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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Count of outcome** | **Column Labels** |  |  |  |  |
| **Row Labels** | **canceled** | **failed** | **live** | **successful** | **Grand Total** |
| film & video | 11 | 60 | 5 | 102 | 178 |
| food | 4 | 20 |  | 22 | 46 |
| games | 1 | 23 | 3 | 21 | 48 |
| journalism |  |  |  | 4 | 4 |
| music | 10 | 66 |  | 99 | 175 |
| photography | 4 | 11 | 1 | 26 | 42 |
| publishing | 2 | 24 | 1 | 40 | 67 |
| technology | 2 | 28 | 2 | 64 | 96 |
| theater | 23 | 132 | 2 | 187 | 344 |
| **Grand Total** | **57** | **364** | **14** | **565** | **1000** |

One conclusion we can draw from crowdfunding campaigns is that around half of crowdfunding campaigns will succeed overall. What we need to realize though is that there is more to this than the general data set tells us. Based on 1000 total campaigns, about 565 of them succeeded, 364 failed, 57 were canceled, while 14 are still live. While one can look at this and believe their campaign has a bit over 50% chance of success or failure, we can take the data a step further and analyze the different types of categories that each campaign falls into and see their percentage of success. In terms of general categories, there were high numbers of crowdfunding campaigns for theater, music, and film & video, with plays as the largest number of campaigns looking at the sub-categories. What this tells me is that a good portion of crowdfunded campaigns revolve around the arts.

When I saw this, I took in some context with the US portion of the data, as we can see how much funding the arts get on a scholastic level. For example, at my school, our music department decreased in size due to lack of funds, and we couldn’t get a lot of new equipment when it came to art supplies for classes and clubs. It’s important to take into consideration the context and demographics of where these campaigns are being conducted. Based on this data, if I were to start a music campaign in the US, I can see that it may have about 56% chance of success, but in the DK, while the sample size is relatively small, only 1 out of 6 music campaigns succeeded.

Continuing with this idea, we can also filter the data to show us the success of each campaign based on their outcome goal. While one would look at this graph and see a 100% success rate for campaigns that have an end goal between $15000 and $20000, we can’t take this graph at face value, which leads us into the second conclusion in that just looking at a graph can be deceiving. The graph suggests that there is a 100% success rate when crowdfunding campaigns set their goals somewhere between $15000 to about $20000. Looking at that, someone could say “Wow, that means I should start a crowdfunding campaign and aim somewhere between $15000 and $20000, and everything will work out” but what the graph doesn’t show is the number of data points that make up that range compared to other points of the graph. For example, while this shows a 100% success rate, what isn’t shown is that for this data set, there were only 17 out of 986 that fit in that range. Compare that to a lower success percentage, like the range of $1000 to $9999, which success percentage fluctuates with a low of 52%, but a high of 83% success rate. While this percentage doesn’t provide as much confidence as seeing the data points in the 100% range, we can note that this data range shows most of the number of campaigns, a combined 546 projects out of the listed 986. I do not believe we have enough data to support a continued 100% success rate when campaigns are at $15000 and above, but we can see that many projects in the 1000 to 9999 range have had their fair share of success, and that degree of probability across a larger sample size is something I would have more confidence in.

Another conclusion we can take note of based on the data is that there appeared to be a slight increase in success and dip in failure of a crowdfunded campaign if it was set up in the middle of the year. June and July saw the biggest deviation from a successful campaign to a failed campaign. Why is that? That is something that we should further investigate, the context around the middle of the year. Is there a reason why more people are willing to participate and donate to a crowdfunding project during these months versus any other time of the year? Based on the data there are two specific categories that really push that narrative, which are music and theater. We can also ask if the length in time of each project will influence its success.

What are some limitations of this dataset?

One limitation that I want to take note of is that lack of data regarding demographics. While we have the countries for where the campaigns started, there’s no information regarding where specifically in each country the backing comes from. For example, we can highlight that there was a high number of plays being crowdfunded. In a place such as New York, where historically people have gone to chase their dreams of acting, people may be more inclined to support the arts and the crowdfunding of plays, whereas a different state such as Arizona, which is one of the states who funds the arts the least, may have a local population that wouldn’t be as supportive. There’s a lot of contextual factors that play a role in this, but having a better understanding of who the supporters are would greatly enhance the usability of this dataset.

Another limitation is not knowing the means by which some of these crowdfunding campaigns went about spreading their message and their project. There could be some parameters that each successful campaign followed and based on that; would we be able to utilize that data to better understand why some of these crowdfunding campaigns failed? For example, someone with a successful campaign, who was able to reach over 100% of their goal, could have been utilizing all means of social media to get the word out, or they could have gone to local businesses and advertised there, or if it were for a play, to a school for arts which may have a more supportive population. Having the means in which their projects were advertised, and the strategies they used to reach their goals would be a good set of data to use to better understand why some of these projects ended up failing or never really had a chance to start.

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

We could create a Box and Whisker graph or a graph that would show how skewed the data is. On the last portion of the challenge, we calculated the mean, median, minimum, maximum, variance, and standard deviation. With this we can see if there are any outliers and where the data is more centralized. This can be utilized for outcome by category or a step further and with sub-categories, giving it more than just the number of backers we have vs success or failure.

Something I would have liked to personally explore if I were working on this project and creating this data set would be to find out the tactics that each crowdfunding campaign utilized in order to reach their goal. There are a lot of ways to get the word out about the campaign from word of mouth to on the internet through social media. Creating a data set that compares the outcomes of a campaign to the tactics used may paint a better picture of how the results changed for each crowdfunding campaign.

Use your data to determine whether the mean or the median better summarizes the data.

Between the mean and the median, the median may be better for summarizing the data in this case. To focus on the successful campaigns, at first glance we see that 201 is the median or the middle of the data set, but the max is set at 7295. Then when you check the mean, which is approximately 851, one may see it as closer to the max number and choose that. What one must realize, however, is that the mean and the median are not close to each other, with the mean being larger than the median. Even without looking at a distribution graph of these data points, the wide difference between the mean and median indicates a positively skewed data set and not a normal distribution. Using the median will give us a better idea of the center of the data points and will provide us with a good place to spot outliers in the data points, utilizing the variance and the standard deviation calculated from the data.

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

Based on the formulas utilized to find the variance of the number of backers and the standard deviation of the number of backers for both successful and unsuccessful campaigns, one can see that there is more variability with successful campaigns. This makes sense because the data points are so widespread. If the mean and median were close to each other, you would see a much lower number in terms of variance. If that were the case, we would have seen a more normal distributed graph, but in this case, the minimum number is 16, while the maximum is 7295. There is a wider gap between these numbers than the minimum and maximum of the unsuccessful campaigns. Either way, with the variance being so high for both campaigns, it is harder to make a solid prediction based on the data.