An Analysis of the Crowdfunding Excel Data

Crowdfunding has become a popular means of gathering funds in order to launch new projects across a variety of projects. While some projects were able to exceed goals many times over, other projects fail to reach their selected goal, and thus fail to receive proper funding. We have been provided with some sample data, and are tasked with analyzing this data in order to identify trends in the data.

The bar chart shown above shows a relationship between project category and the level of success in reaching the threshold for funding. Theater projects had the largest number of crowdfunding campaigns, and thereby the greatest number of successful campaigns. The categories that had the next greatest number of campaigns were Music and the combined Film & Video category, each had about half as many campaigns in comparison to theater, as well as in number of successes. Technology and Publishing each appear to have performed strongly in comparison to the number of crowdfunding campaigns.

Our next bar chart shows the relationship between the project sub-category and the number of successes. Plays had by far the greatest number of crowdfunding campaigns, and the greatest number of successful and unsuccessful crowdfunding campaigns. A closer look at the original data shows that every item in the Theater category fell in the Plays subcategory. Rock has the next greatest number of crowdfunding campaigns, and the next greatest number of successful campaigns.

The above line chart shows a breakdown of the number of successful campaigns in relation to the month that the campaign occurred. June and July have the greatest number of successful campaigns, while there is a drop in the number of successes throughout the fall and into the winter months. The fall months also have the fewest number of failed campaigns. April through June appears to have the smallest number of cancelled campaigns, but no individual month had 10 failed crowdfunding campaigns.

One limitation of each of the above graphs is that it does not clearly show the percentage of successful campaigns. A solution to this would be to have parallel graphs that show a breakdown by a percentage success/failure rate rather than total amount. It would be helpful to also have on this graph the true percentage of successful and failed campaigns so that we can determine if there is any category, sub-category, or time frame that outperforms the population success rate. This would allow us to determine if there is any category that has a higher success rate that has to this point been under-targeted.

Another limitation is that this lumps together data across a 10-year time frame. Having a graph that analyzes the number of successes by category or sub-category as compared to the year that the project took place may be helpful in identifying if there is an emerging field in which projects have been successful recently, to predict which fields are more likely to be successful in the near future. This would also be helpful as it would provide information as to if there is something that was initially successful but has recently suffered from a larger proportion of failed campaigns.

A third limitation to this data set is specifically in relation to the second chart. Plays overwhelmingly outpaces the other sub-categories because Plays are the only sub-category within the Theater category. It would be helpful to see a chart with Plays are not included so that we can better analyze the remaining sub-categories against each other more easily. This would help us to see which other categories have performed strongly in relation to the number of success and total campaigns.

|  |  |  |
| --- | --- | --- |
|  | Successful | Failed |
| Mean | 851.146903 | 585.615385 |
| Quartile 1 | 128 | 38 |
| Median | 201 | 114.5 |
| Quartile 3 | 1280 | 784.5 |
| Minimum | 16 | 0 |
| Maximum | 7295 | 6080 |
| Variance | 1603373.73 | 921574.682 |
| Standard Deviation | 1266.24395 | 959.986813 |

The above chart shows a statistical analysis of the number of backers in relation to the success or failure of projects. In both instances the data is heavily skewed to the right, as the maximum in both situations is more than 3 standard deviations above the mean, indicating that there are outliers present in both data set. In such situation, the median is the better measure of center, because we are guaranteed to know if half of the data lies above the median, whereas less than half the data lies above the mean. As such, the median shows the typical number of backers for successful and failed projects.

In measuring the variability of projects, we would analyze the variance and the standard deviation. The standard deviation of the successful projects exceeds the standard deviation of failed projects by over 300 backers, indicating that successful projects have greater variability. The variance of successful projects exceeds that of failed projects by more than 600,000 units, further showing that successful projects have greater variability. This makes sense, as it is possible that projects can be funded by a few well-to-do benefactors, or for successful projects to be funded by a large number of patrons, whereas it is less likely for a large number of people to contribute to a project and that project end up failing.