Jay Manion

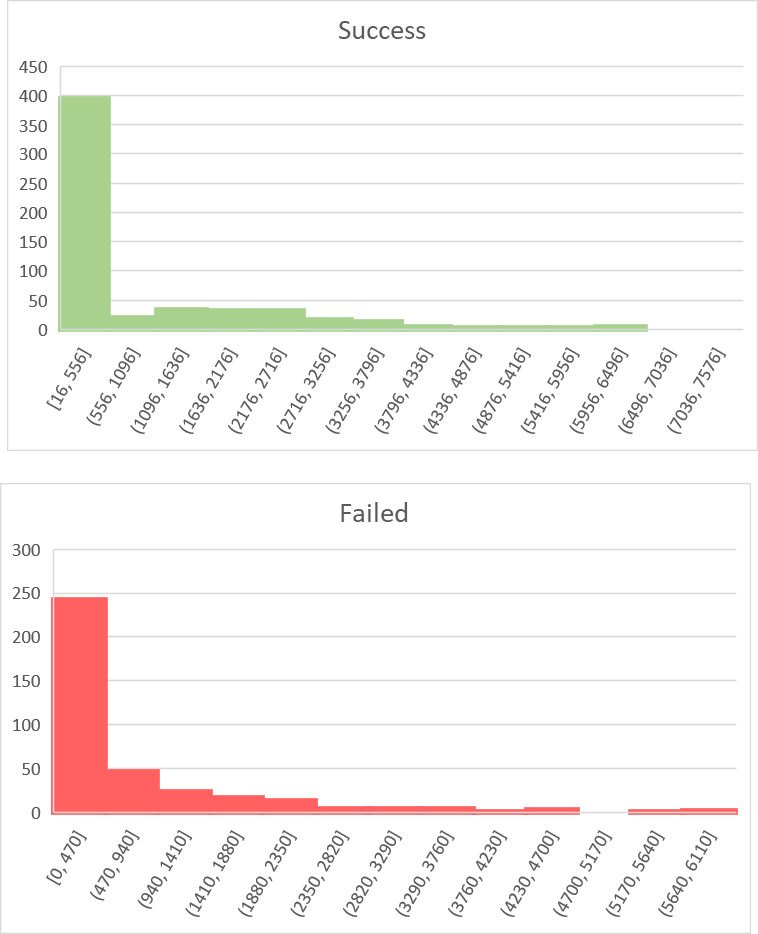
12/17/2023

Data Bootcamp

Module 1

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + Three conclusions we can gather from the data are explained below.
    1. The highest count of parent categories that were used in this data set are Film & Video, Music, and Theater.
    2. This breaks down into the subcategory largely typing to Plays being the most used by a long shot.
    3. Finally, we can conclude that the most successful times appear to happen around the summer months.
* What are some limitations of this dataset?
  + One limitation to this dataset is that we are only looking at the count of outcomes vs when the dates were started. This is not a representation of what was the most successful in terms of currency pledged. While it appears to be any type of play during the summer being successful, that might not be the highest gain towards the goal.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + We need to review a few other aspects such as profit gained per parent category, subcategory, and by month. This will either enhance the direction of the story that the initial view appears to show, or it will be more telling of different types of categories and months that yield better results.





* Use your data to determine whether the mean or the median better summarizes the data.
  + I would have to say that the median better summarizes both Success and Failed data sets. Looking at the histograms, it appears most line items fall within the first set of numbers, (16-556 for success, 0-470 for failed). When looking at the median for each they are within both ranges. Taking the median will help eliminate the outliers that inflate the mean and give a better direction in terms of assessing future data. Another indication that fits this thinking is that the variance for both is very large. This means that the average distance from each point to the mean squared is very high, denoting that there is a large spread between data points.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  + There is more variability within success than there is failed. This does make sense. One reason that fits this is when looking at the percent funded. If it is under 100% then essentially it was a failed goal. However, looking at all the successful goals, the percent funded could be anything from 100% to goal all the way to the highest 2339%. This would mean more money was donated, more backers helped, the average donation was larger. All of these create a much larger variability within the data set.