Caleb Cutrer

Data Analytics Bootcamp

Professor Thomas Bogue

4/4/2024

# Part 1:

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

1. Crowdfunding campaigns seem to be an especially popular means of acquiring funding for projects that fall into the category of theater/plays, as campaigns in this category made up over 34% of the total number of campaigns in the dataset. This trend was true in all countries.
2. Campaigns that launched in the summer months of June and July were the most likely to have successful outcomes.
3. Campaigns with moderate financial goals ($15,000-$30,000) had the highest rates of successful outcomes, while those with high financial goals (over $50,000) had the lowest rates of successful outcomes.

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

One limitation was that the dataset sampled several countries. This becomes an issue when we are looking at the outcomes of campaigns based on the categories they fall into, as different countries have different cultures. So, for example, if things pertaining to “film & video” are not popular or valued in a country, then campaigns in this category that are launched in said country would be less likely to have a successful outcome (and there would be fewer potential backers). Relatedly, there are different currencies that fall into the “Average Donation” column, so the numbers are somewhat meaningless. Another potential limitation would be the overrepresentation of campaigns in the “theater/plays” category. There were 344 such campaigns in the dataset, while there were far fewer campaigns in several of the other categories (i.e. there were only 4 campaigns in the “journalism” category). This makes it hard to make generalized conclusions about the data, as factors that are unique to campaigns in the theater/plays category would have a disproportionate influence on the overall data.

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

I think it would be interesting to create standardized “average donation” and “Goal” categories/columns in which all forms of currency are converted into a single common currency. Then, we could recreate the stacked line chart that looked at outcomes based on goal and get a better understanding of that relationship.

# Part 2:

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

The median better summarizes the data on backers for both unsuccessful and successful campaigns, as both distributions are positively/right skewed. The mean would be impacted by this skewness, and “pulled” towards the high outliers (such as the 1 successful campaign that had over 7,000 backers). The median, however, is not impacted by the presence of outliers. Lastly, upon making a histogram of the backers in both the successful and unsuccessful groups, it is visually apparent that the median is a much better measure of central tendency than the mean.

**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 greater variability in the number of backers of successful campaigns than unsuccessful campaigns, as indicated by the larger standard deviation. At first, I didn’t think this made sense because the sample size for successful campaigns (565) was larger than the unsuccessful campaigns sample (364). However, this makes sense when considering the larger range of the number of backers of successful campaigns, and more importantly, the presence of more severe high outliers in the distribution.