Module 1 Challenge Questions

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

Three conclusions that can be drawn from the given data are:

* 1. That it is more likely for your project to be successful than to fail
  2. Most projects are in Theater and more specifically, plays.
  3. The Summer Months of June and July showed the most success on crowdfunding projects.
* What are some limitations of this dataset?

Some limitations of the dataset are:

* 1. There is only 1000 data points and so the sample size is relatively not very large. This could be a bad set of data or not significant. We don’t have enough knowledge to know the value of the analysis we collected.
  2. This dataset uses multiple types of currency. This means that you truly cannot directly compare pricing as the value of each funded is not equivalent to the numbers presented.
  3. This dataset is over multiple years and so we don’t know if there was some significant outside factor in the success or failure of these projects. An article could have come out before one of these promoting them or the website could have gotten an update. We truly don’t know if some had a better chance than others besides the “Staff Picks” and “Spotlight” descriptions.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Some other tables/graphs we could create are:

* 1. A visualization based off the spotlight feature. That would allow us to determine whether success was more or less likely based on a spotlight.
  2. The staff pick feature. We could see if the projects gained more success from becoming staff picks or not and see if that should be promoted.
  3. A Visualization of the outcome based on how much currency was asked for. This could help us determine what the sweet spot is for asking for money. Whether it be percentage or just flat amount I believe both could help

Statistical Analysis Answers

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

In the Successful outcomes, the Median better summarizes the data than the mean because of how skewed the data is. There are multiple outliers and so the mean would be swayed by those. Therefore, the median more accurately represents the data.

In the failed outcomes the Median better represents the data because the data is again skewed by outliers causing the mean to be a lot larger than a good summarization of the data. With the removal of outliers, the mean would more accurately summarize the data.

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

After looking at both the successful and failed statistical analysis, it appears that the successful outcomes have more Variance. This makes sense because there are more successful campaigns and so it is more likely for the data to have larger outliers with that amount of data. The Maximum value is also significantly larger which causes the possible amount of data inside to have a larger chance of being farther away.