Module 1 Challenge

Crowd Funding Homework

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

1. The first conclusion we can draw from the data about crowdfunding campaign is that the parent category theater was the most popular type of campaign for crowdfunding as it was about 34.4% of all campaigns.
2. The second conclusion we can draw is that out of the 1000 campaigns 56.5% were successful, 36.4% failed, 5.7% canceled, and 1.4% were live. Also, only 2 of the 24 subcategories had 100% success rate which were audio and world music.
3. The third conclusion we can draw is that over the years the month that averaged the most successful campaigns was June with a success rate of about 63.9%.

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

Some of the limitations of this dataset is that it is a limited sample size of only 1000 campaigns over the course of 10 years. Another limitation is that there are campaigns included that had zero backers so the only option for them was failure. The blurb for each campaign could have been more descriptive in order to take the categories for the campaigns and break them down even further to create more categories that will accurately depict each campaign.

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

Another graph/table we could create another stacked bar chart that would look at the average donation and the outcome of each of the parent campaigns to whether they were successful, live, failed or cancelled. This additional value would show if having a higher average donation lead to a successful campaign or not. Another three graphs we could crate are three different stem and leaf plots with the backer counts and a graph for successful, cancelled, and failed. This would really show us the outliers for the outcomes of the campaigns.

**Does the mean or median better summarize the data?**

The median better summarizes the data as there are a large number of outliers for both successful and failed campaigns. As for successful campaigns, out of the 565 successful ones, 389 of them had 500 backers of less and the mean is showing as 851 which does not accurately represent the data. The outliers are causing the mean to become so high, and it causes it to stray extremely far away from the middle number of 201.

**Is there more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

There is more variability with successful campaigns. This does make sense due to the large difference between the mean and the median. The variance for successful campaigns is showing that the spread of backers is extremely high especially when the minimum is 16 and the maximum is 7295. The reason this makes sense is because while the median is 201 and there are 157 successful accounts that have over 1000 backers. This shows that the spread from the median to accounts over 1000 backers is very large and creates more variability.