Module 1 Challenge

**Written Report**

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

Overall, the majority of crowdfunding campaigns are related to theater, more specifically plays. The “theater” category accounted for both the most “successful” and the most “failed” campaigns, however this was also the largest category overall, nearly double the second-largest category “film & video”. The smallest category of campaigns is “journalism” with only 4 campaigns launched, although it should be noted that all of these campaigns were successful. The time of year that a campaign is launched appears to have some affect on the outcome of the campaign; the months of June and July have the highest number of “successful” campaigns launched, with relatively standard rates of “failed”, “live”, and “cancelled” campaigns launched in the same months. Additionally, the total goal amount for the campaign appears to have a correlation with the outcome for the same campaign. In campaigns where the goal amount was “15000 to 19999”, “20000 to 24999”, and “30000 to 34999”, 100% of campaigns were successful. Additionally, for campaigns where the goal amount was “25000 to 24999” the success rate was also relatively high with 79% successful campaigns. Only for two goal amount categories was the “failed” campaign percentage higher than the “successful” campaign percentage: “10000 to 14999” and “greater than or equal to 50000”.

* + What are some limitations of this dataset?

A significant limitation of this data is the fact that this data is from several different countries around the world. While this does give us a more global perspective at crowdfunding trends, the “goal” amounts are not listed under one standardized currency; there are seven different currency types included within the data set. As different currencies have different exchange rates, this can significantly alter the trends in data and ultimately does not properly represent the data set in a standardized manner. Additionally, the dataset includes a significant proportion of campaigns from the “theater” category, specifically the “plays” subcategory, and therefore this category may contribute to an over representation of data from this category, skewing the data related to other categories. This dataset also only gives the total amount “pledged”, allowing us to calculate the “average Donation” per backer, but it does not identify if there are any significant outliers in the pledged data, such as instances where one backer pledged a significant amount to the campaign leading to it’s success.

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

Other tables and/or graphs to create would include a pivot table (and a subsequent line chart) that analyzes the “outcome” of the crowdfunding campaign based on the number of backers for the campaign. This would be best done by creating categories of a range of backers (0-5, 6-10, etc.) and using the COUNTIF function similar to the Goal Analysis done in this assignment. This would help provide insight if there are any trends in the number of backers required in order to run a successful campaign. Similarly, a table and graph related to the “Average Donation” (and a subsequent line chart) would allow for campaign organizers to identify the ideal donation amount, and potentially therefore goal amount, if they know the approximate size of the population they will have backing their campaign.

**Statistical Analysis**

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

In both the “successful” and “unsuccessful” campaign data, the median is a better summarization of the “backers\_count” data. The measures of variability, the variance and standard deviation, for both sets of data are very large, and even larger than the mean. The large variability in both sets of data indicates that there are potentially significant outliers which could skew the mean, meaning that the median is a more appropriate summarization of the dataset itself.

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

The measures of variability, variance and standard deviation, are both lower in the unsuccessful “failed” campaign data compared to the “successful data” for the “backers\_count” category. Logically this makes sense as the “failed” campaigns are likely to have fewer backers, which is ultimately why the campaign was unsuccessful. Because these campaigns typically have fewer backers, then the data will be closer together and have a lower variance overall. Alternatively, the “successful” campaigns could have any number of backers, and likely have more backers, contributing to a greater variability in the data overall.