign was successful, canceled, failed, or currently live. Then I calculated percent funded, as a campaign is only deemed successful once the goal has been met, and I used a gradient scale to easily identify those percentages. I calculated average donation to see what each backer for the campaign was contributing. I split the categories and sub-categories out into separate columns to better aid in more detailed analysis. Lastly, I converted the date created and date ended data from the Unix timestamps to dates for better readability.

Once cleansed, I began by using various pivot tables to aggregate this large data set into smaller groupings to help determine any possible trends. I first grouped by category and looked at the aggregate count of outcome. I then filtered by country to determine any possible trends amongst geos. Next, I created a similar pivot chart utilizing sub-category instead of category. I added additional filters of country and parent-category.

The analysis concludes that Theater is the most successful category (by count), the United States has the most Kickstarter campaigns (by count), and May is the most successful month (by count).

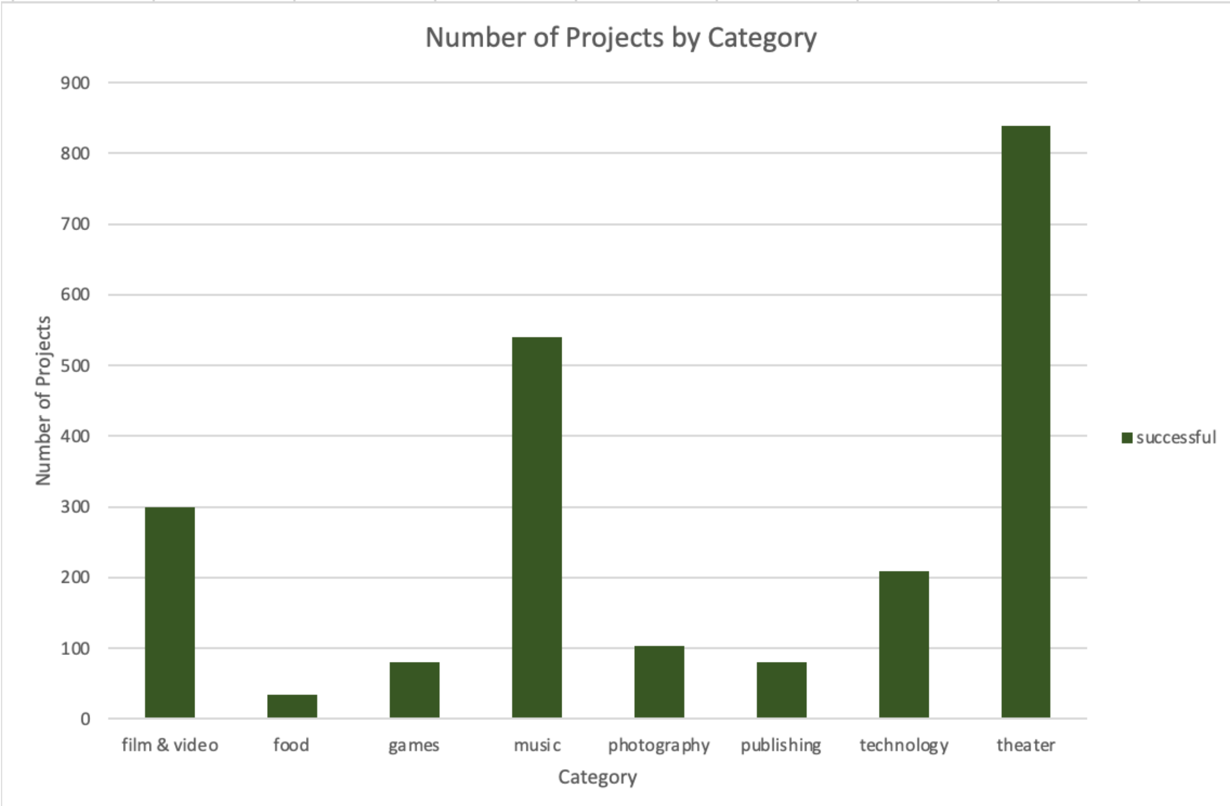

Figure 1 Number of Kickstarter Projects by Category

Figure 1 displays the successfulness the theater category for Kickstarter campaigns. There were almost 1400 theater campaigns with 839 being successful. This category also held the most failed campaigns (Figure 2), highlighting the enormity of this type of campaign overall. Dissecting the theater category even further, the plays sub-category is much more successful than the other sub-categories of musicals and spaces as shown in Figure 3.

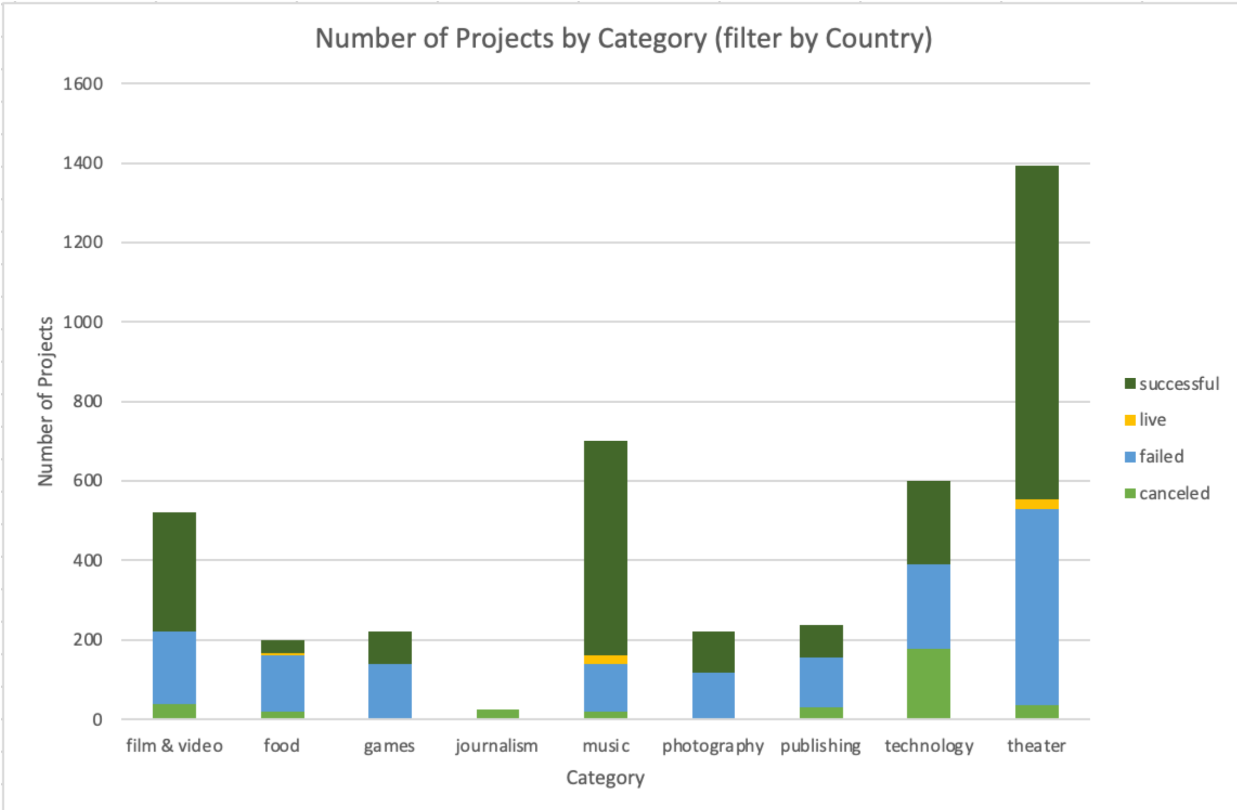

Figure 2 Number of Kickstarter Projects by Category

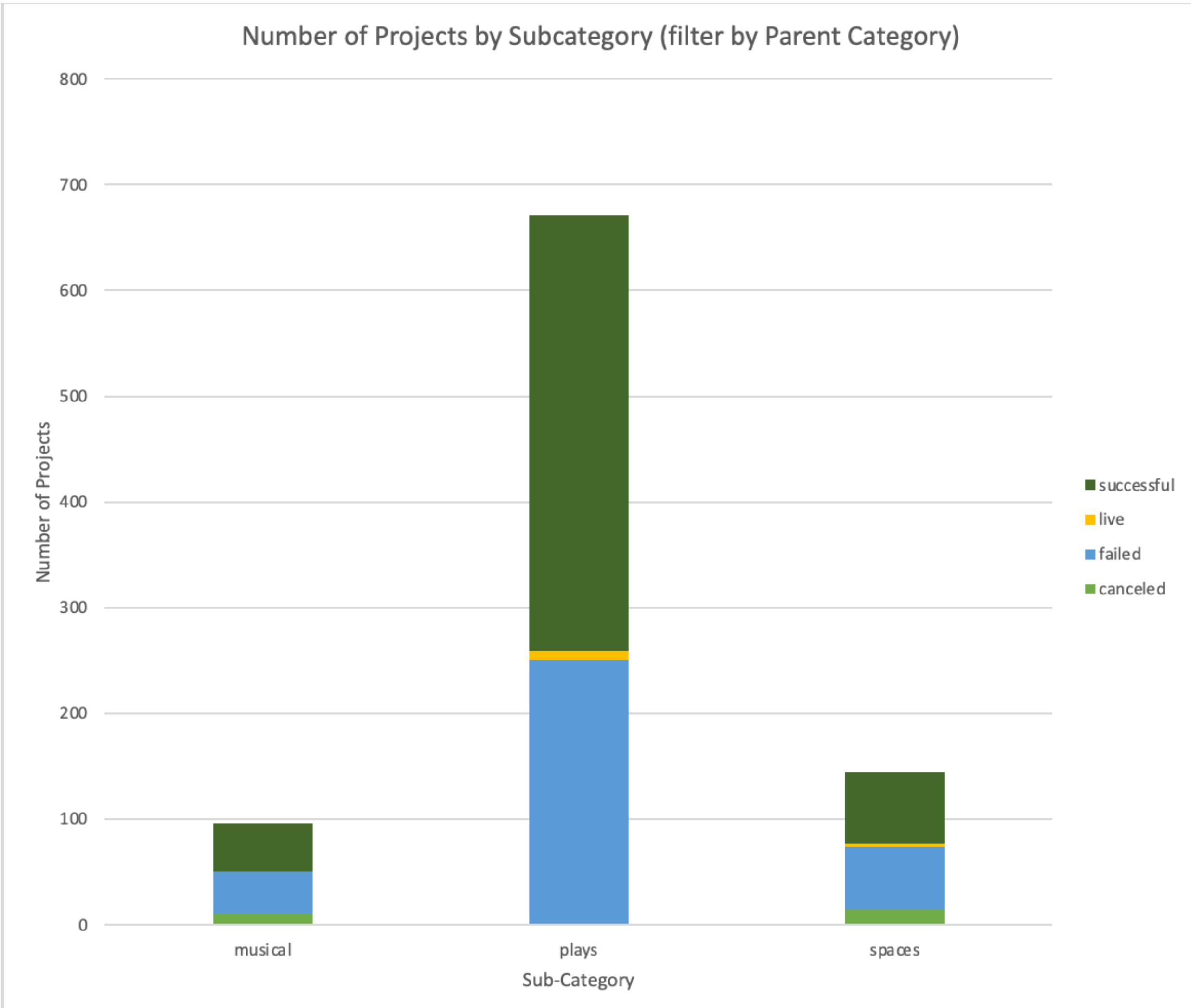

Figure 3. Number of Projects by Sub-category (filter by parent Category)

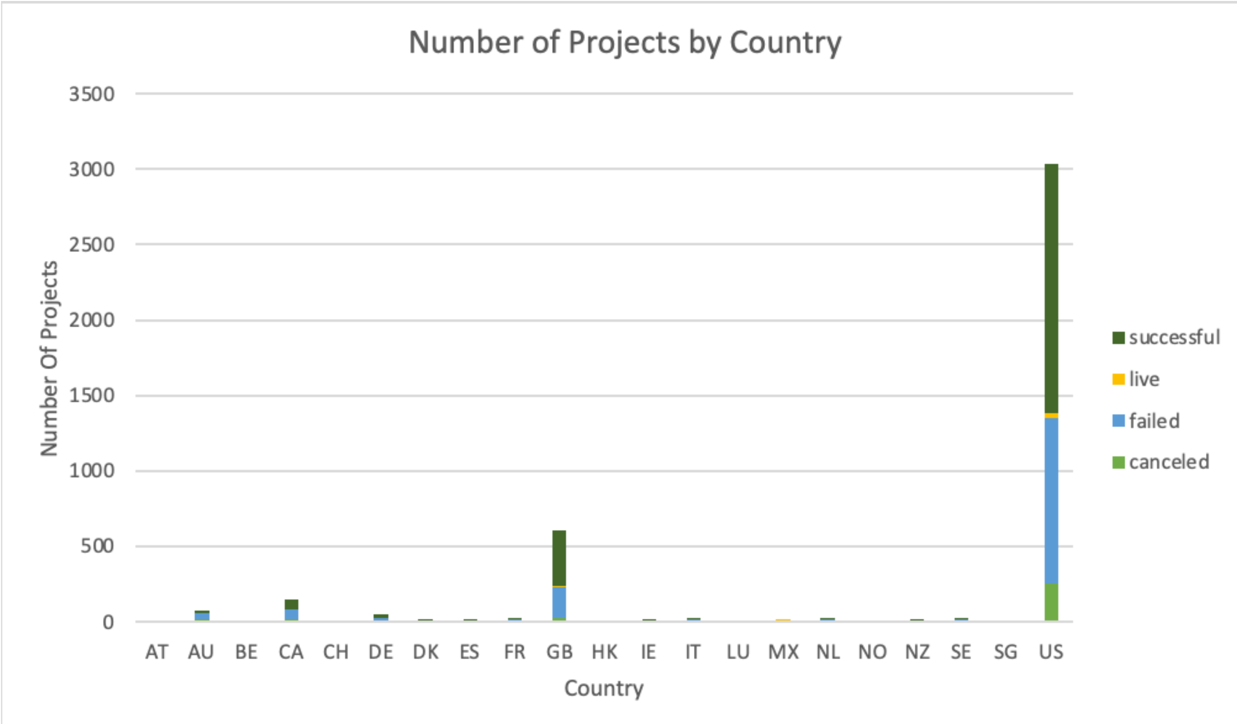

Figure 4. Number of Kickstarter Projects by Country

Country influences success as shown in every category of this data set. The United States reigns supreme in terms of the overall number of Kickstarter campaigns but more importantly, the number of successful Kickstarter campaigns (Figure 4).

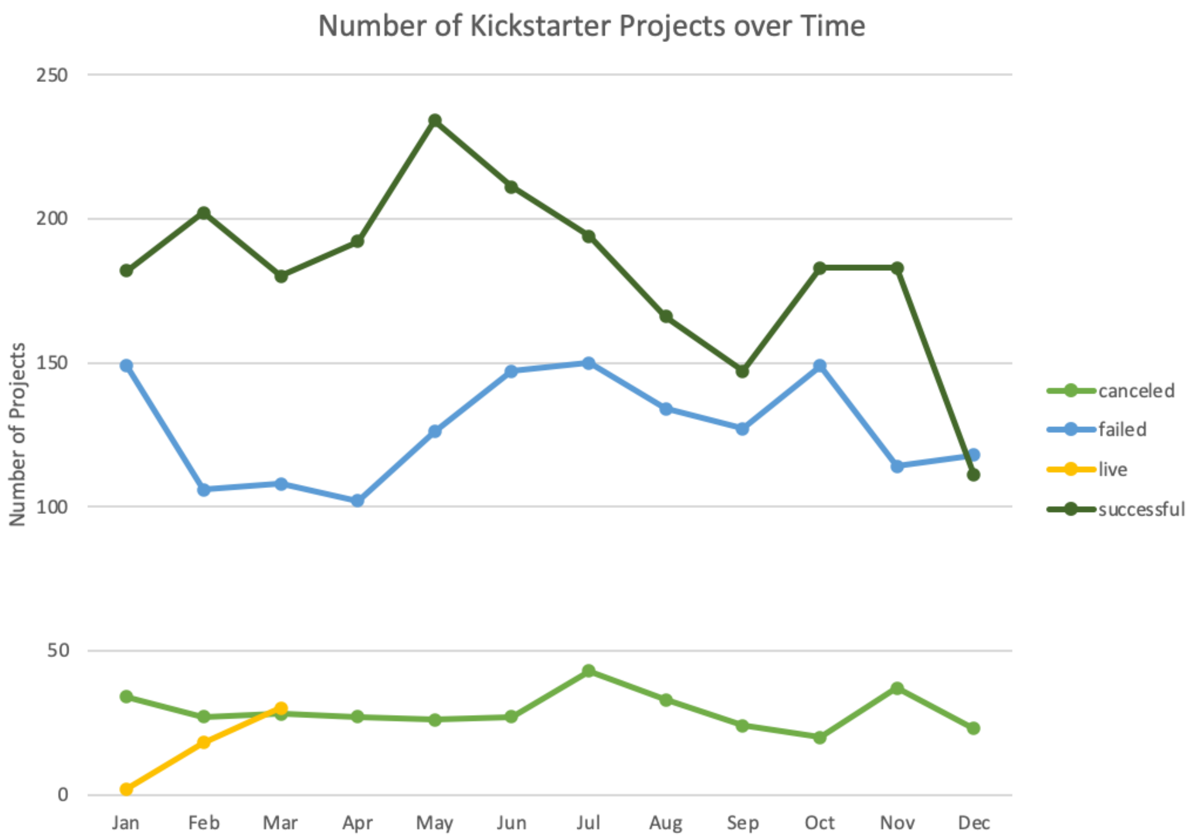

Figure 5 Number of Kickstarter Projects over Time

Lasty, month plays a role in a Kickstarter campaign's success as shown with the overwhelming success in May (Figure 5). This trend of success in May is also reflective in the highly successful theater category (Figure 6).

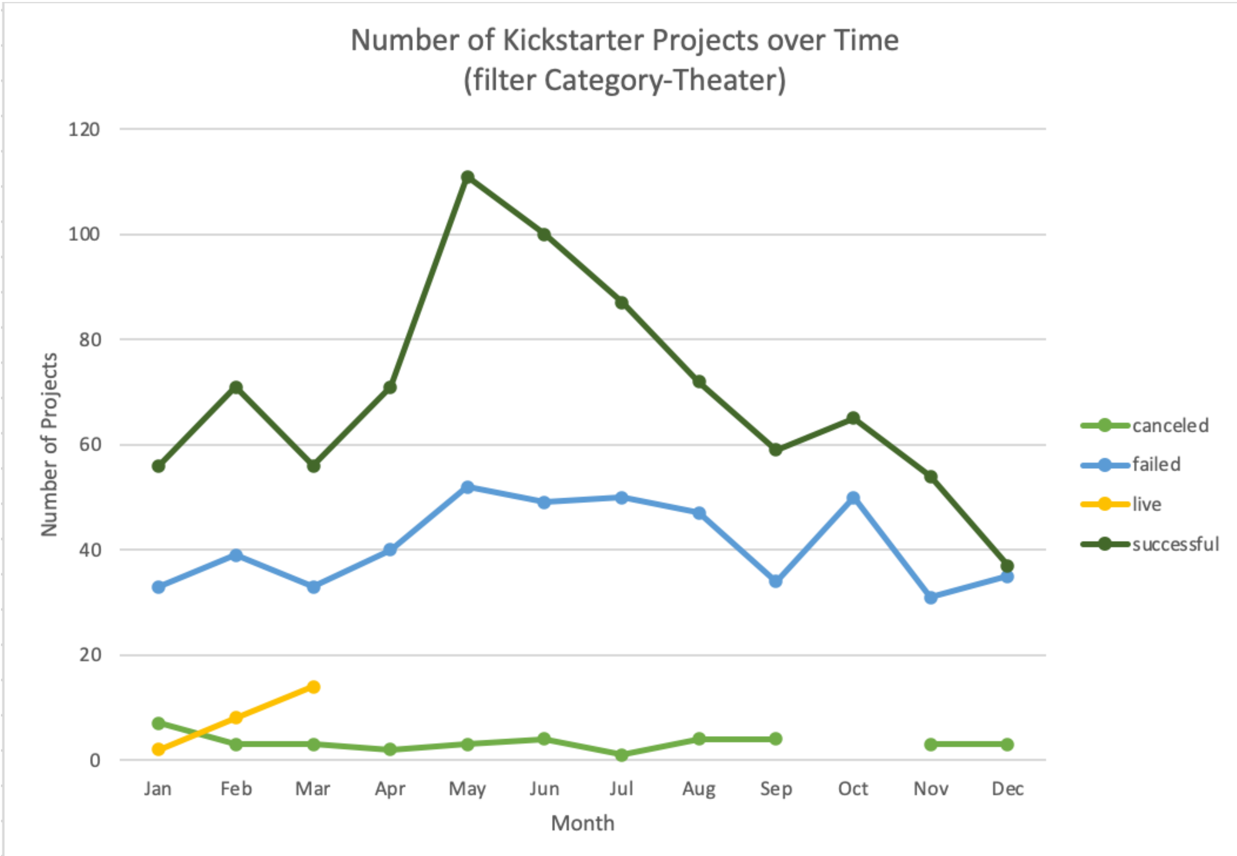

Figure 6 Number of Kickstarter Projects over Time (filter for Theater)

Based on these findings, my recommendations to ensure a successful Kickstarter campaign are to launch in May, in the United States, under the theater category. The trick to finding success in a Kickstarter campaign is to follow these trends now knowing that month, country, and category greatly impact a campaign's outcome.

These recommendations need to be taken in the context of the sample data as there are limitations. There is a bias to the theater category in terms of Staff Picks. A campaign identified as a Staff Pick is a campaign highlighted by a Kickstarter employee for no certain criteria other than preference. Therefore, staff could be leaning towards only seeking out theater projects as theater is shown as the category with the highest staff pick in this data set. The data is mostly pulled from the United States and is not typical of that of a global survey. In addition, the data set includes data from over a decade ago. Sampling campaigns from a more current time period would provide for greater accuracy. Utilizing the raw counts as the marker for analysis is a limitation within this data set. Since there is a bias in terms of the number of theater campaigns, utilizing percent successful rather than raw count would be a more accurate measurement for analysis.

Future works includes reperforming analysis utilizing success rate as a percentage rather than the raw count on a more current data set from a more representative sample, making a predictive model, and looking at regression.

Bonus Statistical Analysis

Outcome	Backers_Mean	Bakers_Median	Backers_Minimum	Backers_Minimum	Backers_Variance	Backers_Standard_Deviation
Successful	194.4251716	62	1	26457	712840.9867	844.2991098
Failed	17.70980392	4	0	1293	3773.221669	61.42655508

Table 1 Successful vs. Unsuccessful (Failed) Campaigns

As identified in Table 1, the median is a better representation of a campaign than the average because of the right-skew in the data set. The median is not as heavily influenced by the outliers as the mean is.

There is more variability within successful campaigns due to the spread of the backers. This makes sense especially when looking at standard deviation. Ninety-five percent of the successful data falls between 0 and 1688 backers, magnifying this huge spread of backers; whereas, in 95% of all failed campaigns the backer count only goes up to 122.