Natalie Lollin

Data Science Bootcamp

Module 1 Challenge

2/20/23

Crowdfunding Data Analysis

Using the crowdfunding data provided, along with the created graphs and charts, it appears that slightly more than half of the crowdfunding campaigns in the sample were successful. Assuming that this sample is representative of all crowdfunding campaigns, this would mean that generally a company/organization would expect a campaign that they start to have a probability of success of approximately 50% (or somewhat higher). Further analysis in three categories can be found below in an attempt to determine if there is an association between the success of the campaign and each of the following: category of the campaign, location of the campaign, and the time of year the campaign started.

**Crowdfunding Results Based on Category**

Each crowdfunding campaign was assigned to both a category and a more specific subcategory. Overall, the most common category of crowdfunding campaign was theater (and within that the most prominent subcategory was plays). The success rates did not vary greatly between categories, with two exceptions. The category of journalism had a 100% success rate; however, only 4 of the 1,000 campaigns sampled fell into this category so that is not a large enough sample from that category to be able to draw a meaningful conclusion. On the other hand, the category of games had the lowest success rate (both subcategories of video and mobile games had more failures than successes). While nothing is certain, it seems that the category of gaming would be the least likely to be successful when starting a crowdfunding campaign.

**Crowdfunding Results Based on Location**

The crowdfunding campaigns originated in several locations around the world. With that being said, the most common location of these campaigns was the United States (763 of the 1,000 campaigns sampled were from the United States). When analyzing the segmented bar charts of each country separately, there again did not appear to be any drastic differences between the success rates in each country. It did seem that the success rate in the United States may have been slightly higher overall when compared with the other countries.

**Crowdfunding Results Based on Campaign Start Date**

In looking at the line graphs comparing the success rates of the campaigns based on what time of year they were started in, a relationship between the two was not immediately noticeable. Further analysis of the difference in the number of successes and the number of failures each month (filtering the graphs both by year and by category) did show that the months of April – June tended to have the biggest difference between the number of successful and unsuccessful campaigns (with there tending to be more successes than failures during those times). Exceptions to this observation do exist from year-to-year, but there does seem to be an moderate association between the time of year a campaign starts and whether or not the campaign is successful. The general recommendation would be to start the crowdfunding campaign sometime between April and June.

**Limitations & Other Items to Include in Future Analysis**

The analysis above was solely based on the requested graphs and additions to the spreadsheet (the only additional graphs provided that were not requested are the segmented bar charts that displayed relatively frequency – which allowed for an easier comparison of the success rates within each category/subcategory). The data provided could also have been used to create visual displays that would have allowed for analyses in the following areas that could have been useful. First, while the charts were able to be filtered by location, it would have been helpful to have created a segmented bar chart in which the success rates could be compared between the locations directly. Additionally, an analysis of the duration of each campaign to see how quickly goals were (or weren’t) being met would have been interesting. Lastly, from the data that was provided, it could be beneficial to try to determine if there is a relationship between the goal amount and whether or not a campaign was successful (In other words, this would attempt to answer the question ‘is a lower goal more likely to be met than a higher goal?’ and then perhaps perform a separate analysis on the higher goal amounts and the lower goal amounts.)

An additional piece of information that could be requested to attempt to help companies make informed decisions about whether or not to attempt a crowdfunding campaign (and what strategies to use if they do decide to attempt one) would be how much (if any) money was spent advertising each campaign to investigate if there is a relationship between money spent and money earned. Aside from this piece of information, the other relationships that were previously mentioned all could be analyzed with information already provided in the spreadsheet.

**Addendum: Notes on Bonus Portion of Assignment**

After analyzing the distribution of the number of backers for both successful and failed campaigns and finding the summary statistics, the following observations can be made. Both distributions are unimodal and strongly skewed to the right (displayed in the histograms that can be found in the spreadsheets). As a result of this skew, the median number of backers (both for successful and failed campaigns) is a better representation of the central number of backers for both groups (this is because the mean is being affected by the skew while the median is resistant to the skewness). The median number of backers for successful campaigns was 201; for failed campaigns the median number of backers was 114.5 (or approximately 115). Based on the standard deviations, it was typical for the number of backers of successful campaigns to deviate from the mean by approximately 1267 backers and for failed campaigns it was typical for the number of backers to deviate from the mean by approximately 961 backers. In other words there was more variation in successful campaigns than in failed campaigns, which is logical since successful campaigns either met or exceeded their goals so they may have had many backers donating small amounts or fewer backers donating large amounts.