**GaTech BootCamp – Data Analytics**

**HW1 – Written Report**

***Jason Hanlin***

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

1. All categories and subcategories some level of success. That success appears to be relative to each other. However, theatre, music, and film/video campaigns are much more prevalent than other categories and enjoy more success of a total number of campaigns funded.
2. The ratio of successful to unsuccessful campaigns grows in June and July but shrinks in August and over the winter holidays in November and December. There may possibly be a correlation between holidays and vacations with success but more data is needed to confirm.
3. Generally, crowdfunding goals should be set between $15k and $50k for the greatest chance for success. Goals in this range had a 67% success rate or higher.

*What are some limitations of this dataset?*

The data set appears to be pretty complete and clean however, more investigation could be done to show if there are any unreasonable outliers.

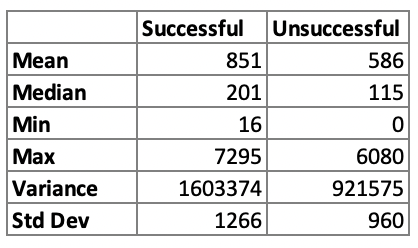
As always, more data could be used. Even though it is a data set with around 1,000 records there is a variety of information that is being evaluated. The last data set occurred in 2020 so more recent data is needed.

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

I would chart success rate against “Staff Pick” or “Spotlight” to see if there is a correlation between the campaigns that are highlighted on the crowdfunding websites. In theory, these campaigns should get more attention and should see a higher success rate.

I would also plot success and failure rate over the period of 2010 to 2020 to see if there are correlation over the life of crowdfunding.

Statistical Analysis

  
*Does the mean or median better summarize the data?*

The median appears to better summarize this data set. The data suggests that there are an average of 851 backers that support a successful crowdfunded campaign. Or an average of 586 backers that support an unsuccessful campaign. And half of the data points occur before 201 backers while half occur after. However, successful campaigns have occurred with up to 7,295 total backers. This represents a pretty big right skewed distribution with data bunched around the low end of backers and a long tail out towards very high numbers of backers suggesting that they are rare. The median is closer to the low end than the average value which suggests it’s a better characterization of where the backer activity is occurring. The same situation applies to the unsuccessful campaign. It appears that there are a lot of backers supporting, or spread out over, a wide variety of campaigns as opposed to most bidders supporting a few campaigns.

It also appears that not much can be drawn from correlation of this data, or this method for analyzing the data, between the number of backers and whether the campaign is successful or unsuccessful. It appears that all campaigns are being supported but it is more dependent on the where the goal is set. However, since most of the data supports successful campaigns, I think we can conclude that the crowdfunding structure, or model, is successful. Funding success even with campaigns that have small number of backers.

*Is there more variability with Successful or Unsuccessful campaigns? Why or why not?*

There appears to be slightly more variability with successful campaigns than for unsuccessful according to Standard Deviation values. According to the metric, 95% of the data falls within plus and minus 2,532 backers for successful campaigns versus within plus and minus 1,920 backers for unsuccessful campaigns.