**Crowdfunding analysis and questions**

**Cindy Hansel**

**10/3/2023**

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

**Considering categories and sub-categories**

Based on created tables of ratios of successful to failed campaigns, the category that performed highest was that of photography, which had a 2.6 ratio. That category had only one sub-category, photography books, which then also showed that same ratio. However, that was not the highest performing sub-category. The three highest performing sub-categories were audio, web, and world music, scoring ratios of 4, 3, and 3, respectively. Caution must be taken when observing the low sample sizes of audio and world music, but web had a substantial sample size.

The lowest performing category, based on the same ratios, was games. Of the games categories, the sub-category of mobile games had the lower ratio of 0.5 which also was the lowest overall in successful to failure ratio consideration of sub-categories.

**Considering goals:**

There were several goal amount ranges where the failure rate was higher than the success rate. Those ranges were between $10,000 and $15,000, and above $50,000. Results suggest that success is more likely with goals between $15,000 and $25,000, and also between $30,000 and $35,000.

**Considering months/years:**

Observing the successful to failed ratio of all of the months and years, those campaigns that were kicked off in June, July, and September performed the best. June and September shared the highest ratios of 1.96. August was the lowest performing month overall, with a successful to failed ratio of 1.17.

**Considering countries:**

Though the sample sizes of other countries compared to the US are small, observations could still be made based on ratios of successful to failed campaigns. The average ratio was 1.51. China, which did have the lowest sample size, performed the most successfully with a ratio of 2. The second highest performing country was the US with a ratio of 1.59. The lowest performing country was Canada with a ratio of 1.16.

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

The sample sizes of some of the categories/sub-categories are very small and therefore do not offer much weight, or representation, to the whole dataset analysis. It is also difficult to evaluate the data by country, since the US was far above the most popular country to kickoff campaigns. It was not clear, by the title of the column, what the scope was on the country category. An assumption was made that this indicated where the campaign kicked off. It is unknown if these campaigns were limited to those countries.

Any data that would indicate incentives (product offers, company equity, etc.) offered might help to analyze the success/failure of campaigns.

Data that would indicate how the campaign was conducted (how was it “marketed” or pushed through social media) might help to understand the success/failure of campaigns.

It would be beneficial to know if there were minimum or maximum limits of donations to understand the distribution of donation. This would help to demonstrate if a campaign goal was set appropriately and perhaps give some added insight to the audience behavior/passion/interest.

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

To observe the ratios of successful campaigns to failed campaigns is beneficial. The stacked graphs gave us a good visual, however, in order to more precisely evaluate the data, the computations are helpful. Ratios and percentages are a good representation of the data because it normalizes the data. I created ratio tables and graphs for some of the data categories and was able to observe the averages of the ratios. From this, I was able to evaluate how far from the average, with numbers and color, the data category elements performed.

We could also do more statistical analysis and tables/graphs/charts on the ratios of successful to failed campaigns. This would help us determine the standard deviation and variances of this more normalized data.

Additional data would be helpful, as mentioned in the above paragraphs on data limitations.

**Use your data to determine whether the mean or median better summarizes the data.**

Of primary consideration when determining mean/median should be the objective of the analysis. In this problem, success is defined as meeting a goal. Though the data shows those cancelled or still live, the emphasis is on the binary decision of successful or failed. We are looking for a recipe for successfully meeting a pre-defined goal. If it were based on the various amounts that were actually raised by a crowdfunding campaign, this might affect this decision. For this reason, and along with data that shows a large variance and perhaps outliers, the preferred measure of central tendency would be the median.

**Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

There is more variability with successful campaigns. This makes sense because the range defined for success is unlimited (there exists a lower limit, %100 funded, but no upper limit), whereas the definition of the range for failure is contained (the lower limit is %0 funded and the upper limit is <%100 funded). The defined range for success is [%100 to infinity], and this data set shows a range from [%100 to %2338.83]. This data set shows a range for failed campaigns as [%0 to %99.68]. With the higher scope definition of successful campaigns, it is logical to find a wider range of number of backers, total amounts, etc. in the data.