Frances Jay

Hobert Bush

Georgia Tech Data Analytics Bootcamp

December 8, 2022

Module 1 Challenge Report

**Dataset/Graph Analysis**

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

Theater is, by far, the most popular category of crowdfunding campaigns based on this dataset. There are a total of 344 theater campaigns, when the second most popular category, film & video, clocks in at 178 total campaigns.

Despite the overwhelming popularity of theater campaigns existing on crowdfunding sites, the trend appears to be that around half of all campaigns in each category are successful, with the remainder of the campaigns mostly failing. I’m removing the journalism category from this analysis because it has a very small sample size of campaigns (only 4 total, and all 4 were successful), and can be considered an outlier.

When compared to the results based on sub-categories, plays are by far the most popular, but this is also the only sub-category available for theater parent category campaigns.

*What are some limitations of this dataset?*

The dataset is mostly made up of US-based campaigns (763 out of 1000), so many conclusions drawn will relate more to US trends. When filtering the US out of the graphs, the parent categories trends appear to stay relatively the same.

The sub-category trends also stay relatively the same as the whole population. However, by filtering the data, we can see some clear distinctions. I compared the graphs when including and excluding the United States-based campaigns, and removed the plays sub-category as an outlier. When excluding the United States-based campaigns, the pattern is far less variable. The rock sub-category completely dominates when the US is included. When the US is not included, it is still the most popular (disregarding plays as the only option for the wildly popular Theater category), but it is more closely followed by other sub-categories.

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

Compare the popularity of each category/sub-category in each country by outcome and by all campaigns. We might see some trends in what types of campaigns flourish across North America, Europe, the UK, and Australia.

**Statistical Analysis**

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

The median better summarizes the data, as both are right-skewed. The mean number of backers is heavily influenced by the smaller number of campaigns with a large number of backers.

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

There is a greater amount of variability with the successful campaigns than with unsuccessful campaigns. I would expect a high amount of variability among all campaigns, and especially if they were successful.

The popularity/success of a campaign depends on a significant number of factors. These factors can include, but are not limited to, marketing, a backer’s perceived return on investment (will this affect their own community? Will they feel a greater sense of virtue through their magnanimous donation?), the amount needing to be raised and how much has already been raised. If a campaign hits close to home for some backers, they might be inclined to donate far more than the average backer, thus leading to a spike in variability.