Challenge 1 Excel

Kirsten Rain

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

The Arts categories, including film & video, music, and theater, have the highest success count. However, they also have the highest number of attempts and the success rate, in contrast to the success count, was not actually that high (57, 57 and 54 percent respectively). The technology category had fewer attempts and hence fewer successes, but a higher success rate than the arts.

There does not appear to be a correlation between the time of year and the number of failed or cancelled projects. There is a slight uptick in successes during the summer months. The difference in success rates between countries does not appear to be significant either.

All of the projects seem to have blurbs about technology rather than arts, food, music etc. So, either we have bad data or there is a confounding variable like the projects being tech-driven regardless of the industry or category.

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

Charisma is a variable in any campaign that is hard to measure, but a crucial factor. There are several subjective factors to consider: How compelling was their pitch? Was it thorough and well written? Are they likeable? Is race/ethnicity or social class a factor? Are there underlying variables like serving children or religious affiliation? It might not be possible to easily capture all of these, but we could capture the word-count of the essay, quantity and content of photos, and key words.

Outcomes could also depend on their social connections or affiliations. Social media connections could be driving backers. There is also correlation between charitable giving and sense of community ties (which is often attributed to religion). Our data set does not include any community ties whether they be business, school, or faith community. We could attempt to capture these with key words.

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

Given the data we have, we could also look at the relationship between country and outcomes, but there doesn’t seem to be a strong relationship there. Either we didn’t collect the right data as outlined above or we didn’t collect a large enough sample or there is something wrong with the data we did collect. If we were able to collect new data, we could include charts and graphs to analyze those other qualitative factors like scatterplot of social media followers versus percent funded.

**Include a brief and compelling justification of 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?**

There are slightly more backers in successful campaigns, but there is a lot of variability in both sets. Both data sets for number of backers for successful and failed campaigns have many large “outliers” that are pulling the mean higher than the median. For this reason, the median is a better descriptor of the data. There is more variability in the number of backers for the successful campaigns. This makes sense as there are many other qualitative and subjective factors that contribute to success.