Sarah Demmon

Data Analytics Bootcamp – Module 1 Challenge

August 24, 2023

Crowdfunding

Pivot Tables

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

We observe that crowd-funding campaigns are more successful than not. Through our construct of pivot tables we displayed that campaigns' were predominantly successful across different parent categories, sub-categories and launch times. Additionally, our calculation of the outcome from money pledged vs campaign goal yielded a result of 565 successful campaign fundraises compared to 364 failed campaign fundraisers. The most successful campaigns belonged to the theater and plays category while the least successful belonged to theater and plays. We can infer that the event itself does not impact popularity.

What are some limitations of this dataset?

A limitation for the dataset is within the category and sub-category section. I think having some information regarding the demographic breakdown for the backers belonging to each category. This way, we can gain information about what audience these specific categories attract (age, gender, ethnicity, income) so we can create more tailored strategies to attract more backers within a category or focus resources elsewhere. I also think that city should be included so we can get a picture of how population density and demographic belonging to regions can impact fundraising. I also think that the blurb of the crowdfund is not easily understood for those unfamiliar with the industry so if there is a way to standardize the blurb into more broad sectors, that would be helpful.

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

If I were to expand on the information provided, I would want to create a stacked bar chart displaying backer count and the correlated category and sub-category. This would be a difficult task because categories and sub-categories overlap but I believe that this chart will provide information on the count of people that a category can attract. Fundraising campaigns can also be supported by name recognition.

Backers

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

The median would more accurately summarize the data for both successful and failed outcomes. There the maximum values are drastically higher than minimum value which suggests a widespread. This is also evident through the large standard deviation. Means are better for data that follow a symmetric distribution because they can capture outliers that skew the data while medians is mor useful when data is distorted by outliers. In this instance, central tendency is better captured by the median.

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

We observe that there is more variability with successful campaigns. This makes sense successful campaigns generated more money from backers, so it is more likely that backers contributed varying pledge amounts. This inference is supported by the variance level which is higher for successful campaigns than failed.