Victoria Sanders

**Module 1 Challenge Questions**

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

The vast majority of projects that used crowdfunding campaigns are for plays. This leads me to think that plays in particular have high upfront costs and low likelihood to draw in revenue. The same could be said of the next 2 largest categories, film and music.

Most of the campaigns were launched in the USA. If this was a comprehensive dataset, you might conclude that the United States is most likely to have artists without the funds to start projects or the least infrastructure to support artists.

Relatively few campaigns are fully cancelled. It seems like people who use crowdfunding platforms would rather leave the campaign open until the end date and fail to reach the funding goal instead of just cancelling the campaign.

* What are some limitations of this dataset?

I’m not entirely sure how the data has been sourced, so it could have these issues:

Data might be drawn from a primarily American crowdfunding platform, leading to overrepresentation of American crowdfunding campaigns in the dataset. This would make conclusion #2 above invalid.

It’s possible that some people who start crowdfunding campaigns abandon the campaign partway through. A human considering the situation might say that’s equivalent to cancelling the campaign, but the dataset might not account for that and just consider the funding level at closing date. This would also make conclusion #3 above invalid.

The type of campaign projects may be influenced by the source of the data. If film creators prefer to use Kickstarter and digital artists prefer Patreon, then a dataset pulled from Kickstarter will skew towards the film category. This could make conclusion #1 above invalid.

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

Please see tables on “Extra Pivots” tab. I don’t know for certain what the spotlight and staff picks columns in the original dataset are for, but I assumed that these are categories where the crowdfunding website might have shown some campaigns up on, say, the front page of the website or in a special category to shout them out. If that was the case, I wanted to see if being in those categories actually increased chances of campaign success. Turns out, not that much of an impact, and very campaigns were in the staff picks category at all.

**Statistical Analysis**

Mean vs. Median: For the number of backers on successful and failed campaigns, I would think the median is a better summary of the data for both sets because both sets have many outliers shown in the box and whisker plots. Since, for both data sets, there are many outliers on the upper end, I think the data is skewed to the left. The mean would be highly affected by the existence of the outliers.

Variability: The backers count of failed campaigns has a lower standard deviation (959.987) than the backers count of successful campaigns (1266.24), so the failed campaigns have overall less variability. I don’t think this is all that surprising because most of the campaigns have goal funding amounts less than 10,000, so I bet the failed campaigns mostly have funded values less than 10,000, while the successful campaigns could have funding values wildly over the goal, adding to outliers.