**Week 1 Challenge – Crowdfunding**

**Data Analysis**

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

1. Looking at the “Stacked Outcome”, it’s easy to identify that the “theater” campaign outnumbered all other categories with 344 in total, followed by “film & video” (178) and “music” (175). This indicates that these three categories are fairly popular when it comes to crowdfunding campaigns.
2. In terms of the campaign performance, looking at the “success/failure” ratio, overall achieved 1.55 with “photography” the highest, 2.36, and “games” the lowest, 0.91. Taking a deeper look at the “theater” category, which only has one sub-category, “plays”, though it has the greatest number of campaigns, its “success/failure” is 1.42, below the overall value. Another consideration of the performance is “success rate”, Successful/Grand Total, but considering the size of each category campaign, “technology” and “photography” outperformed others. We can find that these two categories were the most effective ones
3. Based on the summary of “outcome by years”, it demonstrates that June and July of each year achieved more successful campaigns, compared to other months. In this case, campaign runners can consider launching more campaigns around late spring and summer time.

* What are some limitations of this dataset?

1. We have only 1000 data points for 7 countries, which all are developed countries. This does not provide enough evidence to draw any general/concrete conclusions that can be applied to Crowdfunding campaigns. Sample size/dataset (can include developing countries’ date as well) should be further increased.
2. To better analyze the success/effectiveness of each campaign, more external factors should be considered, i.e. any tools or methods that campaign runners leveraged, social stability, culture, economy growth

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

1. Histogram/Boxplot: adding histogram or boxplot for our current campaigns can better illustrate the distribution of the whole dataset, allowing us to better understand the mean, median and outliers in one chart.
2. Pie chart for each country: adding pie chart for each country with “campaign outcome by categories” to better understand the portion of successful cases in different categories for each country.

**Statistical Analysis**

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

Comparing the mean and median, the median would be a better fit for summarizing the data. As the “Side-By-Side Boxplot” and Skewness indicate, both “Successful” and “Failed” boxplots are right-skewed (skewness>0), meaning that both “means” are largely affected by a few outliers and the whole distribution is pulled out to the right. There are more data points at the left side of the distribution. That also explains why the mean is largely greater than the median. In this case, the median represents most data points better.

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

According to the calculation of variance and standard deviation, there is more variability with successful campaigns. But that does not fully make sense judging from the size difference of 2 samples. Based on the “Crowdfunding Goal Analysis”, there are 565 successful campaigns (each campaign with different number of backers) whereas the number of failed campaigns is way less, 364. With larger sample size, successful campaigns’ variability is highly enlarged. To better compare the variability between the two groups, the unsuccessful campaign’s sample size should increase.