Devorah Zweig

December 22nd, 2022

Module 1 Challenge

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

One conclusion that one can draw from the data about crowdfunding campaigns is that the projects that were directed in the summer months had a drastic increase in success. Another conclusion one could draw is that from all the sub-categories, the “Web” sub-category had the highest rate of success, with about three times more successes than failures. Lastly, it can be concluded that, while journalism had a very low project rate, they had a 100% success rate. It would be interesting to see what would occur if more journalism projects were funded.

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

Some possible limitations of the dataset are that some projects were given enough time to reach their goals, while other were extremely limited on time. In addition, some projects that failed did not have a significant amount of backing. The lack of funding and support ensure that the project will not be successful.

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

Some values and graphs that I believe could provide additional value to the study is creating a new column that finds the ratio of successes relative to failures. In two of the Pivot Tables, it is easy to notice that one or two categories contained a higher number of successes than the others. However, they also contain some of the highest number of failures. I believe it would be beneficial to create that column and then create a graph to show the percentage of successes to failures for each project. In addition, I believe that it would be beneficial to create a new column that shows the ratio of successes to how much financial backing the project had. This will allow us to determine if the two are correlated.

Statistical Analysis Questions

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

After plotting the data for both successful and unsuccessful campaigns as a scatter plot, one can see that the data is bottom-heavy for both. After comparing where many of the points lie on the graph, one can conclude that the median better summarizes the data. In addition, the bar graphs for both datasets indicate a randomly distributed distribution, which generally indicates to use the median, while a symmetrically distributed distribution will usually implicate that the mean is a better indicator.

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

The data suggests that there is more variability with the successful campaigns, which has a variance of 1603373.732. This makes sense because the successful campaigns generally had a higher number of backers. These high numbers will likely skew the data and raise the variance.