Tia Scott September 27, 2023

Module #1

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

1a. Crowdfunding at the theater level, whether failed, cancelled, or successful is almost twice as popular and in some cases triple or quadruple as popular as other sources of crowdfunding.

2a. Specifically, plays are by far the most popular source of crowdfunding and have a nearly 55% chance of being successful in funding.

3a. Statistically, once crowdfunding has begun it likely has a 60% chance of reaching a successful outcome and 40% chance of failure across all categories, depending on how you assign cancellations. If you assign cancellations as neither success nor fail, this statement holds true. If you were to assign cancellation as a failure, (as it did not result in a successful campaign) the statistics change slightly to 57% success to 42% failure.

1. What are some limitations of this dataset?

2a. Given that nearly all the data needed to be modified, changed, or have formulas added to it, and given that it is a rather large dataset, it would be hard to easily make assumptions or predictions about the dataset. It is also possible that without clear indication of what is attempted to be interpolated from the data, there could be a wide range of generalizability of the results.

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

3a. One thing I would have done differently would have been to utilize the currency column and to have changed the currency to one single currency instead of the multi-use currency as it is now. I believe the dataset did not convert from Danish Kroner, Pound, and Australian Dollars and has thus skewed the dataset into the current conversion regardless of the true conversion into a single unified currency. The actual single currency, after conversion may be much lower or higher than the predicted rate of pledged funds and average donations. If we had decided not to change the currency in the dataset, I would have used a separate table to more accurately explain the variance in pledges given the exchange rates and location codes. This would have more accurately reflected whether we met the overall goal of pledges.

3b. I would have utilized a table or a heat map for donations globally. This dataset clearly indicates that while the US was statistically more likely to source crowdfunding, there may be other “hot spots” that have a high success rate of funding, even though there were fewer cases. This information could lead to learning what demographic areas are likely to support crowdfunding globally. Or this data may also indicate that while average rates of funding were not only successful but also that funding multipliers exist, therefore being a better demographic area to target in future campaigns. This information would also apply (applying the same logic), if the country had a low rate of success. It would indicate that crowdfunding was likely to not be an efficient way to target those prospects and that it may not be a fruitful venture in the future.

QUESTIONS FROM STATISTICAL ANALYSIS SECTION - ALSO IN THE EXCEL FILE

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

4a. Median is better for summarizing this data. As there are many different types of campaign funding categories and each category has a different success margin, and no cap on number of backers, the median summarizes the outliers more effectively for the large margin of data variables.

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

5a. There is a larger variability with successful campaigns. We can see this in the range between the minimum and the maximum (it is higher in the success category). There is also a larger amount of variance between the two sets and the standard deviation is higher in the successful category. This makes sense to me as there was a higher degree of success to failure from the previous graphs and there was a higher number of backers and amount of successful projects over the amount of failed projects and backers.