Daniel Rose

Module 1 Challenge

**Conclusions and Further Examination**

One conclusion we can make from this data is that worldwide, Crowdfunding campaigns are more likely to succeed than fail. Out of the 1000 campaigns analyzed, 56.5% of projects were successfully funded. Great Britain had the highest nationwide success rate, with over 58% of British projects being funded. Canada had the lowest success rate of all countries analyzed, with only 50% of their projects successfully funded. The United States was the most popular origin for crowdfunding campaigns, with over 76% of all campaigns analyzed originating from the US.

Another conclusion we can make from this data is the most popular month to start a crowdfunding campaign is July. July has also been the second most successful month to start a campaign, with only June seeing a higher percentage of campaigns funded. The least successful months to run a campaign were December and January, with 52% and 53% of projects funded in those months respectively. This could be due to people having less disposable income to give around the holidays, but this hypothesis would need to be researched further.

The arts are the most popular type of venture to attempt to fund via crowdsourcing, with 74% of campaigns falling under either the Theater, Music, Film & Video, or Photography categories. Specifically, Plays are the most popular venture, with over 34% of all crowdfunding campaigns belonging to that subcategory alone. The campaigns that achieved the highest success rates at funding their ventures were Technology campaigns, with Web (71% success) and Wearables (62% success) being the first and second most successful projects across all subcategories where over 40 ventures were measured in the data.

When comparing campaigns across categories and subcategories, however, we encounter some of the limitations of this dataset. Though 1000 entries of data are substantial enough to draw conclusions across all crowdfunding campaigns, when we separate the campaigns down to category, subcategory, or by month, it is harder to draw conclusions, as there are less examples per group. For example, while this data shows that crowdfunded Journalism campaigns are 100% successful, this conclusion could be misleading as there are only 4 examples.

While we have determined the Average Donation per campaign by comparing the total amount pledged to the campaign to the total amount of backers, it is possible that this Average Donation total could be misleading if we are trying to determine the behavior of an average backer. We only have the total amount pledged; we do not have the actual data sorted by individual donations per campaign, so we cannot know whether the distribution of data is normal, or if the average backer contributes a smaller donation, and the Average Donation data is being artificially boosted by outliers.

There are many other ways we can compare the data to examine this dataset even further that we haven’t yet created. We could generate a table using the Date Created and Date Ended columns to study the relationship of campaign length vs. campaign success. It could also be useful to use the Average Donation column that we created to attempt to measure which categories/subcategories receive the largest average donations, or to compare in what seasons the largest contributions are made.

Some graphs we could create for this dataset could be a bar graph showing success rate across categories/subcategories. It could also be helpful to have a pie graph filterable per category/subcategory that shows the relationship between funded/canceled/failed campaigns by percentage. If we wanted to study the how effective the Spotlight or Staff Pick designations are to helping boost a campaign, we could create a graph comparing the success rates of Spotlighted campaigns, Staff Pick campaigns, and campaigns that did not receive either designation.

**Statistical Analysis**

When examining the relationship between backers of successful campaigns versus failed campaigns, the data is best summarized by the median. The median is best in this instance because the data is not normally distributed. When this data is examined with a box plot, this data is quite skewed and contains many outliers, which would skew the mean and make it less useful when trying to analyze the data.

In terms of variability, there is a higher variation of results with successful campaigns in comparison with failed campaigns. Logically this makes sense, as it is more likely to have a successful campaign that is extremely popular, producing an outlier result and more variation, than it would be to have a popular campaign that failed.