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

Given that the mean and median backer counts are higher for successful campaigns vs. failed campaigns (266 or 45% higher and 87 or 76% higher respectively), we can conclude that the higher the quantity of backers you gain, the higher the likelihood of success becomes.

Looking at the average donation amount, the difference in the mean and median is relatively small (5 or 8% higher and 4 or 6% higher respectively). Therefore, we can conclude that the average donation amount typically doesn’t differ greatly between successful and failed campaigns and is not an indicator of whether a campaign will or won’t be successful. This helps to reenforce the previous conclusion about backer count being a bigger driver of success: if the average donation amount is consistent, then the real driver of funding will naturally be the quantity of backers.

When looking at just the successful and failed campaigns by parent category, Chart 1 below shows the data as a % of the total, sorted from highest % successful to lowest % successful. At first glance, this shows that journalism is the winning category with 100% of the campaigns being successful. However, when you look at Chart 2 which reflects the data as the count of campaigns that are failed and successful listed in the same order as Chart 1 (sorted from highest to lowest % successful), it becomes apparent that there are very few journalism campaigns (only 4!) compared to the total count of all categories of campaigns (929) – this category represents only 0.4% of the total number of campaigns. Since the sample group is so small, I would argue that the technology category is perhaps the most meaningfully successful category as it makes up roughly 1/10 of the total number of campaigns and has the highest % successful once the sample size reaches that amount.

A graph of blue and orange bars

Description automatically generated

Chart 1: Successful vs Failed Campaigns as a Percentage of the Count of Campaigns by Parent Category

A graph of blue and orange bars

Description automatically generated

Chart 2: Successful vs Failed Campaigns as a Count of Campaigns by Parent Category

## What are some limitations of this dataset?

Since there is no cap to how far over the goal a campaign's donations can go, this results in there being several campaigns that are outliers in this regard. You would have to limit the definition of what "success" means in order to narrow in the data to be more meaningful and less varied when looking at what the drivers for success would be for the more typical campaign. For example, you could look at the data in pieces that you would need to define - e.g. comparing unsuccessful campaigns (not meeting their goal, <100%) vs low-level successful campaigns (100%-110% of the goal), mid-level successful campaigns (111%-200%) and high-level successful campaigns (>200%). The aforementioned breakpoints are an arbitrary example; a method for determining reasonable break points for something like this would need to be thought through. Alternately, outliers could simply be ignored.

There is no measure of the quality of the campaign from a marketing perspective. For example, did the campaign just send the funding link to their family and friends and then rely on word of mouth from there? Or did they advertise on social media, reach out to corporate donors, etc. Essentially, how aggressive were they in getting the word out – to whom, how often, did they make it to their target audience, etc. Subjectively, I believe this is likely to be the key differentiator for the campaigns that went well over 100% of their goal. This study would benefit from quantifying these “softer” qualities of the campaigns and then analyzing those metrics.

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

The “staff\_pick” and “spotlight” columns reflect should be looked at in order to see if the value of “TRUE” results in a meaningful impact to the likelihood of campaign success. This could also similarly be compared to the number of backers.

The length of the campaign is another factor that could impact success. We could chart the number of calendar days the campaign was live versus whether or not it was successful to see if there is a difference. Similarly, we could look at the total number of backers versus the number of calendar days the campaign was live to see if leaving it open longer allows for the backer counts to meaningfully increase.

In addition to the above two suggestions, you will see when reviewing the excel file used with this report that I ran several tabs in addition to those that were requested to check additional aspects of the data set. I used these additional tabs to draw conclusions about the data discussed in the first section of this document. The required tabs are highlighted yellow in order to differentiate them from the extra tabs I ran for this module challenge.