**Module 1 Challenge**

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

1. Majority of the crowd funding projects end in success.
2. Majority of crowd funding projects are done for theater projects
3. Majority of crowd funding projects are done for plays

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

* Some of the crowd funding projects are still live. Once they are completed, they could change the trends we observed from the current data set.
* A small number of projects were followed. This means that the trends observed might not reflect the actual trend observed of all crowd funding projects in the world.
* A vast majority of the projects are in the United States. This means that this data is heavily bias towards trends in the United States and may not represent other countries accurately.

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

1. It would be interesting to see what Parent category and sub-categories were most like to have amounts of money pledged that was higher than their goal. This could show what kind of projects receive the most interest/ which project people are more likely to fund.
2. A graph showing the percent of the campaign funded vs country. Are there countries where a higher percentage of Crowd Funding campaigns succeed? (This would probably require more data on other countries because the United States is represents a larger proportion of the data than the other countries).

**Bonus Statistical Analysis**

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

In both the successful and failed crowd funding outcomes the median would be better to summarize the data. Both of these data sets have very higher standard deviations meaning that there is a non-normative distribution of data and shows the data will most likely have a high presence of outliers. These outliers will have heavy influence on the mean and therefore Median is a better representation of the middle of the data because it is not influenced by the non-normative distribution of data.

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

Seeing as the successful outcomes data set has a higher standard deviation, it therefore has a higher variance. This is most likely due to the fact that failed outcomes more consistently a lower number of backers in total in comparison to successful outcomes (this is one common reason as to why they fail). Having less backers overall will push down your average mean and lead to lower variability in the data. Another reason is that these campaigns do not have a limit on the number of backers they can have. This could lead to an extremely high number of backers to certain campaigns that people support or have a good marketing strategy. Having campaigns with an unusually high number of backers will lead to a larger variance of the mean and inturn higher variability.