Module 1 Challenge

Joshua Baltzell

3/25/2024

Conclusions and Limitations

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

The first and second conclusions that I was able to make are, one, that the theater/plays categories are the most frequently sought after for crowdfunding; and two, despite this large volume it is not the most successful as a percentage of the total projects started. Out of 1000 project, Theater had 344 total projects put forward. Looking at the success rate of the parent categories as a percent of the total projects within the category, rather than a percent of the grand total, changed the story.

Technically, journalism was the most successful (100% success rate), but a sample size of 4 is not significant enough to draw conclusions from. The parent category of technology was the most successful at 66% success, followed closely by Photography (61.90%) and Publishing (59.70%).

The third conclusion I was able to form is that projects were more likely to be successful if they were started in the early summer months (June and July had a 62.46% average success rate, compared with 55.37% average for all the other months).

One last item that I noticed that was surprising to me (assuming this data is taken from a real source) was that the project status as a spotlight of the crowdfunding website had little impact on its success. If I had to make a prediction, I would have assumed that being a spotlight of the website would increase the likelihood of success. However, non-spotlight projects had an almost 3% higher success rate compared with spotlight projects, 57.10% and 54.85% respectively.

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

One of the limitations that I thought of while looking at this data was the lack of data about specific donations to projects. After calculating the Average Donation column, I noticed a similar problem to addressing the difference between median and mean (statistical analysis below). This lack of data makes it so we cannot be sure if certain projects were successful because of one large donation as opposed to many smaller donations.

Another limitation to this data is the effect of social media on how many backers are received, along with the pre-existing popularity of a project creator upon the launch date of the project. Crowdfunding could be more likely to be successful if a company is already well established and has a following on social media. This would also account for the skewing of the data regarding the variance in the backers between successful and failed projects.

Finally, it would be interesting to see how frequently certain individuals back multiple projects. Some backers may be interested in backing a certain type of project and are more likely to back multiple projects that are correlated.

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

I mentioned this earlier in my conclusions, but adjusting the pivot table to focus on the rate of success as a percent of total projects would create a better representation of whether a category is worthwhile for crowdfunding. Additionally, disaggregating the data further to show a degree of success would also be beneficial. A chart that focuses solely on the successful projects would show not only the most beneficial crowdsourcing, but also the potential effect of public opinion. A project that raised 100% of its goal is a success, but a project that raised over 1000% of its goal would be far more successful than the former (by ten times).

To take this a step further, we could look at the length of time a project was open for crowdfunding. A project that reaches its goal after only a month of being open would be far more lucrative than a project that took over a year to reach its financial goal.

**Statistical Analysis**

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

In this context, I think the median best describes the data. In both cases, the variance in the data is high relative to the values in the backer’s column. This high variance shows that the average will be skewed to look higher than it is because of the few projects with many backers (>5000) compared with the many projects with a small number of backers (<1000). For the successful projects, only 13 had more than 5000 backers, but 408 projects had fewer than 1000 backers. Similarly, only 3 failed projects had over 5000 backers, and 292 had less than 1000. The median is lower than the average in this context, which gives a better representation of the distribution of backers.

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

There is more variability with successful projects compared with failed projects. The variance for the successful projects was 1,606,216.6, while the variance for the failed projects was 924,113.5. This does make sense, since the project creators choose their intended goal for the backers. A successful project could have a low monetary goal that is easy to achieve with a small number of backers thus making it a higher chance of success. At the other end of this spectrum, a popular project could have a higher goal set, but the popularity will get them a larger number of backers.

For failed projects, regardless of the set goal there will more than likely be a small number of backers. Rarely can you expect to see many backers on a failed project. This can be represented by the median in both contexts: successful projects had a median number of backers at 201, while failed projects had a median of 114.5, almost half that of successful projects.