Crowdfunding Analysis

One conclusion that can be drawn is that over the summer months of June and July crowdfunding has a higher success rate and the month of August has the highest failure rate. This can possibly give insight to the best time of the year to crowdfund with the best probability of success while also showing which months to avoid. A second is that the most popular category to crowdfund is theater followed by Film & video and music. Using this knowledge, you can see that these categories have a much larger number of successes, but their percentage of successes compared to the total number of projects is relatively the same as every other category. This shows that the category that is crowdfunding does not have much effect on the success or failure of the project. Finally, the overall percentage of successful crowdfunding campaigns is 56.5% in our dataset. If you look at the individual countries that are included in the dataset, the percentage of successful projects varies from 50 to 58% which is very close to the mean of all of the included countries. This means that whichever country the crowdfunding campaign begins in, the outcome has relatively the same chance of success.

Some limitations that come from using this dataset are that it only includes crowdfunding campaigns up until 2019 which is a year before covid-19 hit so this dataset does not include possible changes that could have occurred because of that factor. Also, this dataset does not show the efforts these people made to accrue funding. The methods used could have influenced whether or not they were able to gain funding and could provide insightful information on what other should do. As well as converting all of the currencies to the same currency such as USD to more easily compare the funding goal to every project’s outcome. Lastly, I believe having more information about the founder’s support system, motivations, and experience level would be helpful in order to see some other confounding factors and see if they play a role in the success rate.

Other graphs that could be created to further break down the dataset would be creating a histogram with the length of time the project was open compared to the outcome which would show me whether a campaign had a higher probability of success if it was open for a longer period of time. I would also have liked to create a pivot table with columns as outcome, rows as parent category, and values as goal (we would need to convert all the goal currencies to the same currency). This would provide insight on how the goal affected the outcome and which categories required more funding and could be a useful guide for someone trying to start a crowdfunding campaign.

Statistical Analysis

The median is a much better way to summarize the data than the mean because the data for both success and failures is strongly left skewed. There are multiple outliers in both datasets which when calculating the mean, it increases the value. The median is a way to calculate an average of the data while disregarding any skew or outliers that could pull the number away from the true average.

Success has a much larger variability than failure. You can tell by looking at both the variance and the standard deviation. This makes sense because the more backer a campaign has the higher chances of success so there are not going to be as many large numbers for backers under the criteria fail. Because of this the maximum is a lot lower for failure than success and with that the variance and standard deviation are less as well.