Module 1 Challenge Written Report

January 3, 2023

John Geng

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

1. Based on the analysis of sub-categories, we can see that plays are the most popular sub-category of funds in terms funds started with more than 7x compared number of funds of other sub-categories. However, without calculating the percentage successful, we can tell from the visualization that only about half of the funds were successful which is on par or even worse than most of the other categories. From there observations, we can see a bias towards starting funds involving plays, however this bias may not be
2. From the analysis of the parent categories, we can see that theater has the greatest number of funds and it is driven by plays, the only sub-category in theater. The rest of the parent categories are also not popular with the exception of music and film & video. However, even with the two categories combined, it is still not as numerous as the top parent category theater. From these observations, we can conclude that most funds are started in categories involving art.
3. The analysis from date created and outcome tells us that number of successful outcomes of funds’ goal tend to stay at around 40-50 per month. Wit the exception of June and July when number of successful campaigns in those two months reaching 55 and 58. As observed from the line graph, there is no significant correlation between successful, failed, and canceled funds.

What are some limitations of this dataset?

This set of analysis does not tell us the whole story. The dataset directs our focus on number of successful and unsuccessful campaigns regarding the category of the fund. Although this can tell us which funds are most created and the distribution of the outcomes of these funds, it does not tell us why these outcomes occur. Another limitation of this dataset is the analysis on outcome by date, it does not really tell us anything insightful but just a count of outcomes in a particular month. Doing the same analysis with the duration of a fund could give us better insight to how successful a fund is based on time as a variable.

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

We could further our understanding of this dataset by creating tables and graphs of the percentage of outcomes of every parent and sub-category paired with percent of goal that is funded. These visualizations could provide better insight to what categories do backers care about more and how close funds are able to get to the goal.

Another graph that could potentially provide insight to this dataset is the visualization of percent funded paired with average donation. By comparing these two datasets with ranges similar to the bonus component, we can see how much each individual backers were willing to contribute to a fund and if that has any correlation how successful a fund is.