Matthew Hagarty

Columbia University Data Analysis Bootcamp

April 20th, 2023

**Excel Challenge Report**

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

1. Most crowdfunding campaigns seem to go towards the arts, with the top three categories of theater, film & video, and music, comprising 697 of the 1000 sampled campaigns, over 2/3rds.
2. Even so, those campaigns have a successful track record. Only food (22 out of 46) and games (21 out of 45 finished campaigns) have under a 50% success rate, so those are statistically less likely to succeed than the other categories. Although with such a small sample size the statistical error might be relevant.
3. Among music folks, most are rock fans, with 130 out of 175 campaigns in the music category belonging to rock (85) or indie rock (45). Though all music categories have more successes than failures of campaigns overall.

* What are some limitations of this dataset?

Besides the limitations of a small sample size when there are potentially millions of campaigns, the charts and graphs we have created do not seem to indicate any real statistical difference between campaigns that fail or succeed. While a majority succeed, there is no indication of why those succeed more often than others, or that any which have succeeded do anything differently than those that do not. This seems unsurprising, as country, category, and sub-category do not determine the worthiness of a project. I did expect the year to make a difference, depending on whether the US was in a recession at the time or not, but it does not appear to be that way.

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

Additional information which might make a difference would be the initial amount asked for, as well as how many similar campaigns were ongoing when the project launched, and how long the campaign was live for. A campaign which lasted for two weeks when there were 30 other projects of similar types competing for a large sum of money does not seem like a campaign that would succeed, but maybe it would, and we could look at the campaigns that succeeded in more unlikely scenarios and see specific information about what caused that campaign to succeed, seeing if there’s a pattern among them.

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

Based on the histogram of campaigns, it seems that the median better summarizes the data, as the data is skewed far to the right, meaning that most of the campaigns have far less backers than the mean. The few campaigns backed by thousands upon thousands of backers make the average backer count extremely high, when the vast majority of campaigns have under 300 backers, which the median more accurately reflects.

* 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 (and thus the standard deviation) of the successful campaigns are greater than that of the failed campaigns. To me, this makes sense as there are not only more successful than failed campaigns (which would usually cause a higher variability because there is more data), but also because a failed campaign can only have so many backers before the campaign becomes a success, so the possibility that a successful campaign has a large amount of backers is greater than the same for a failed campaign.