**Crowdfunding Report**

*Alexandra de Roos*

Based on the data provided in the Crowdfunding workbook, over 50% of the campaigns were successful with theater parent companies having the most successful campaigns. While theater parent companies had the most successful campaigns, they also had the most failed campaigns and the most canceled campaigns. This is probably due to the fact that theater companies had the most campaigns overall, with almost double the next most campaigns.

Another conclusion from the data is that the most successful sub-category was plays. When filtered by parent company, all the play campaigns are in the theater category. Since theater had the most campaigns, along with the most successes, fails, and cancels, it would make sense that any subcategory under theater would have similar numbers.

A third conclusion is that there is no clear benefit to having a campaign created in a given month or year. All the years had relatively the same amount of campaigns and the same amount of successful campaigns. Between the months, the campaigns are within range of each other.

One limitation of the dataset is how many people were interested in a campaign, but didn’t donate. A campaign can be interesting to someone, but they can’t afford to donate or there aren’t many donors, so they are discouraged from donating. Another limitation is how the campaigns were marketed. Some campaigns were very successful getting over 200% of their goal, while some got none of their goal. How were the campaigns marketed. There could be a data collection about how the campaigns were marketed (social media, emails, newspapers)

The graphs right now show the outcome of a campaign. There could be a graph for the percent funded. This way one could see how close a campaign was to being successful or how much a campaign went over. Another optional graph could be how the average donation to a campaign compares across industries. Maybe some industries have more donors or more money from the average donor. That could give answers to why some campaigns were more successful than others.

Based on the data from the number of backers in each campaign, either successful or not, it seems that the median is a better measure. Median is a robust measure, so it is not affected by outliers or extreme values in a dataset. Both of these data sets appear to have a lot of variability and many outliers. Many of the values are three standard deviations away from the mean in both the successful and failed campaigns. Median is not affected by the values of the numbers, it is the number in the middle. Median would be a better indicator of the central tendency of both datasets.

From the datasets, successful campaigns have more variability than failed campaigns. There are more outliers in the box and whisker plots. More important, the standard deviation is bigger in the successful campaigns. Standard deviation is the average amount of variation a value is from the mean. Higher standard deviation means a value would be farther from the mean. If a value is farther from the mean then it is more variable. It makes sense why there is more variability in successful campaigns. If a campaign is really popular and already funded then people may be more likely to want to add to the fund. Sometimes a campaign doesn’t have guaranteed rewards if it is not fully funded. An already successful campaign takes away that risk and more people will want to fund it if they want what’s in the campaign. This can lead to a lot of variation in the amount of backers for a successful campaign. A failed campaign means not enough people funded the project. A campaign isn’t funded by an amount of people, but by money. Fewer people donating to a campaign means more money per person to fund a campaign. This would make it more likely that there is less variability because the campaign isn’t getting enough people to donate.