Matthew Channell

Vandy Data Analytics Boot Camp

11 August 2023

Crowdfunding Written Report

Given the provided data, here are three conclusions we can draw from these specific crowdfunding campaigns:

1. The two-month span from June to July had a higher success rate for crowd funding relative the average success rate over each month throughout the year.
2. The largest parent category within this set of data was that of theater crowdfunding, with 150 more projects for crowdfunding than film and video, the second most popular.
3. Within subcategories of this data, film was by far and away the most popular with just shy of 350 crowdfunding projects, while no other subcategory had over 100.

However, drawing these conclusions only apply to this data set alone, and not crowdfunding in general due to the limitations of the data. This data set only has 7 countries included and leaves out a vast array of other crowdfunding projects in different areas of the world. This data set also only started occurring data starting in January of 2010, leaving anything before this time not included in the data as well.

Finally, another table/graph that would be useful would be one that compares the goal of the crowdfunding project, the average donation as well as if it was successful of not. Looking at the data, it seems to be a trend that the extremely large goals seemed to fail more often, and the extremely small goals seem to be more successful, baring a few outliers that is. None the less, this table would provide another way to look at the data of crowdfunding.

Statistical Analysis

When it comes to summarizing the data, in this set of data relative to the number of backers for a crowd funding project, the mode is a better tool to better understand the set. In most cases, the mean is better to use to get an understanding of data, however that is in cases with normal distributions. In this case, we have something called a positively skewed distribution, meaning that the larger backer counts, for both failed and successful, pull the bell curve to the right of where the true center of the data is. Also, looking at the standard deviation of successful and failed campaigns, it can be concluded that successful campaigns have high variability within them. The minimum values are almost the same in both while the largest successful campaign has over 1000 backers more than the largest failed campaign. This goes to show why the standard deviation and variance are higher in the successful campaigns as well.