Hannah Kollmann

Challenge 1

Vanderbilt Bootcamp

March 7, 2024

**Module 1 Challenge**

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

One conclusion I can draw about crowdfunding campaigns is that within the parent category, theater is the most prevalent category with 344 total campaigns. Film & video, followed by music were the other frequent categories. Another conclusion I can make (solely from this graphic) is that theater was also the most failed campaign.

A graph of different colored bars

Description automatically generated

Another conclusion I can draw about crowdfunding campaigns is that plays were the most prevalent category within the sub-categories. This suggests that people are more likely to donate to the theater category.

A graph with different colored bars

Description automatically generated

A third conclusion that you can make from the data is that there tends to be more campaigns during the summer months (July and August). January had the highest number of failures. This could be because people have more free time in during the summer months. In January, people might not have as much time or money since it is the holiday season. New Year’s and Christmas can have a major effect on this.

A graph of different colored lines

Description automatically generated

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

I noticed that there are no metrics/data that describe the success of each category as a whole. It would be helpful to see which category/sub-category does well or poorly during the relevant time periods and why. This might help us account for external factors while analyzing the data.

While there are “blurbs” for each campaign, it would be helpful to see more detailed information into what each crowdfunding platform hopes to accomplish in terms of “actionable items.” It might not be worth including campaigns with ambitious/expensive goals and campaigns with smaller/cheap goals in the same study. That could skew the data.

Lastly, the data does not provide any information on which projects bring in the most money after being launched, which could inadvertently affect the success of the campaign. There is also no data on the socioeconomic status of those who pledged. This could also contribute to how successful a campaign is considering wealthier people/companies have more money to influence and donate. I’d also like to know if donors receive anything in exchange for their donations (e.g. a product prototype, tickets to a production, t-shirts, etc.). It could impact a donor’s decision to donate if she/he believes she/he will receive something significant in exchange for her/his donation. This, in return, could affect the number of backers for different projects.

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

There are many different tables and/or graphs that we could create. I mentioned above that it would be interesting to know the success of each category as a whole to see which category/sub-category does well or poorly during the relevant time periods. Therefore, a table showing detailed data into each category/subcategory’s level of success would be good. Some possible metrics include success rate per category/sub-category; descriptive statistics for the average donation per category/sub-category; descriptive statistics for length of crowdfunding campaign timelines per category/sub-category, etc.

I would also like to see a table visualizing the outcome of projects by parent category and by country. It could show the popularity/priority of projects according to geographical location.

Finally, I want to see a display that excludes significant funding goal outliers. It would be interesting to see how that affects the data.

**Statistical Analysis (Conclusion)**

In this analysis, the median summarizes the data more meaningfully, because the mean is skewed by outliers. Although the median looks at the at the middle value, it is not influenced by outliers.

There is more variability in a successful group compared to a failed group due to higher upper outliers. This is evidence with the higher variance and standard deviation.