Homework 1 – Scott Sanders

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

1. Theater, film & video, and music are the most common fundraiser categories.
2. Tech fundraisers have the highest likelihood of success.
3. The number of successful campaigns peaks in June and July.

**What are some limitations of this dataset?**

The data says nothing about the size and age of organizations meaning we cannot see if larger companies have a greater chance of success. Furthermore, there is no information on why certain campaigns were cancelled. Data showing when funds were collected during the campaign would also be helpful in showing when most of the funds are pledged.

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

Tables should be created to show success and failures based on country, campaign length, and whether the campaign was a staff pick or spotlight campaign. Analyzing these factors with pivot tables and stacked bar charts (since we have categorical data) can help to further reveal which factors are important in making a campaign successful.

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

Both data sets for ‘failed’ and ‘successful’ campaigns are heavily right skewed as indicated by the histograms created on the excel sheet. Most campaigns have fewer than 500 backers, but outlier campaigns with thousands of backers exist in both data sets. Since the mean is heavily affected by outliers (non-resistant measure), the median should be reported as the measure of central tendency.

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

The backer count for successful campaigns has a greater variance and standard deviation than that of the failed campaigns. This makes sense as the maximum value of successful campaign backers is greater than that of the failed indicating a greater range. Also, there appear to be more high-number outliers in the successful campaign backers count which makes sense as successful campaigns are more likely to have many backers. Failed campaigns are more likely to have fewer backers meaning the data is more clustered around the left end of the histogram.