Excel Challenge Report

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

By looking at the data comparing the category of crowd-funding campaigns to their outcome, we can see that theater, music, and film/video campaigns are by-far the most popular campaigns, but do not necessarily have a greater chance of success than other less popular categories like photography or publishing. Most categories seem to have more successful outcomes than failed outcomes, and this is also seen for the most popular campaign themes. This means that although certain campaign activities may seem more successful due to their popularity, they do not have any advantage over other less popular activities for successfully raising money.

It also appears that the number of successful campaigns, seems to peak in June and July, whereas the number of failed campaigns seems to drop down during these months. Therefore, in order to maximize the likelihood of success for campaigns, it seems that they should be planned for June/July (the summer months).

Finally, a comparison of the success and failure rate of campaigns based on their fundraising goals demonstrates that campaigns who aim to raise between 15000 and 45000 dollars seem to have the highest success rate and lowest failure rate (the percentage of canceled campaigns remains low across all fundraising goals). Campaigns that aim to raise below 15000 dollars have lower success rates and higher failure rates. The data does not go past 50,000 dollars and thus we cannot see how the success and failure rates trend beyond this point, but there is a steep drop down in success and an increase in failure between 45000 and 50000 that would lead me to believe that campaigns with higher fundraising goals have a greater risk of failure and lower chance of success.

1. What are some limitations of this dataset?

The data set does not convert the currencies for each country into a single currency. It maintains the original currency that the crowdfunding was done in. This is problematic for our analysis as it does not allow us to accurately compare the overall goal to the overall amount raised for successful and failed campaigns. For example, we are treating DKK and USD as the same currency and not accounting for the significant discrepancy in their overall worth – this is not an accurate comparison.

Additionally, it would be useful to see the amount of money donated by each backer on top of the number of backers. While we were able to calculate the average amount donated by each backer, this is likely not an accurate value as the number of backers was not necessarily proportional to the goal or the outcome. It would be useful to see if different types of campaigns drew in different levels of funding from backers to get a better sense of the demographics.

Finally, it would have been useful to see when the donations were made. The data set simply shows when the crowdfunding began, when it ended, and the total amount pledged. However, it is most likely that there were variations in the amount of donations made throughout this time period. Being able to see how the donations varied throughout this time might provide insight into the success or failure of the campaign.

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

I would want to do a more in-depth analysis of how crowdfunding campaigns varied within different countries. The dataset pulls information from a variety of countries like Canada, Australia, Great Brittani and Switzerland. The unique sociopolitical climate and demographics of each country would likely affect crowdfunding campaigns differently in each country, so a more narrowed focus would provide useful insight into their different dynamics. We could make a bar chart to compare how successful crowdfunding campaigns were in different countries, or make a pie chart to see which crowdfunding categories were most successful in each country.

I would also like to perform statistical analysis on the current charts to see if the differences that are observed in the success and failure rates are significant. While we can visually observe changes in success or failure, we did not statistically calculate whether these changes are significant and thus cannot make statements on whether or not different factors influenced the success of a campaign.

Finally, it would have been useful to compare the percent funded to the different fundraising factors (time of year, parent category, goal, etc). Just looking at the outcome (success vs fail) doesn’t provide a very nuanced analysis of how well the campaign did. Sometimes the campaigns did not fail by such a large margin and came close to reaching their campaign goals. The percent funded would allow us to see a scale of less successful to more successful campaigns and what factors potentially affected that. For this data, we could do make a line graph or bar chart to see how the percent funding changed based on changes in different fundraising factors.

Statistical Analysis

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

The data seems to have many outliers and doesn’t follow a normal distribution, so I believe the median would be a more accurate representation of the center.

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

The success statistics have greater variance and standard deviation than the failed statistics. This makes sense because there is a greater difference between the mean and minimum values for the success statistics than there is for the failed statistics. The fewer outliers, the less variability there would be.