Module 1 Challenge

Aliha Ahmed

May 4, 2023

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

* Of all parent categories, the Theater category has the highest number of crowdfunding campaigns, with 344 campaigns out of a total of 1000 in the provided dataset. As such, the Theater category also has the highest number of failed campaigns (132 out of 364 total failed campaigns) and the highest number of successful campaigns (187 out of 565 total successful campaigns)
* On a relative basis, the Journalism parent category has lowest proportion of failed campaign outcomes, with 0% of all 4 Journalism campaigns failing. The Photography category has the second lowest proportion of failed campaign outcomes, with 26.19% of the 42 Photography campaigns failing, and the Technology category has the third lowest failure rate (29.17%). On the contrary, the Games category has the highest proportion of failed outcomes (47.92%) and thus the lowest proportion of successful outcomes (43.75%). Naturally, the Journalism category had the highest proportion of successful campaign outcomes, with 100% of all 4 Journalism campaigns succeeding (4 out of 4).
* The line graph which illustrates the relationship between campaign outcomes and the month in which they are created indicates that there is a potential relationship between the two variables. As shown in the line graph and its associated pivot table, there appears to be a sharp uptick in the number of successful campaigns per month in the months of June and July (55 and 58 successful campaigns respectively), with a range of 41-49 successful outcomes per month in every other month of the year. With further research to prove out this linkage, researchers may have something to say about when crowdfunding campaigns/efforts are most effective and should be targeted.

1. What are some limitations of this dataset?

* Sample size: With a larger data set, the conclusions which we are able to draw will be more accurate and refined. Furthermore, having data across other campaign categories (e.g. healthcare/medical campaigns) can help draw further conclusions regarding which category of crowdfunding campaigns are most successful and resonate the most with backers/donors.
* Lack of data on type of funding: There are several types of crowdfunding models, including rewards, donation, debt and equity. Having data on the type of crowdfunding model utilized can help draw conclusions regarding which models most effectively incentivize backers/donors to support campaigns.
* Lack of demographic data on donors/backers: Possessing demographic data on the donors that are backing these crowdfunding campaigns (e.g. age, gender, income/socioeconomic status) can help determine key demographic attributes associated with individuals who are most likely to back/donate to crowdfunding campaigns. Through this, one can create a more compelling story around their campaign and target/engage with the demographic that is more likely to resonate with and pledge to their cause.
* Lack of engagement/investment data on campaigners: Having data on variables such as a campaigner’s advertising budget and amount of engagement with target audiences (e.g. number of channels/outlets utilized for engagement) can help draw conclusions about the quality, as well as the quantity, of efforts required for successful campaign outcomes.

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

* Country/geographical analysis: Creating a pivot table and stacked-column pivot chart of campaign outcomes by country can help draw conclusions regarding which regions tend to see a higher proportion of successful outcomes, which can then allow us to theorize why this may be the case.
* A scatterplot demonstrating the relationship between goal sizes and average donation made: If a trend line is visible, we can draw conclusions regarding whether the average donation size increases or decreases as campaign goal sizes increase. Thus, we may be able to determine whether there is an ideal/target campaign goal range which may receive larger donations on average.

**Statistical Analysis**

1. Use your data to determine whether the mean or the median better summarizes the data.

The summary statistics calculated for both successful and unsuccessful campaigns show that the mean is significantly greater than the median in both cases. As such, the distribution of data is positively skewed for both successful and unsuccessful campaigns. Applying the SKEW.P Excel formula, which is used to calculate the skewness of a given dataset, to the number of backers of successful and unsuccessful campaigns demonstrates this result as well. This formula retrieved a value of 2.17 for successful campaigns and 2.69 for failed campaigns, confirming the positive skew observed from the mean being greater than the median in both cases. As such, the median better summarizes the data and should be the preferred measure of central tendency in this case. Given the skewness of the data, the mean is more likely to be manipulated/influenced by extreme values/outliers in the data. The median is more resistant to these outliers and provides a more useful measure of central tendency.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

The variance and standard deviation are both higher for the number of backers of successful campaigns. Given there are more successful outcomes (566) than failed outcomes (365), the larger number of data points for successful outcomes may be introducing more variability in the data.