Benjamin Fish

December 20, 2023

Written Report and Statistical Analysis Challenge #1

WRITTEN REPORT

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

1. The most successful crowdfunding campaigns typically occurred during the 2nd quarter of the years (April to June). At the same time, that period yielded the lowest canceled rates as well.
2. According to the data provided, the most popular crowdfunding campaigns by **category** were theater, film and video, and music. With that being said, they were not the most successful when comparing the success rate to the overall data for those categories. Theater campaigns were successful about 54.4% (successful/grand total), film and video campaigns were successful 57.3%, and music campaigns were successful 56.6%.

The most successful campaigns by category ended up being journalism (100%), technology (66.6%), photography (61.9%), and publishing (59.7%). While they have the highest overall success rate, it is important to note that those campaigns had significantly fewer campaigns compared to the ones mentioned above.

1. The United States narrowly missed being the most successful crowdfunding country (57.1%). Great Britain was number one with a success rate of 58.3%, although the sample was significantly smaller than the United States.

I also decided to compare the data of the United States to that of all the other countries listed in our dataset. The United States was more successful (57.1% vs 54.4%)

What are some limitations of this dataset?

1. We don’t have more data
   * When filtering data by countries, it is clear that majority of the data we do have comes from the United States. It is much harder to make comparisons by country. That is why I also decided to compare United States vs All Other Countries
2. What is the source of our data?
   * Is our data truly a random sample?
3. There is not a universal currency in the data collected

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

1. Outcomes Per Duration of Campaign
   * Would see if campaigns were more or less successful based on the duration of the campaign being run. I believe this could be easily done with the DATEDIF function.
2. Outcomes Per Backing Donors
   * Would see if campaigns were more or less successful based on the number of backed donors.
     1. Additional factors to consider include how and when the goal was determined. Did any of the campaigns know the number of donors they would have? Did any of the campaigns alter their goals because they exceeded their initial goals?

STATISTICAL ANALYSIS

After calculating the mean, median, min, max, variance, and standard deviation, I determined that the median better summarizes the data. Once my statistical analysis was completed, I decided to generate a histogram to show the distribution of the data. This was done for the successful campaigns as well as the failed campaigns. Upon examining the distributions, both distributions were **right skewed**. In the successful campaign, 395 projects out of the 565 projects (approximately 69.9%) have 556 or fewer donors per campaign, which is significantly less than the mean. Likewise, in the failed distribution, 243 projects out of the 364 projects (approximately 66.8%) have 470 or fewer donors per campaign.

Graphs of the statistical analysis are shown below.



