Module 1 Challenge

Kelly Hendre

Georgia Tech Virtual Data Analytics Bootcamp

Dr.Arrington

January 30, 2023

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

Based on the provided data, we can conclude that theater campaigns are the most successful types of campaigns; specifically plays. Theater campaigns, specifically plays, are also the most canceled. We can conclude that July is the most successful month when it comes to the success rate of all campaigns, while August contains the most cancelations. America also contains the most campaigns out of all the countries involved.

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

A limitation of the dataset includes the lack of knowledge of donors’ employment status. Knowing this could give some understanding as to why some categories have better funding as opposed to others. One could take this a step further and include the average salary of donors to see if there is correlation between goals and pledges. Another limitation is the lack of knowledge of country-wide dates and common practices. For example, hypothetically speaking, schools in the US begin in August, while many in Europe begin in September, if I am made aware of this, I can form new conclusions as to why a campaign I want to launch does not perform well in August in the US, but well in August in Europe. I could hypothesize that American parents are spending more money on their children’s “back to school” supplies in August, and not spending money on crowd-funded local events, whereas parents and students in Europe are still on vacation, and willing to spend money in August since school does not start until September.

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

We could create box-and-whisker graphs to determine any outliers that could skew data. We can better understand true success, failure, and cancelations among campaigns to determine which campaigns we need to focus time and resources on most. We could also create scatter plots to see the relationship between dates and average donations; adding a trend line would display what time of year people tend to donate most, when donations dwindle, and when they cease. With that information, one could plan things like marketing material distribution more accurately.

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

The data is skewed, so in this case, the median better summarizes the data because it helps us better understand the distribution of the failed and successful backers. Since the data is skewed, we can consider that outliers are involved, and therefore, the mean value could be misleading.

**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 more variability with successful campaigns because their variance and standard deviation are higher than the failed campaigns. This makes sense because successful campaigns met more goal categories than the failed, resulting in a higher chance of risk along the processes.