Included is the word doc to answer the questions under the “instructions” section as well as the questions in the “statistical analysis" section.

**What are 3 conclusions that we can draw from the data?**

1) We tend to see more successful campaigns when they start in the summer months (June and July). In all 10 other months, there tend to be fewer successful campaigns; however, we do not truly know if this shows true causation (i.e. the percentage of successful/unsuccessful campaigns could be relatively uniform across each month)

2) We tend to see fewer failed campaigns in the fall months (September, October, and November). Again, though, this only means the overall number is lower in these months; the true percentage of failed campaigns in these months could still be the same as every other month.

3) The amount of canceled campaigns remains relatively constant across each month; we do not see, in our line graph, much difference on a month-to-month basis in the amount of canceled campaigns.

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

As I have previously mentioned, we do not know the percentage of successful/canceled/failed campaigns in each start month/category. These values are extremely crucial in signifying which month (if any) or which category is more successful on average than the rest. If we only know the overall numbers, the categories and months with larger samples will always appear to be “more successful” because they have more campaigns (they will also seem “more unsuccessful” because of this). True percentages will give us the detail we need to start making verified/data-driven conclusions. In addition, we also need summary statistics such as mean and median. These values would have the same impact of percentages because we can find out how much more (or less) successes there are in certain months/categories than others in relation to the rest of our data. For example, we have 49 successes in the month of January; now, if our mean and median values for successes in a month are around 40, then we can claim that campaigns started in January have more successes than compared to other months. But, if the mean and median are around 60, then we can conclude the opposite; and, if the mean and median are roughly equivalent to 49, then we can conclude that January is very representative of the dataset.

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

We could filter count of outcome by years or by category and create line graphs using those as our x-axes. These would give us value by informing us which year or category of crowdfunding campaigns results in the most successes or the highest percent of successes. We could also create pie charts for grand total, successful campaigns, failed campaigns, and canceled campaigns. This way, we could see which months/categories/years take up bigger portions of the pie in successes/failures/cancels than in the grand total pie, showing us which are therefore more or less successful in comparison to the rest of the data. We could also create data tables with summary statistics for outcomes in relation to month, outcome in relation to year, and outcome in relation to category; this would also allow for us to compare each month/year/category value with the rest of the data (via mean, median, standard deviation values) and allow us to form more educated conclusions. We could go in many different directions with extra tables and charts, but these are just a few examples.

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

We see in both successful and unsuccessful outcomes that are means are much greater than our medians. This is normally a sign that our data (for both sets) is positively skewed—meaning that most data points are concentrated closer to zero or smaller numbers, but there are a few extremely large values that skew the mean from properly representing the data. So, this tells us that the median in both sets more properly summarizes the data since they show where most of the data lies.

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

From our summary stats, we see that the variability and standard deviation for successful campaigns is greater than the respective values for unsuccessful campaigns. This makes total sense as unsuccessful campaigns are more likely to receive no backers or very few. On the other hand, successful campaigns could receive 1 very generous backer or thousands of casually-donating backers. This means that the number of backers for unsuccessful campaigns is more likely to stay similar (at lower/smaller values) and not vary all too much whereas the number of backers for successful campaigns is more likely to vary monumentally from campaign-to-campaign.