Excel Challenge: Kickstarter Analysis Douglas High

5/2021 git repo and local folder changed to KickstarterXL, file names changed, no code changes made.

Initial observation of our first pivot table shows three categories with the highest rates of success; film & video, music, and theater. People like to be entertained and so, the most successful campaigns are ones that provide such. These are the only categories with a success rate above 50%, with music standing out as the most successful at 77%. Further analysis of their subcategories reveals a dichotomy of success. Film & video is the strongest example of this; all projects involving animation or drama have failed and all science fiction projects have been cancelled, on the other hand all documentaries, television, and shorts projects have been successful.

This dichotomy of success is prevalent within each category, not just the most successful ones. The food category is the most unsuccessful, with a failure rate of 80% (70% failed, 10% canceled). These failures all come from the two subcategories, food trucks and restaurants. The remaining category, small batch, has had an 85% success rate and currently shows a 15% rate of live campaigns. Another example is the technology category; one might think that any project involving technology would have a fair chance of success. Of the six subcategories, hardware stands out with a 100% success rate, space exploration at 67%, after which the next most successful subcategory has a success rate of only 10%.

In consideration of what time of year to start a new campaign, December is the worst month, with a success rate of 44%. February, March, April, and May have been the most successful months, with rates around 60%. June and November are close behind at 55% and the rest are all relatively close to 50%. So, all in all, it doesn’t make much difference when you start a campaign. Most people have steered away from December, perhaps from the psychological perspective of not wanting to start something new at the end of the year; there is a spike in campaigns in the beginning of spring with a slow decline towards summer.

The dataset from which these extractions were made is a sampling of just over 1% of the full population of campaigns. The population consists of approximately 300,000 projects and would not be too large to analyze, a strong consideration if there was actual concern for interpretation of the data. Socioeconomical conditions could be a consideration for choosing what type of project to start, or if even to start one. This is something that analyzing metrics of the data does not show.

There are, obviously, many ways we can look at this data. Simple things like evaluating percentages as opposed to counts or choosing different charts and graphs to help us visualize information are things we can do to help notice trends. We could consider how projects were named or described in the blurb, simple metrics on perhaps lengths of such, but most probably a more manual analysis would be necessary to find any correlations in language usage to success. The spotlight field is set to true for every successful campaign and to false for every other, making it either redundant information or a seemingly guaranteed indicator for success. A staff pick of false has a fairly even chance of success but a staff pick of true yields an 87% chance of success. Originating country could play a role in success, or even perhaps the currency type.

In conclusion, statistically you would fare better from a kickstarter campaign if you choose something that entertains people. Regardless of this, you would be wise to look at the success rate of subcategories when choosing a project, as they generally indicate a strong likelihood of success or failure. Starting your project in the spring may help with success, after all spring is when things generally come to life. Personally, I’d find who these staff are that pick projects and I would see if I could run my idea past them, if someone picks it, odds are it will be successful.

Excel Challenge: Bonus Analysis Douglas High

When considering to look at the metrics of successful campaigns with relation to the number of backers, there are two things to consider. Firstly, something to consider regardless of backer relation is do we count cancelled projects as failed. We may not know why a project was cancelled and whether it is related to any variable but ultimately it was an unsuccessful campaign. Secondly, do we consider projects with no backers. I have provided four sets of metrics in my BackerData sheet; the first column shows two sets of metrics where cancelled is not considered failure, the top set considers campaigns with no backers and the bottom set only campaigns who acquired backers. The second column of metrics is the same set but considers cancelled projects as completed and unsuccessful, in other words, failures. All four of these groups show the same metrics for successful campaigns (as there were no successful campaigns without backers).

Adding cancelled campaigns as failures or removing campaigns with no backers both increase the average, as a mean, and also increase the variance and standard deviation. This is the result of both failed and cancelled groups having a large percentage of campaigns with no backers (both around 20%).

When considering whether the mean or median best describes the average for number of backers, we need to take a closer look at the breakdown of this variable. I have included some quick breakout charts showing how well centered the mean and median are for the various groupings.

With regard to successful campaigns, the mean number of backers is 194 and the median is 62. Looking specifically at these numbers, 82% are below 194 and 18% above; compared to 50% less than 62 and 49% greater (and the remaining 1% equal to 162). Obviously, in this instance, the median is a pretty accurate measure of average, whereas the mean is not. The details of the other three categories I have analyzed can be seen in the BackerData sheet but they all show the same basic statistics for mean and median. Not only is the median a better measure of ‘center’ here but it is also pretty well dead on.

With regard to variability, the variance of successful campaigns is about 100 to 200 times greater than unsuccessful campaigns, depending on whether we consider canceled campaigns and ones without backers, and the standard deviation is approximately 10 to 14 times as large. We have already seen that median is a good measure of center in this case. The median for successful campaigns is 62 and for unsuccessful they are 3,4,5, and 6, again, depending on whether we include 0’s and/or canceled.

Considering the median number of backers is so low for unsuccessful campaigns, we would not expect to find a large amount of campaigns with exceptionally larger numbers for backers. On the other hand, the median for successful campaigns is 10 to 20 times larger than the unsuccessful and when we look at the campaigns with the largest number of backers, we see a 20 fold difference there as well. Because of this we would expect to find successful campaigns having more variability.

From a logical perspective, successful campaigns are ones that investors feel good about and generally investors are looking at similar metrics when deciding what to invest in and campaigns that seem sound will seem that way to a multitude of investors, just as unsound ones will be viewed similarly and not invested in. Many successful projects could likely have a large number of investors while other successful projects may only have a few, or even just one, whereas unsuccessful projects are likely to not be invested in by many backers, creating a smaller variability of this data.