**Module 1 Challenge: Crowding Funding**

**General Questions**

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

With the data provided, we can draw the following three conclusions:

* Overall, there were a total of 1000 campaigns of which 565 were successful, 60 failed, 11 canceled and 5 were live. The success rate was 57% and 61% if compared to failed campaigns with canceled and live campaigns excluded. In looking at the funding for successful campaigns, 270 raised between 100% and 200% of their goal while 295 raised 200% or more of their goal. Amazingly, 24 raised over 1000% of their goal.
* The category with the most campaigns was theater with 344 followed by film & video with 178 and music with 175. These campaigns represented 34%, 18% and 18% of all campaigns respectively. Journalism had the fewest campaigns with 4 (less than 1% of all campaigns) followed by photography with 42 (4%), games with 48 (5%) and film & video with 46 (5%).
* The largest number of campaigns (93) were launched in July followed by January (91), June (86) and March (86). June had the highest success rate (64%) with 55 successful campaigns, 28 failed and 3 canceled, while April has the lowest success rate (49%) with 41 successful campaigns, 35 failed, and 8 canceled.

*What are some limitations of this dataset?*

One of the main limitations is the relatively small sample size of campaigns. With a dataset of 1,000 campaigns with 9 categories and 24 subcategories, there is limited data to analyze the success or failure of campaigns based on various factors including the category, sub-category, month, year and country. As an example, there were a total of 4 journalism campaigns and the average number of campaigns for all sub-categories excluding plays was 29. Another limitation is information regarding the backers such as the date and amount pledged, demographics, and information on why they backed the project. This would provide valuable information to analyze the success of a campaign based on individual pledges and attributes of backers.

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

The following tables and graphs would be useful for further analysis:

* A duplication of the existing 3 charts and pivots calculating the % of the campaigns for each outcome instead of the number of campaigns. This was calculated manually for the above conclusions and would give more insight into the success and failure rates for the different categories, sub-categories and months.
* Analysis of the outcome of a campaign based on the length of the campaign – This would require an additional column to calculate the length of a campaign by subtracting the created/launched date from the end date. This would help assess the potential impact of the length of a campaign on its success or failure. Additionally, you could calculate the average donation per day of each campaign and compare it to the campaign’s outcome.
* A box and whiskers chart to compare the amount pledged vs the goal for successful and failed campaigns. This would delve deeper into their relationship beyond the calculation of the “Percent Funded” column that was added on the data worksheet.
* Further analysis of the outcome based on the staff pick and spotlight to see if any of these are an indicator of a successful or failed campaign.

**Statistical Analysis**

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

In general, the median seems to best summarize the data given it’s skewed by some successful and failed campaigns with a significantly large number of backers. This increases and skews the mean. For successful campaigns, this is clearly indicated by a median of 201 with a minimum of 0 and maximum of 7295. This can be clearly demonstrated by creating a box and whiskers chart for both campaign outcomes. Both are skewed with many potential outliers. That said, there are situations in which the mean may be more useful in summarizing skewed data. It often depends on the question being investigated and answered.

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

There is more variability with successful campaigns given the standard deviation for these campaigns is 1266.24 and is greater than the standard deviation for unsuccessful campaigns which is 959.99. This makes sense when analyzing the other measures. One, the maximum of successful campaigns (7295) is significantly larger than that of failed campaigns (6080) while the minimum of each differs by 16. This alone represents a larger spread of data. Two, the difference between the mean and the median is greater for successful campaigns (851.25 – 201 = 650.25) than for failed campaigns (585.62 – 114.5 = 471.12). This is an indication of more heavily skewed data and would tend to indicate more or larger data points above the mean. Thus, this indicates more variability and a larger standard deviation from the mean.