HW1: Excel Diana Kennen

3/12/2021

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

While the theater category had the largest number of successful projects (839), the music category had the largest percentage (77%) as compared to film (58%) or theater (60%).

Within the subcategories of the music projects, rock and indie rock projects were, by far, the most successful with 240 successful rock projects and 138 successful indie rock projects.

More successful projects were created in May (234) than in any other month and the number of successful projects generally decreased afterwards through December.

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

The data table does not provide much in the way of context for why some projects were successful and other projects were not. What were the characteristics or qualities of the projects themselves that lent to their success or failure? We also do not have information about advertisement of the Kickstarter projects other than the spotlight feature.

1. *What are some other possible tables and/or graphs that we could create?*

We could create a graph for the difference between the data created and date ended for successful, failed, and canceled projects to discern any differences in the length of time that each category of project is posted for within each outcome group. We could also create a histogram for the percent funded for failed and canceled projects to investigate the distribution and what proportion of projects were close to being successful.

I am not too familiar with Kickstarter, but if the spotlight feature is something that draws attention to certain projects within the platform, it would be interesting to do an analysis to see what impact that had on the likelihood of success. An independent t-test with number of backers or pledge amount as the dependent variable and the use of spotlight delineating the groups would indicate whether or not there was a significant difference in the number of backers or pledge amounts between the two groups.

**Bonus** 2:

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

In both successful and failed projects, the distributions have a positive skew and the mean is much larger than the median. This indicates that a small number of projects with high numbers of donors are impacting the mean as outliers. As a result, the median number of backers would be a better measure of central tendency in each case.

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

The variance is much higher for successful Kickstarter projects (7.1 E 5) that for unsuccessful project (3.8 E 3). This makes sense, because it is highly unlikely that many of the failed projects would have a large number of small donation backers to increase the spread of the distribution, otherwise they would not have failed. Successful projects can be funded either with one large donation or many small donations, resulting in a wider variance in the number of backers.