**Crowdfunding Data Analysis**

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

* 1. Theatre had the most campaigns across all categories, followed by film & video and music. However, journalism, photography and technology had the highest success rates, even though they had fewer campaigns.
  2. Most campaigns were below $9,999 and had varied success rates, all above 50%. However, the campaigns with the highest success rates set higher goals. While there were far fewer campaigns at this level, those with goals between $15,000 to $34,999 averaged a success rate of 94%.
  3. The most successful time of year for campaigns to be created was in June or July, only to be followed by August which was the least successful time to start a campaign.

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

While this dataset does include a lot of valuable information, there are still limitations to what it can tell us. While we can analyze the trends of what leads to a successful or failed campaign, and look for indicators such as the time of year it was posted, or the goal that was set, or the number of backers, this data does not take into account other external factors such as the economy, the location (within the listed country) or social media. This dataset spans ten years, during which the economy was recovering from a recession, only to be hit by a pandemic with even more economic upheaval. Additionally, while kickstarter campaigns occur virtually, perhaps there is more local momentum to support theatre projects in rural communities as opposed to urban ones with many theatre options. However, it is hard to know, as the location within each country where these campaigns occurred is not listed. Finally, a large factor that I think is neglected within this data set is the presence and impact of social media. So many kickstarter campaigns are promoted heavily through social media, but the number of “shares” is not indicated on the dataset. Moreover, social media has transformed tremendously in the last ten years, and may have had more of an impact in 2018, than it did in 2010. These limitations affect the overall analysis of dataset.

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

* 1. Some further analysis that we could conduct using tables and/or graphs could be around the “staff pick” or “spotlight” categories. I am wondering what impact (if any) these behaviors may have on the outcomes for campaigns and/or specific categories.
  2. Additionally, while we analyzed the years that campaigns were put on, we did not analyze the duration of this time and if it had an impact on the outcome of a campaign. I think this analysis could be fruitful to think through the outcome of a campaign, depending on how long it was posted.
  3. We calculated the average donation, but did not cross check this against the category and/or sub-category for each area. Which categories had larger donations per backer or less? How did this impact the overall success of the campaign?
  4. Finally, we spent some time looking at the categories and their success rates, but we did not spend as much time looking at success rates within a category. I created an additional pivot table after the first, just looking at the percentages by category. This was helpful in drawing conclusions about how a category effects the overall outcome of a campaign.

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

In this particular data set, where there are larger outliers, the measure of central tendency that better summarizes this data is the median. I can see this because for successful campaigns, the minimum number of backers is 16, the maximum is 7,295, the median is 201 and the mean is 851. The difference between the mean and the median shows that the mean is skewed by the outliers to the right, and therefore the median is a better summary of the number of backers per campaign. The data for the failed campaigns demonstrates a similar preference, with a minimum of 0 backs, a maximum of 6,080 backers, a mean of 585.62 and a median of 114.5. These large maximum values indicate outliers, for which the mean is susceptible, and therefore the median is a preferred measure to summarize 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?**

There is more variability with successful campaigns than with failed campaigns, in terms of backers. This is evident because the variance for a successful campaign is 1,603,373.73, with a standard deviation of 1,266.24, while the variance for a failed campaign is smaller at 921,574.7 and a standard deviation of 959.99. This makes sense because on the whole, successful campaigns are more likely to have more donors than unsuccessful campaigns. However, when accounting for all of the factors that contribute to a successful campaign, such as the category of the campaign, the goal set, the time of year it was posted etc. this leads to a wide range of variability between the number of backers per campaign. In some higher goal set campaigns, there were fewer donors who contributed large amounts. Conversely for other campaigns, there were many donors who contributed very little. This results in a wide range of variability.