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

Three conclusions drawn from this set of data can be drawn if we look at all the data and graphs that we have assembled. First, from the entire dataset, Having a high backer count does not guarantee having successful crowdfunding campaigns. In fact, from this dataset there is a trend that occurs often where the campaigns that were highly backed were not always successful as they were either failed or canceled. The second conclusion from this dataset is that Theatre crowdfunding campaigns had the highest number of successful campaigns out of all the categories. This is interesting because the Theatre category is also the most numerous when it comes to campaigns and this most likely played a role in the big difference between successful campaigns in comparison to the other campaign categories. The third observation that I noticed about crowdfunding campaigns was that the Technology industry had the highest success rate based on the total number of successful campaigns in relation to the grand total of Technology campaigns. In comparison to other categories, Technology had the highest success rate for crowdfunding campaigns.

1. What are some limitations of this dataset?

Some limitations of this data set include how external factors such as industry growth and the country where the campaign raised funds are not entirely measured. The effects on crowdfunding due to industry growth and the locations of campaign crowdfunding are not present from this dataset. Location and industry growth have large roles in the overall success of crowdfunding campaigns and could lead to general misunderstandings about crowdfunding in different industries. Another limitation of this dataset is that the journalism category had only 4 campaigns in the entire dataset. This sample size for the Journalism category is not nearly large enough to gather any significant conclusions that could help us learn more about crowdfunding campaigns in his category. The journalism category could be considered an outlier as it is nowhere near the mean of the dataset.

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

An additional chart that we could create that could help us understand the trends of crowdfunding in this dataset would be a Pie Chart. This chart would be more valuable when analyzing the outcome and success rate of campaigns in comparison to other categories instead of bar graphs. A pie chart would give people a better understanding on how much of an effect each sub/parent category had on the overall dataset.

Statistical Analysis

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

Using the data, I feel that the mean is what we can use to summarize the data that we have gathered best. The mean of financial backers for successful campaigns was significantly higher at 851.15 than failed campaigns at 585.62. This is compelling because this means that successful campaigns had much higher support numbers on average than failed campaigns. Observing the data allows us to see that higher means of financial backers may be associated with a higher count of successful campaigns. This may not allow us entirely to conclude that having higher backer counts will always lead to more successful outcomes as there are failed campaigns with high numbers of support. However, the minimum count of backers a successful campaign had was 16.

Failed campaigns had a minimum of 0 and there were plenty of campaigns that had less than 10 backers. Overall the average counts of support successful campaigns received over the entire subset of successful outcomes allows me to conclude that a high backer count on average definitely contributed at least at a minute level towards having a successful campaign. On the other hand, It is more likely that a high average for backer count in tandem with other external factors such as country of operation or category is more likely to impact success.

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

Using the data, there is more variability with successful campaigns than unsuccessful campaigns. This means that successful campaigns are on average farther from the mean than failed campaigns are. This makes sense because the data of successful campaigns for each cell contains data that is usually vastly different from the mean. Also the number of successful outcomes is larger than failed outcomes and this affects our statistical analysis when calculating the mean and variance.