Kevin Nazario

Module 1 Challenge/Homework- Written Report

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

There are several conclusions that can be drawn about crowdfunding campaigns within the context of this sample of data. First, overall, campaigns were more likely to be successful more than half of the time. Second, this broad conclusion held true across parent categories, with the exception of “Games”, which was due to the particularly poor rate of success for the “Mobile games” subcategory. And finally, although there appears to be a slight increase in the amount of successful campaigns in the months of June and July, this general trend varied significantly when looking at individual years, so the time of year does not seem to have a strong impact of the likelihood of success.

* What are some limitations of this dataset?

There are various limitations that can be considered within this dataset. Depending on the conclusions that we are trying to draw, it is difficult to ignore that over 75% of the data is in the United States, which potentially skews any inferences you can make (or your ability to generalize any conclusions). Additionally, about one-third of the dataset is within a specific sub-category (plays), which can speak to it’s popularity, but also may skew and broader conclusions you can make. This also means that there is a very limited sample size in some of the categories, which may not be representative of larger potential trends. Finally, the data point that a goal was met, provides some information about the underlying project, but does not necessarily equate to if that project continued success in the future. All of these factors would need to be considered as you analyze and interpret any conclusions.

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

Additional graphs and tables could provide some additional insight on trends. One additional table and graph would be to specifically filter the successful campaigns and build a table that shows the percent funded per parent and sub-category. This may reveal some additional trends on which campaign categories had the most success relative to their goals. A box plot may be useful here as well to consider outliers as some of the datapoints had funded rates of over 1000%. Another graph that would be useful would be campaigns over the years. I was able to see the individual years by filtering the third graph, but in that view, it was not as easy to see if there were any trends or changes in the types of campaigns across the years (i.e. maybe there was a year were a certain category was most popular).

**Statistical Analysis (Backers- Successful vs Unsuccessful)**

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

Regarding both the successful and unsuccessful backers statistical analysis, the median provides a more representative summary of the overall data. In both cases, there are several projects that have a much higher number of backers (over 3000), where the majority of the data is between 0 and 1000 backers. These outliers create a positive skew for both sets of data and pull the mean much higher than where most of the dataset is. Because of this, the median values are much closer to the majority of the data and a better summary of 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?

Reviewing the data, while both successful and unsuccessful campaigns have a lot of variability in the number of backers, the successful campaigns had more variability (which can be seen with the higher variance and standard deviation values). This makes sense because the mean for the successful campaigns is significantly higher than the mean for the unsuccessful campaigns, meaning there are more datapoints in the successful dataset bringing the mean up and creating a more spread out distribution of the data.

Intuitively, this could also make sense because in general, you could imagine that campaigns that ended up being successful would have a wide range of backers (relative to the goal size, average amount donated…etc.), whereas unsuccessful campaigns would have less backers.