Homework 1   
Tim O’Donoghue

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. When looking at success/failure by Parent Category, we can see that every category has a number of both successful and failed campaigns. This indicates that one parent category does not necessarily attract more funding than another.
   2. With a quick glance at the data (and without running a statistical analysis), it seems that the strongest corollary to success is the number of backers in the campaign. It suggests that campaigns with more backers are more likely to succeed in achieving their funding goals.
   3. We can use the data created and date ended columns to assess how economic conditions affect fundraising targets. This would take an extra step, but by contextualizing the campaigns within economic conditions we would be able to assess how elastic donors are in their giving relative to their financial flexibility.
2. What are some limitations of this dataset?
   1. We know very little about each campaign. When the organizer is an organization, are they national or local organizations? When they are individuals, are they first time entrepreneurs, long time philanthropists, well connected, etc.? Where are these people from? Knowing this could help assess the variables that go into a campaign’s success or failure.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. We could create a scatterplot that shows the relationship between pledged amount and target fundraising goal, which would show how causes did relative to their stated goal. We could analyze the average donation per campaign, which would demonstrate donor engagement. We could create a sub-category next to name that clarifies if the fundraiser is an organization or an individual, and then create a graph showing success based off individual/organization, which would give insight into donor preferences.
4. Use your data to determine whether the mean or the median better summarizes the data.
   1. There is a large discrepancy between the mean and median for both outcomes. The large discrepancies suggest that there are outlier data points, which have skewed the mean. When this occurs, it indicates that the median is a better summation of the data.
5. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   1. The data shows that there is more variance in successful campaigns than in failed campaigns. This makes sense. The reason for a campaign being successful could be highly variable- it could take off on social media or attract a particular niche market. Unsuccessful campaigns, on the other hand, are likely to have recurring themes, such as poor marketing, failure in their messaging, or just be raising money for a shoddy campaign. All of this would result in a lower backer count and a failed campaign, which we expect to have lower variance.