Jenna Kutz

September 16, 2023

Module 1 Challenge: Analysis of Crowdfunding Success

**Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

The first conclusion we can draw is that the most popular categories, no matter what country the crowdfunding campaign was performed in, are theater, music, and film & video. These categories represented 69.7% of the total number of campaigns conducted during this time period.

A second conclusion we can draw is that the success rates are largely similar across the categories. Six of the nine parent categories have success rates between 40 and 60%. Of the top three categories of theater, film & video, and music, the success rates are 54.4%, 56.6%, and 57.3%, respectively.

A third conclusion we can draw is that, according to the pivot table based on time (looking at the data from all years), the theater category sees spikes in number of successes in March, June, and September, with drops in May and August. The music category sees spikes in February, July, and November, with a drop in April/May and October. Finally, the film & video category sees spikes in February, August, and November, with drops in April and September. This indicates that the time of year may influence whether a campaign is successful or not, though this will depend on the category, the country, and the year, as there might be a trend in more recent years that differs from the overall trend.

**What are some limitations of this dataset?**

The top three categories (theater, music, and film & video) had relatively large sample sizes compared with the other categories, which almost all had less than 50 campaigns to analyze. For example, journalism and world music both have a 100% success rate, but each launched only 4 and 3 campaigns respectively. With such a small sample size, it’s difficult to determine whether the success rate is merely a fluke or if there was something special that made these successful.

Another limitation where the samples were collected. Using the top three categories and subcategories as an example, most of the data comes from the United States alone. If we were trying to organize a campaign in another country, such as Australia, it would be difficult to use this data analysis as a tool to inform your campaign, as there simply is not enough data specific to that part of the world included.

Additionally, we don’t know whether any of the campaigns were advertised or promoted before launch and, if so, to whom and how far in advance.

**What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

There are many more graphs and tables we could include, depending on what the most demanding questions are. Some of these include:

1. A graph summarizing the success based on the number of backers or the average donation. This would help determine whether, within a given category, there are usually only a few backers who donate, or if there is a larger group who are each able to donate a little.
2. A visualization of the percentage of successes (or failures). While theater has the highest sheer number of campaigns, its success rate is only about 54%, while the technology category holds the highest success rate at 67%. Using percentages will allow us to see whether certain categories or subcategories tend to have higher success rates, and will enable further research into why.
3. A bar graph showing the length of time of the campaign versus whether it was successful or not. Some of the campaigns are a month or two; some are only a couple of weeks. By using this graph, we could determine whether there is an ideal length of time to keep a campaign open.

**Statistical Analysis**

In this analysis, the median better summarizes the data. For both the successful and unsuccessful campaigns, the data is right skewed (the median is lower than the mean) and the variance is very high. With such unsymmetrical data, it is better to use the median as a summary statistic instead of the mean, as it is a better description of what the data looks like.