Module 1 Challenge: Crowdfunding Book

Questions

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. Looking at the first two pivot tables and charts, we can conclude that a little more than half of all campaigns were successful.
   2. For the first pivot table and chart (parent category), we can also determine that the three most successful parent categories were “film & video”, “music”, and “theater”.
   3. However, the second pivot table and chart (subcategories) show that the subcategory “plays” was far more numerous and successful than any other, making up about a third of all campaigns, regardless of parent or subcategory.
2. What are some limitations of this dataset?
   1. One limitation might be the validity of these recorded values. If I were creating an analysis for a client, I would need to be sure that this dataset is valid, but there is nothing provided that might prove it.
   2. Another limitation could be that there is an unequal number of campaigns in each parent category (total film and video campaigns numbered 178, food numbered 46, etc.), and this leads to unintentional skewing. If we were to analyze equal amounts of each parent category, it might give better insight into how well each did, especially since a single category, “theater”, out of nine provided about a third of all data. We don’t know if this data was pulled at random or not, and this might lead an analyst to believe that it wasn’t.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. We could create line graphs that plot the number of backers of each campaign against each parent category or subcategory to determine popularity. So instead of showing how many campaigns in each category were successful or not, we could illustrate how many actual backers donated to each category. This might be useful in determining popularity between categories (larger number of backers equals more popular). Because the initial tables and charts only depict whether a campaign made enough money to be successful or not, we are unable to determine if a single backer or handful of backers contributed larger amounts of money to that campaign, rather than a large group of backers contributing smaller amounts.
   2. We could also create line graphs that show the average donation amount for each parent category and subcategory. This would provide insight into another version of “popularity” between categories (more money per average donation equals more popular).
4. Is the mean or median better?
   1. In this case, the mean is better to summarize the data. Because the range of minimum and maximum number of backers is so wide (16 to 7295 for successful and 0 to 6080 for unsuccessful), the median doesn’t consider how spread apart some values might be from each other. For example, the median of the successful campaigns is 201, showing that half of these campaigns had backers equal to or less than 201, but does not really show the campaigns that had backers in the thousands. The mean, however, does, which is depicted in the larger value for both successful and unsuccessful campaigns.
5. More variability with successful or unsuccessful campaigns?
   1. There is more variability with the successful category, most likely because the range of the number of backers is far wider than the unsuccessful category. The standard deviation for the unsuccessful category is also far less than the successful, which supports this finding.