Crowd Funding Analysis

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

* The three most common categories for project release are Film & Video, Music, and Theater. (Parent Category Sheet)
* The Parent Category projects that appear to be more successful than failures are Journalism, although it has only four projects, Technology, and Publishing. (Parrent Category Sheet)
* Throughout the year, the projects appear to be more often unsuccessful, failures or canceled, if they were created in the months of May, August, and December. Having the success count stay the same or go down, and the failure and canceled count go up in these months. (Month Created Sheet)

1. What are some limitations of this dataset?

* One of the limitations of this dataset is its size, even though it has 1000 rows, we need more information to make more accurate conclusions. For example, the Journalism category was a success for all the projects in that category, but the number of Journalism projects is only four, a number too small to make an accurate judgement if the next Journalism project would be a success.
* Another limitation is the small number of projects per country, most of the countries have less than 40 or 50 projects overall, only the US has a significant number of projects.
* Research has shown that the way we ask for donations matter, so having more information on how the description and goal of each project is organized or put together would help us in predicting which projects have greater chance of success.

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

* Creating a table and graph that show us the percentage of successful and unsuccessful projects is a helpful way to see in a more precise manner which kind of projects are more successful than the others.
* A graph and/or table that divide the projects into a group of goal of smaller amount and a group of greater amounts, since there are some projects that need more than $100,000 dollars. That might be an interesting correlation to look at to know if having a smaller goal influences backers to donate.
* A table showing the relationship between “Spotlight” and “Outcome”, to see how much a project being on the spotlight helps or detriment it in being successful.
* A table and graph for the relationship between the “Average Donation”, “Goal”, and “Outcome” to see if the these projects were more successful due to larger donations on average.

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

I believe that using both the mean and the median is the better way to summarize the data. Both provide important information on how the data behaves, but if I were to choose, I would pick the Median to summarize this data set. It gives us a more accurate number of backers that most projects have, that at least half of them have less than a few hundred backers. Since the mean in this case is being pulled towards a much higher number due to a portion of the projects have a very large number of backers.

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

We can clearly see through the Variance calculations that the successful campaigns have twice the variability of the unsuccessful ones. To a certain extent this does make sense since unsuccessful campaigns should have less backers on average, and since we can’t have less than zero backers, the number of backers gets more condensed or has les variability due to its smaller range. Also, the number of unsuccessful projects with a large number of backer should be much lower than the successful projects.