**Written Report – Crowdfunding Conclusions and Evaluation**

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
* What are some limitations of this dataset?
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Given the provided data on crowdfunding campaigns there are several conclusions one could draw. The first can be drawn based on the pivot table and pivot chart which used the count of the outcomes of a campaign based on the parent category of that campaign. According to the data, the highest number of campaigns were in the film & video, music, and theater categories. Theater had the most campaigns of any parent category and had the highest number of successful campaigns. However, compared to the other categories with higher overall counts of campaigns, there was not a significant difference in the percentage of theater campaigns that were successful compared to other categories. So, one cannot conclude that there is a higher probability of a theater campaign being funded than a campaign in other categories. One can only conclude that of the campaigns being crowdfunded, the highest number fell into the theater category.

Another conclusion that could be drawn has to do with the Date Chart sheet, with the pivot table and pivot chart based on the month a campaign began in and if it was canceled, failed, or successful. The lowest counts of the outcome of a campaign were if the campaign was canceled. This is true for all the months in which campaigns started. Based on the data, one can conclude that regardless of the month in which the campaign began, the least likely outcome of the campaign would be that the campaign was canceled.

Finally, based on the Date Chart sheet, we can draw a conclusion about campaigns started during different months of the year. The biggest difference in the count of campaigns if the outcome was successful was between May (46 successful campaigns) and June (55 successful campaigns). The biggest difference in count of campaigns if the outcome was successful was from January (36 failed campaigns) to February (28 failed campaigns). These differences are not significant and do not occur during the same months. This suggests that the month in which the campaign began did not affect the outcome of the campaign.

One limitation to this data set is that many of the campaigns occurred over time periods of varying lengths. Because of the differences in time periods of the campaigns, the sample is not representative of the issues that campaigns may have faced if they were run for a longer or shorter amount of time to another campaign. Another limitation to the data set is that the campaigns were run in multiple countries and the amount of money collected was in different forms of currency. This is a limitation because it adds to the variability and wide scope of the data that makes it more difficult to isolate variables to compare.

A possible table that could be created in addition to what was created during this analysis would be a table and chart summarizing the number of campaigns that took place in different countries. This would add value because it would visually display where in the world the campaigns took place. This is information that was collected but not analyzed by the charts and tables that were created in the analysis. Another table that would be useful would summarize the count of backers and the average donation by category. This data in this table could be put into a chart that visualized the difference in average donation and/or number of backers by type of campaign. This would be valuable information for someone looking to understand the relationship between the number of backers and the average donations given by those backers as it relates to the type of campaign they are backing.

**Statistical Analysis – Mean/Median Justification**

* Use your data to determine whether the mean or the median better summarizes the data.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

Based on the data provided, the median is a better summary statistic than the mean. This range does not fall on a normal distribution, because the median is lower than the mean in both the successful and unsuccessful campaigns instead of being the same or a similar value. This suggests there are outliers at the maximum side of the range that have skewed the distribution and resulted in a larger mean. Given that skew, the median is a better summary statistic because it better represents the middle of the dataset.

According to the data and the variance statistics, there is more variability in successful campaigns. This makes sense because there was a larger range in the number of backers (16 backers minimum to 7295 backers maximum in successful campaigns versus 0 backers minimum to 6080 backers maximum in unsuccessful campaigns). The mean of the successful campaigns was also further from the median of successful campaigns on the distribution than was the mean from the median in unsuccessful campaigns. Based on these differences, it makes sense that there would be more variability in successful campaigns than in unsuccessful campaigns.