Challenge One Questions and Data Analysis

Chart, waterfall chart

Description automatically generatedThe provided data gives us a few insights into what possibly makes a successful crowdfunding campaign. By looking at the campaign outcomes by category, one can see that the most successful categories were journalism, technology, and photography. While these categories didn’t have the most campaigns, they had the highest proportion of successful campaigns to failed/canceled projects. By far, the largest number of crowdfunding projects were in the theater category.

Chart

Description automatically generatedBy looking further at subcategory data, one can see that more niche topics such as “audio” and “metal music” are more successful, while categories such as plays have a large number of projects, and thus larger numbers of successes and failures. While we can’t say for sure based only on the graphs, one could surmise that niche categories find it easier to reach a specific subset of donors, whereas more broad categories such as “plays” may find it more difficult to reach donors due to the oversaturation of campaigns in their subcategories.

Chart, line chart

Description automatically generatedFinally, by looking at the outcomes of projects by date of creation, it seems that projects created in the summer months of June and July saw more successes on average. Unfortunately, given the nature of the data, I don’t know exactly why any of these observations are the case.

While this is a very large sample to work with, there are still limitations to this data set. I don’t know about the collection process for this data, so I can’t tell if it was randomly selected or not. Additionally, we don’t have a lot of information on the creators of each project, besides their name. there are many other contributing factors to the success of a crowdfunding campaign that can be related to the existing status and wealth of the campaign organizer, which would be difficult to account for when collecting and analyzing this data.

Additional graphs or tables that may be helpful could be a graph of average percent pledged by category and subcategory, as well as a table of average goals by category, adjusted so that the type of currency is consistent. It would also be interesting to look at a stacked chart of how many projects succeeded, failed, or were cancelled based on their duration, in order to see if shorter or longer campaigns were more likely to reach their goal.

Comparison of Backers by Outcome

Chart, box and whisker chart

Description automatically generatedBy looking at side by side boxplots of the number of backers based on the outcome of a campaign, one can see that both distributions are highly skewed. This means that for both distributions, the median would be a more appropriate measure of center. In terms of variability, while both distributions have lots of outliers, the successful campaigns had a higher variance for number of backers. At first this did not make sense to me, as I thought that successful campaigns would have a high median and low variance. However, the number of backers that it takes to make a campaign successful can vary wildly based on the initial goal of the project, so this higher variance does make sense.

Works Cited

edX Boot Camps LLC, *Crowdfunding Data*