Adriana Galindo

Data Boot Camp

June 25, 2023

Module 1 Challenge

Crowd Funding Questions

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

The first conclusion that can be drawn from the first data titled “Count per Category”, is the following: Campaign results vary depending on the category. From this data, we can point out that the campaigns done through theater was the most successful. However, this was result may have been more successful than all the other categories because it was used the most by almost double the amount in contrast to the other categories.

The second data demonstrates a closer look to the subcategories of the campaigns that were ran. This data set “Count per Sub-Category” indicates that the campaigns that ran photography book campaigns were the most successful with over 60% of success in comparison to the small amount of success of the other categories.

The third data set shows an increase of successful campaigns when they are started from the months of May-July. After these months, the number of successful outcomes diminishes.

1. What are some limitations of this dataset?

There is not enough information to draw an informative conclusion about the success of crowdfunding campaigns. The only information that these data sets are offering are the levels of success by categories. This data set is also yet to be finalized because there are campaigns that were still live. Another limitation is the limited dates of this data. This data is from a set time thus these results only reflect this short time. To know more, the dates should be extended or there should be studies of every decade.

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

We could analyze whether the duration of the campaigns affect their outcomes. Are longer campaigns more successful?

We can probably use a scatter plot to analyze whether the geographical location of the pledgers played a part in their success. We would be able to figure out if there are correlations between higher incomes and higher success in these geographical locations.

Statistical Analysis Questions

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

If I create a scatterplot with the successful data, the median summarizes the data better even though the mean should be the one that should summarize the data better. This mean is in the higher numbers because there are also a lot of higher data, however the higher data is not the most common. For the failed campaigns, it is the same situation. The median is also the better summarization of the data. Most of the data is also scattered near the 114.50.

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 standard deviation for successful campaigns is 1267.37 while the failed standard deviation is 961.31. This indicates that there is more variation in the successful backers. There is very high success which indicates that there are more variations. For the failed section, there is not much variety.