Excel Challenge Report

By Shannon Lloyd

Questions to be answered:

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
* What are some limitations of this dataset?
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Conclusions:

Based on the data set crowdfunding campaigns have a higher rate of success than failure, amount of the goal set by the campaigns might influence the outcome of the campaign, and the month the campaign is created may affect the outcome of the campaign.

The overall percentage of successful campaigns in the data set was 57% and the failed campaigns were 36% with canceled and live only showing 6% and 1% respectively. From this I would consider crowd funding as a viable option for raising funds while keeping in mind that a 36% failure rate is far from an insignificant amount of risk.

The financial goals of the campaign tended to cluster within certain ranges 23% of campaigns were within the $1,000 to $4,999 range and 32% were within the $5,000 to $9,999 range. The outcomes of the campaigns were noticeably different in these ranges. The $1,000 to $4,999 range had an 83% success rate whereas the $5,000 to $9,999 range had a 52% success rate. The other goal ranges were few enough that I would not put much faith in the outcome trends. An exception to this is the over $50,000 which had 31% of the campaigns but only a 37% success rate. Taking this into consideration, I would be less inclined to consider crowdfunding unless my goal was within the $1,000 to $4,999 range.

The outcome of a campaign varies depending on the month. 62% of all campaigns started in July were successful, whereas 49% of campaigns started in June were successful. The type of campaign did vary quite a bit though. Theater projects were 65% successful in June yet only 55% successful in July. Film and video campaigns showed the greatest success in February and August. Therefore, if I were planning a campaign, I would take into consideration which month to start the campaign.

Limitations:

Some things to consider about the limitations of this dataset are, the source, sample size, and past performance is not a guarantee of future performance. This dataset was provided by UT for the purposes of this assignment. I have no visibility as to where this data came from, how it was obtained, how were the campaigns to survey chosen, or if it the results have been altered. The results of 1000 crowdfunding campaigns are likely not enough data to have a true representation of all crowdfunding campaigns. Many things can affect future outcomes. Some things to consider are the rise of skepticism because of an increase in crowdfunding scams, a tightening economy, and the possibility of new funding methods.

Other tables and Charts:

Some information that may prove useful would be looking into the length of the campaign and a little more statistical analysis. A comparison between the length of a campaign and the outcome would be useful for deciding if the length of a campaign is worth considering and if so, what is the best length of a campaign. Further statistical analysis would provide a clearer view of the outliers and refine our view of what data is most significant. I intentionally did not include staff picks and spotlights because this is an area in which the campaign would not have much control, but could be very useful for a crowdfunding business.

Statistical Analysis questions:

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

The mean and median both have their strengths and weaknesses. The low median value in comparison to the range of backer count values tells us many of the campaigns have a low backer count. The mean was also quite low in comparison to the range of backer counts. Both indicate a high likelihood of outliers on the upper end of the range of backer counts. From this I would tend towards trusting the median value to be a better estimate for the typical backer count. I created scatter plots and box and whisper charts for both successful and failed campaigns. Looking at these, there are a lot of outliers that would tend to pull the mean value higher. I think if we adjusted for the outliers, we would find that the median is a better representation of a typical campaign.

There is less variability in the unsuccessful campaigns. This makes sense to me because I would expect unsuccessful campaigns to tend to a lower backer count. If they had more backers, they might have been successful. Successful campaigns can easily have more outliers on the high side because they have a chance at greatly overachieving their goals.

Notes:

This site was very helpful in deciding how to get accurate tables for the backers count to outcome analysis.

<https://www.automateexcel.com/formulas/vlookup-display-multiple-matches-or-results/>

Under the Goal Outcomes sheet I created a separate column of range values to practice entering a mix of reference values, text, and calculated values in my first column.