1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. We see an increase in successful crowdfunding campaigns beginning in May and running through July. From July to August, we see a decline in successful campaigns. Based on these observations, we can conclude that crowdfunding campaigns seem to be slightly more successful during the early summer months.
   2. When looking at the outcomes of campaigns by category, it would first appear that Journalism crowdfunding campaigns are the most successful with a 100% success rate. However, we have a very small sample size of Journalism campaigns coming in at 4 total events, which isn’t a large enough sample size to make the conclusion that this category of campaigns is the most successful. The next highest success rate is for Technology campaigns, which has a 67% success rate and a 29% failure rate. We do have a decent sample size of technology campaigns to analyze with a total of 96 campaigns with data. Therefore, we can conclude that Technology campaigns have been slightly more successful in terms of meeting donation goals.
   3. Looking at the sub-category vs outcome chart, we can see that most of the campaigns were plays. However, plays actually had a lower success rate than the majority of the sub-category campaigns, coming in at 54% successful campaigns. We can also see the same issue as in the above bullet point of two sub-category campaigns having a 100% success rate but only a small sample size (Audio with 4 campaigns and World Music with 3 campaigns). The next sub-category with a larger sample size is Web with 51 total campaigns and a 71% success rate. While it would be helpful to have a larger sample size, we can ultimately conclude that web campaigns saw a slightly higher success rate than the other sub-categories.
2. What are some limitations of this dataset?
   1. One limitation in the data is the difference in campaign goals. It is important to keep in mind the variation in campaign goals as we might see greater success in campaigns that had lower goals.
   2. Another limitation of this data set is that the majority of the sub-categories for campaigns were plays. This makes it difficult to draw conclusions around the other sub-categories as they all had much smaller sample sizes.
   3. An additional limitation of this dataset is that it does not consider reviews of the campaigns. It only views success in terms of reaching donation goals. Reaching donation goals is definitely an important factor of success, but public opinion is also an important factor as an event could meet its donation goal but not be viewed as favorably by those who attended and made pledges.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. One additional table that I created looked at the % of success and % of failure by campaign category. I thought this table was helpful as it helped me analyze which category of event tended to see more success in reaching donation goals. While there was a graph that visualized this data, I found it helpful to look at the actual percentages as it was a bit difficult just looking at the bars.
   2. Another table that would be helpful to create would be one that examines average donation by outcome. When looking at the donation goals for the different campaigns, there was a lot of variation between the goals, with some being much higher than others. Therefore, simply looking at the outcome to determine success of the campaign negates that some campaigns did receive high donations and just didn’t hit their goals or were still live/canceled. For example, I created a pivot table to look at average donations by outcome, and it seems that live and canceled events actually had slightly higher average donations than successful campaigns.
   3. Moreover, we could create graphs that show average donations by category to see which categories of campaign receive the highest average donation. Based on this chart, publishing campaigns had the highest average donation and journalism campaigns saw over 50% less average donations than the other campaigns.
   4. We could also create a graph to look at average donations over time to see if there is a time of year when campaigns receive higher average donations.

**Statistical Analysis**

1. Use your data to determine whether the mean or the median better summarizes the data.
   1. In this case, the median is a better value to summarize the data as there are multiple outliers outside of the upper whisker in both data sets that will increase the value of mean, making it a less accurate representation of where the majority of the data falls.
2. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   1. There is more variability within successful campaigns. In part, this could be due to there being a larger number of successful campaigns to examine as there were 565 successful campaigns and 364 unsuccessful campaigns. Additionally, unsuccessful campaigns seemed to be more likely to have lower numbers of backers (as low as 0 backers and with a mode of 1), which would make sense since having less bakers would be more likely to result in a failed campaign. That doesn’t mean that having a high number of backers would equate to a successful campaign as there were 6080 backers in the max value of the unsuccessful campaigns data set. However, overall, it seems that the successful campaigns overall had a larger number of backers with the upper whisker value being 3016 and 41 data points above that while the unsuccessful campaign had an upper whisker of 1910 and 31 backers above that. Ultimately, as a result of the larger number of outliers and larger sample size, successful campaigns saw greater variability in the number of backers.