Tye Smith

Crowd Funding Analysis

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

Those who set a goal below 15000 saw an average failure rate of 31%, and those above 35000 had an average failure rate of 50%. Setting a goal not to high, but not to low in the 15000 to 34999 range showed a greater success rate and a significant drop in percentage of failure at 12%.

The Sub-categories that saw the most success include Rock, Plays, and Documentary. The largest population existing in plays (319) with a 58.6% success rate, Rock (79) with a 62% success rate, Documentary (55) with a 62 percent success rate. I would focus in on these categories for the highest probability of success.

In the three categories mentioned above, the best time to launch can be seen in the timeline date by filtering down to “Music”, “Theater”, and “Film & Video” which are the parent categories of the subcategories mentioned.

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

In the categories we see the most success our population is the smallest, this I believe limits the level of reliability of our data. With a population so small we have to rely on the percentages of outcome, which could be miss leading in a bigger population. Another thing to consider is the type of “plays”, “Documentary” “Rock Music”, this could greatly impact success rate as well.

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

Knowing the average amount given by the backers per backer for each category could be useful. This would allow us to know where to set future financial goals in a campaign trying to raise money. To do this you could create a pivot table that has parent categories in rows, with average amount given by backers per that category.

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

The Median best represents the data as it is more resistant to the outlier’s in the data that could otherwise skew the data if we used the mean. The range of the data sets are so broad that a few scores at the high end or low end can greatly affect the mean, the median fines the natural middle without being pulled by a few high or low scores (Outliers).

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

When you look at the overall data there is more variability with those who have had success then those that have been unsuccessful. This makes complete sense as the number of backers to someone who is successful will be limitless in how high it can go. While in unsuccessful campaigns one thing should remain common, the number of backers was not enough to have success, this can be seen in the whisker plot.