Kurt Dawiec

Data Analytics BC

Columbia University

Analysis and Summary of Kickstarter Data

Kickstarter is an online crowdfunding platform used by many entrepreneurs and innovators to raise capital for their ideas & inventions. This report references the data set in './Kurt-Dawiec-01-workbook' which contained metrics related to nearly 5,000 fundraising campaigns on Kickstarter. Below, I aim to answer 3 questions and supply a bonus analyses of my own summary table. The most recent entry in this set was a campaign launched on 3/15/2017.

Given the provided data, what are the three conclusions we can draw about Kickstarter campaigns?

1.) The arts dominate Kickstarter.

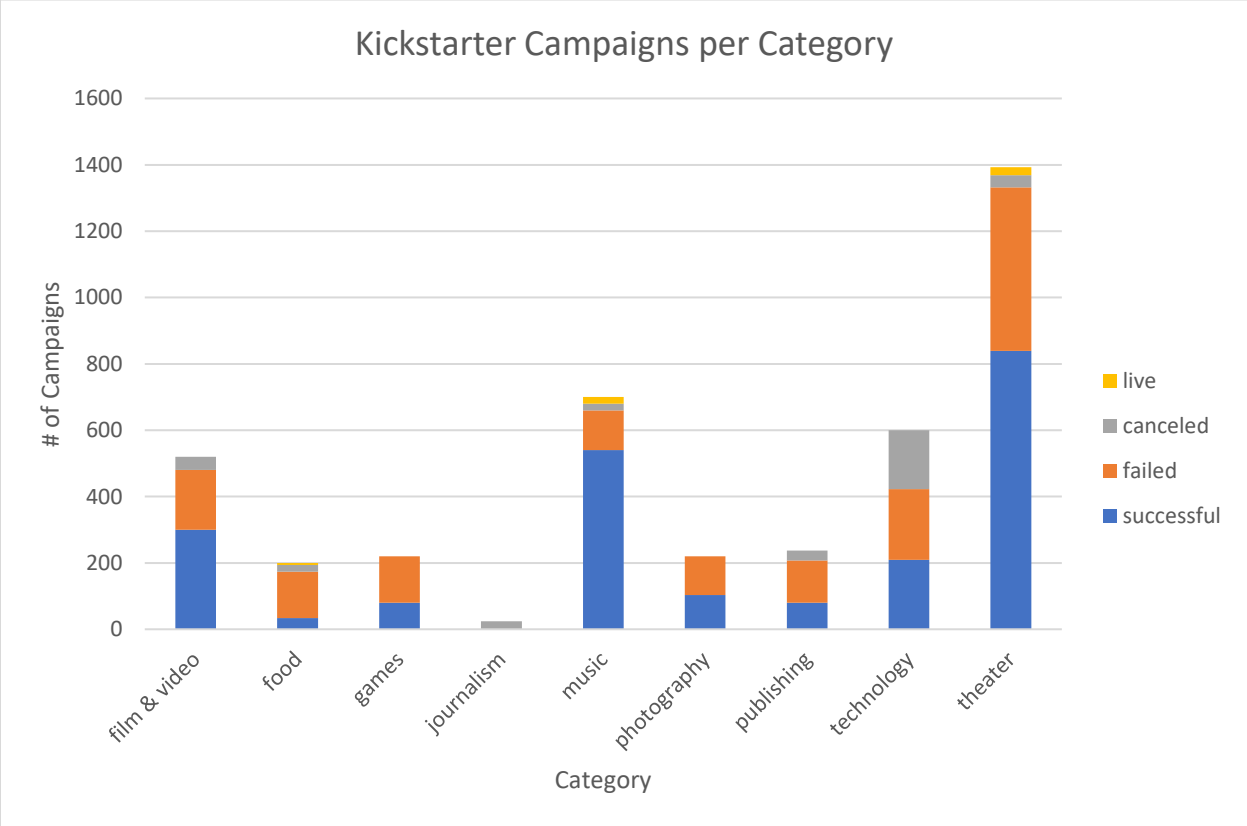

Figure 1: Stacked column chart showing the total amount of campaigns per category. Theater is followed by Music but leads by 700 campaigns.

Most campaigns on Kickstarter between 2009 and 2017 fell under the Theater category. They were most popular within the United Kingdom, and the US. This compounds the effect since 88.54% of all

Kickstarter campaigns started in the UK, or US. Coming in at 2nd and 4th are "Music" and "Film & Video." These 3 giants could be lumped into a general "Arts" category, for argumentative purposes. They are, after all, more like each other than to "Food", or "Technology."

The Arts make up the lions share of failed campaigns, but only because of owning the most traffic on Kickstarter. This is a clear indication of Kickstarter's user base and where their interests lie.

2.) **Campaign outcomes can be reasonably predicted for over 70% of sub-categories. So far, these house campaigns that always flop, or flourish.**

As of 3/17/2017, 20 out of 46 of all sub-categories contained only failed or cancelled campaigns. Additionally, 13 out of 46 sub-categories had only successful campaigns. Therefore, ~70% of these categories seem to be cursed or blessed.

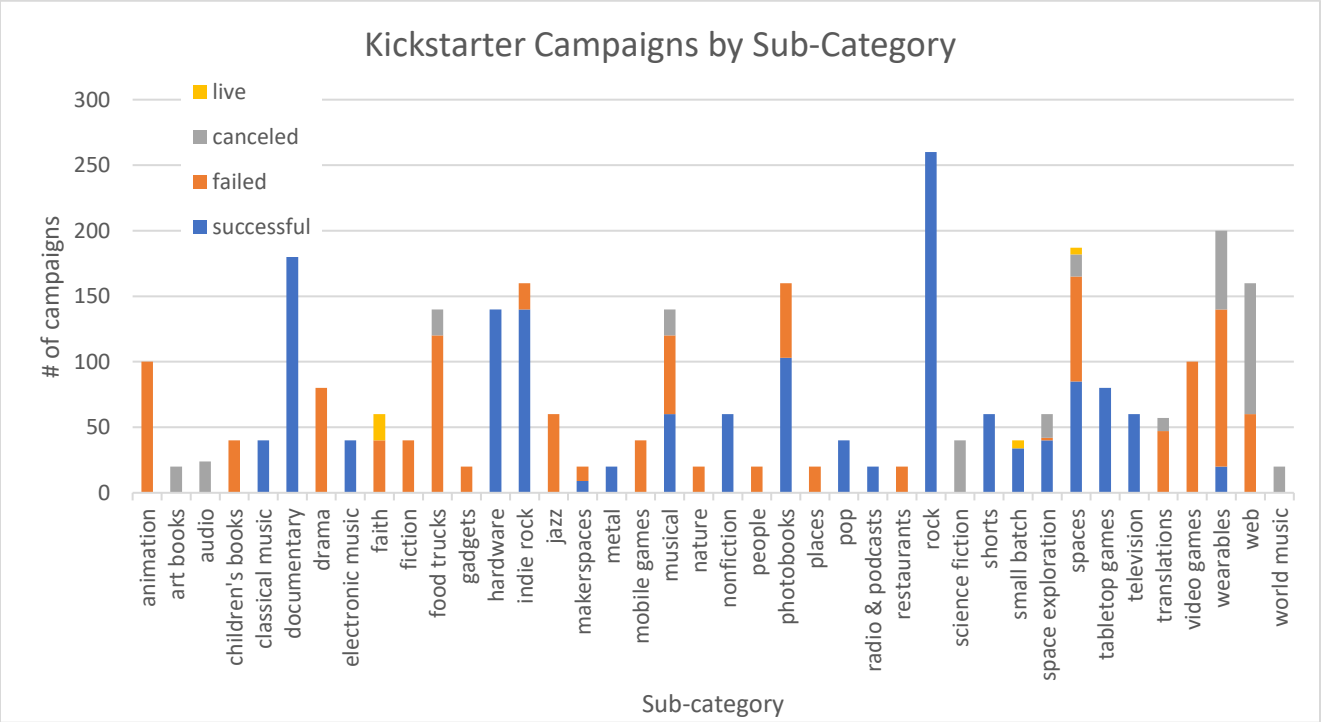

Figure 2: Stacked column chart of campaigns per sub-category. The non-uniform distribution of campaign success rates is made easier to see by omitting the "plays" category.

3.) **There is a seasonal behavior to campaign success rates.**

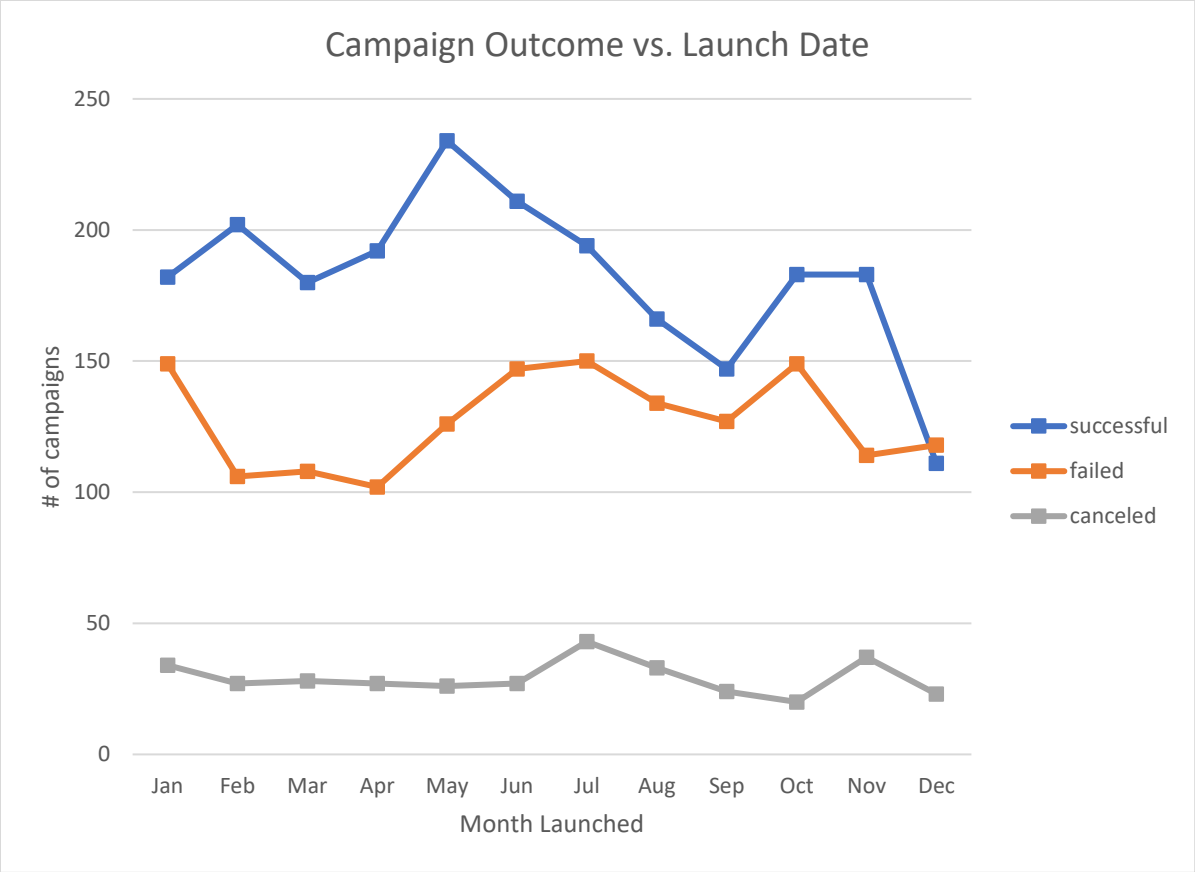

Figure 3: Campaign launch date and outcome from 2009-2017. Failed campaigns only overtake successful campaigns when they are launched in December.

December is the worst month to launch a campaign. Not only is it the month with fewest campaigns launched in total, but it is also the only month where campaigns are more likely to fail than succeed. On the other end of the spectrum, campaigns launched in May see the highest number of successful outcomes.

What are some limitations of this dataset?

Most of these assumptions are, unfortunately, weighed heavily by a lack of a large comparative sample size. It's difficult to draw patterns and conclusions about relationships between campaign categories and other metrics due to theaters dominance. Theater appears to be the bread & butter of Kickstarter, so many of the failed campaigns in other categories don't see enough repetition on the site to be able to draw substantial conclusions.

The time range of this data would also benefit from being updated. Kickstarter has likely seen continued growth since 2017. This data set would be much more powerful with latest numbers. Additionally, most of the data before 2014 is too few and far between to make use of.

What are some other possible tables/graphs we could create?

Kickstarter may benefit from tailoring content for their users based on their country of origin. A helpful graph would be the success rate of campaigns based on category for each specific country.

Additionally, it could be useful to see how quickly campaigns reach their goal. This obviously would require a new column, “date_goal_reached”, for example. This could indicate what sorts of campaigns gain momentum on the platform the fastest.

Bonus

In addition to the above, the following graph and data were generated.

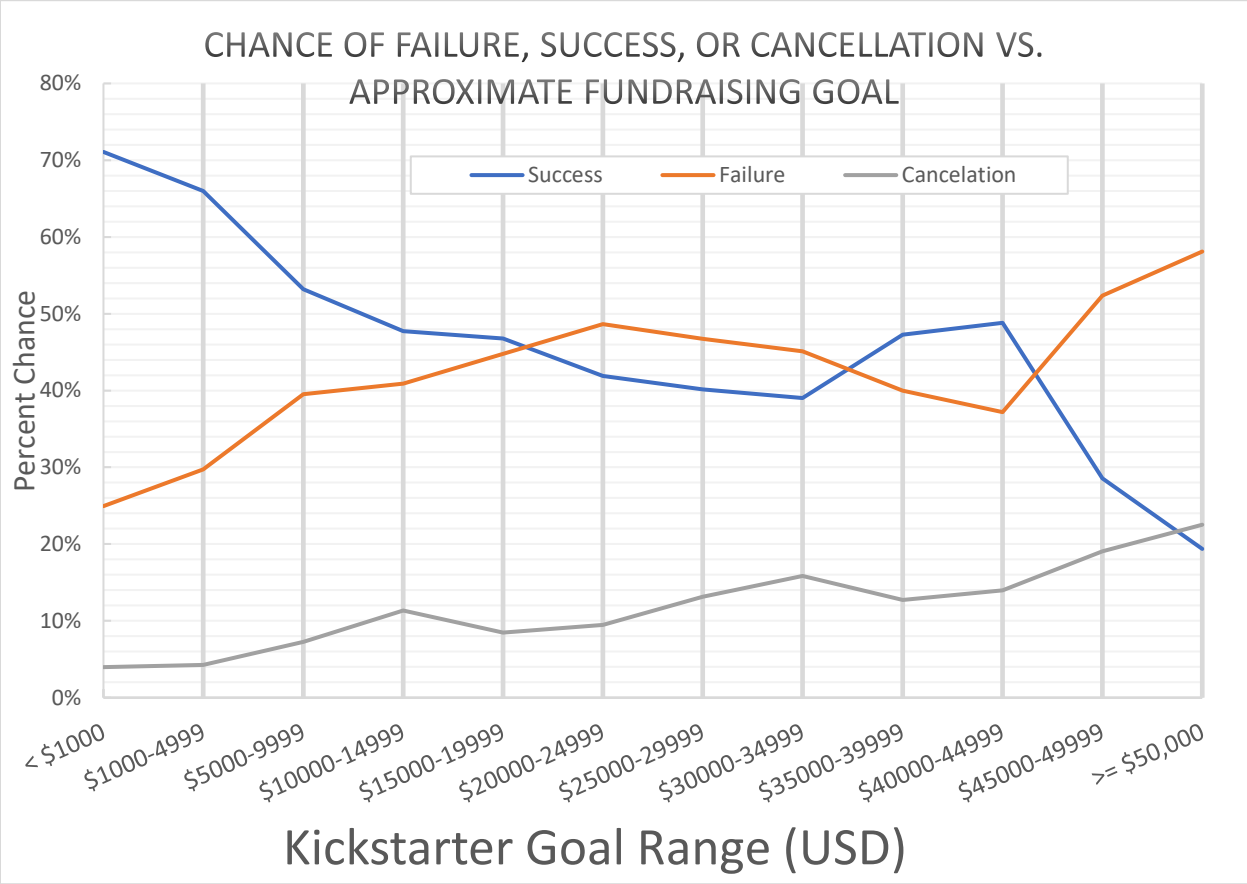

Figure 4: Outcome chances plotted against approximate campaign goal. Divergence indicated at the lowest and highest goal ranges. 3 points can be seen where the success rate and failure rate are the same. The cancellation rate seems to increase linearly with the goal amount.

Values	successful	failed
# of campaigns	2185	1530
Total Backers	424819	27096
Mean backers	194.4251716	17.70980392
Min backers	1	0
Max backers	26457	1293
Variance of backers	713167.3791	3775.689439
StdDev backers	844.4923796	61.44663896
Median backers	62	4

Table 1: Calculated values of backer counts for successful and failed campaigns.

In this case, the median amount of backers for a campaign is less helpful than the mean. The amount of backers for failed campaigns is deflated by the amount of failed campaigns with zero backers. Additionally, there are many successful campaigns with very high backer counts that deflate the median.

This also contributes to the higher variance in backer counts for successful campaigns. Successful campaigns can have few backers who made massive contributions, or many backers who made small contributions. Both of these are equally as likely, since most successful campaigns have smaller goals.