Module 1 Challenge

Crowdfunding Data Set Analysis

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

With the provided data, we can see that:

1. The category to have the largest number of campaigns were theatre campaigns, and the lowest was journalism campaigns. However, journalism campaigns were more successful as all 4 campaigns had acquired sufficient funding.
2. Campaigns that had set their crowdfunding goals at anywhere between 15,000 to 25,000 or between 30,000 to 35,000 successfully met or exceeded them.
3. More campaigns were launched in July than any other month, however the month of June had a higher percentage of successful campaigns over total campaigns launched. (64% campaigns launched in June were successful.

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

1. We don’t know how each campaign procured their backers. We objectively know the outcome of each campaign, but we don’t know why it failed or why it succeeded.
2. We also don’t know how much each backer donated to an independent campaign.
3. These campaigns all took place over different time frames across different countries. We also lack reasoning behind the duration of each of these campaigns and how the timeframe (launch and deadlines) could affect the success rate of a campaign.
4. In the statistical analysis tab with the box and whiskers charts of backers\_counts for both successful and failed campaigns, we can tell the data set is right skewed. This in turn tells us that our crowdfunding data set is not normally distributed and has a lot of potential outliers.

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

1. We could insert a column in the main dataset that shows us the duration of each campaign by using the dates when they launched and when they ended. After doing so, we can create a table that shows us how many successful/failed campaigns there were in a specific time frame. This could tell us if the duration of campaign has any effect on the campaign’s success rate.
2. We can create a table that shows us how many backers existed in each country. A step further could be creating a pivot chart and pivot table that shows us the average number of backers per campaign outcome, with an additional filter on country. This could potentially tell us if the number of average backers have any effect on campaigns’ success.

Statistical Analysis

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

We know that by looking at the count of successful and unsuccessful backers per campaign, there are a lot of potential outliers and that our dataset is skewed. Given the minimum and maximum count of backers for each outcome, our dataset contents some extreme values on both cases.

In my research and understanding of central tendency, it’s important to realize that the mean could be distorted by these extreme values. In our calculations for successful backers. The average amount of backers per campaign is 851, but the median is 201. So, we know that 201 backers belong to a campaign that sits relatively in the center of this list, which would mean that it’s appropriate to assume a typical campaign that is successful would have about 201 backers.

This can be applied to what we know of unsuccessful campaigns as well.

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

If we look objectively at the variance of successful campaigns vs unsuccessful campaigns, the variance is high for the former. The standard deviation is also high for the former. BUT the total number of successful campaigns is different than the total number of unsuccessful campaigns. I don’t think that comparing the variability between these two tables would give us any reliable information.

Larger data sets could give us smaller standard deviations allowing us to draw more accurate conclusions. In this case, just because std.dev for successful backers is more than std.dev of unsuccessful backers, it does not mean that it is any less stable/reliable than the latter.

Source(s)

Website: “Mean and Median – CLINFO.” CLINFO, <https://www.clinfo.eu/mean-median/>