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

The first conclusion we can draw about crowdfunding campaigns is that majority of total campaigns are in the performing arts: (Theatre, Music and film and video). This makes sense as it is probably hard to get signed by a professional performing arts company. Since this is the case, a person’s, or group’s best option is to use crowdfunding to get their art up and running. Then hopefully they will create something that gets them noticed by a professional theatre company, production company or record label. By looking at the sub-category data you can see that an overwhelming number of projects are for plays. This makes sense as they would probably cost a lot of money between renting space to practice, costumes, set pieces and hiring actors/stage crew, etc. whereas you can record music on a phone and create a movie with one camera if you had too.

A second conclusion we can draw from this data is that campaigns are not overwhelmingly successful. There are more successful campaigns then not but only campaigns from the film and video category have a 2:1 ratio of successful campaigns to failed campaigns. So, it must be hard convincing people on the internet to invest.

A third conclusion is that most of these campaigns occur in the United States suggesting this crowdfunding website was created and operates in the United States. This suggests that this funding idea was started in the United States and has just recently been gaining in popularity. In due time the numbers for campaigns around the world should be on even ground with the United States.

What are some limitations of this dataset?

The dataset does not show how each campaign marketed itself to potential funders. Each group or individual most likely created a video or written pitch to hopefully obtain funding. This probably plays the biggest role in whether a campaign was to successfully reach its goal. I would hypothesize that the better the pitch a group gave the more likely they were able to hit their goal. Another limitation is that we do not know how much each group advertised their campaign. Some groups may have been using social media to bring people to their crowdfunding campaign. The groups that advertised the most often and most effectively would most likely be more successful than groups that did not utilize social media.

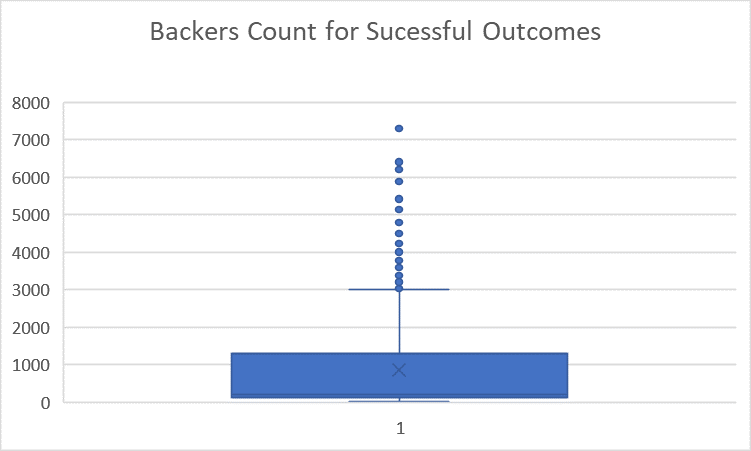
The data set also does not show us when the donations are made. If it did, we could aggregate it to show us a count of donations per day for each campaign. That way we could potentially see a trend shows when most of the donations come in or a realistic time frame campaign to reach its goal.

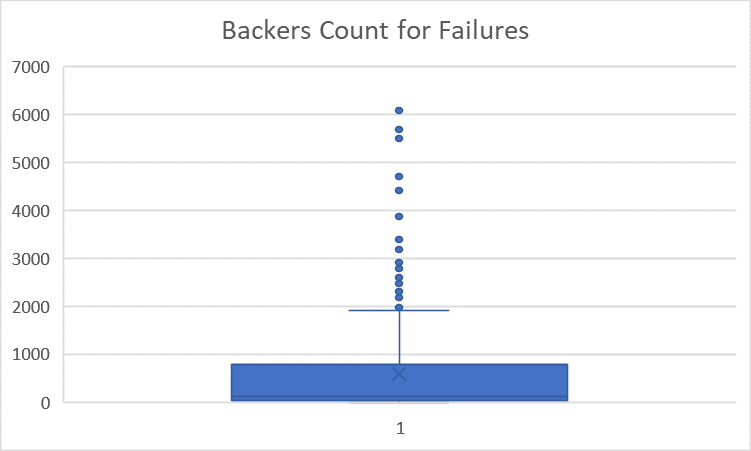
What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

We could create a table and bar graph showing the outcomes based on whether the project was spotlighted or not. That could show us if it was worthwhile for a campaign to be spotlighted and whether campaigns should seek to get spotlighted in the future. We could also take the difference between the project created date and deadline. That could potentially find a common time frame between successful campaigns and give future campaigns an idea of how much of a time frame they will need to hit their goal.

Statistical Analysis

When deciding whether the mean or median is better to summarize the data you have look at the distribution of the data. If the data is symmetrical or normally distributed, then the mean is better to describe the data. When the data is skewed, the mean will be pulled in one direction or the other by outliers. In this case the median would be best to describe the data. By creating box and whisker plots, you can see that in both cases the data looks fairly symmetrical around the mean so I would say the mean is best to summarize this data.





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

Judging by the standard deviation and variance of each data set, successful outcomes have more variability than failed outcomes. This makes sense for failed outcomes because typically they fail because they a) did not get enough backers or b) not enough funding. For successful outcomes however, there is going to be instances where they did not get a lot of backers but the backers, they did get gave a lot of money or they just get a ton of backers. This is evident in that the minimum number of backers for a successful campaign is 16 and the maximum is 7,295.