**Crowdfunding Analysis**

Analyzing the given crowdfunding database of 1000 values from 2010 to 2020 it is possible to conclude:

1. 57% of the projects were successful, and 36% have failed. Considering these numbers, it would be interesting to combine other ways to receive funding for projects. The failure risk is relatively high.
2. Crowdfunding is mainly used to obtain funding for projects whose goals are less than 9999 or more than 50000.
3. Technology and photography had the highest success percentages, 66.7%, and 61.9%, respectively.
4. Theater is the most relevant category, representing 34% of all categories, followed by film & video and music, representing 18% of the projects each.
5. Journalism is the least expressive category, totaling only 4 projects in 10 years.
6. In general, it is possible to identify a successful seasonality in June and July, however, it is followed by a decrease in August.

There are some limitations on the dataset. It could collect more information regarding the backers since they are the ones responsible for the success or failure of the projects.

Adding to this topic, there are data that are hard to work with, like staff\_pick, and spotlight. They need to be better collected, then it would be possible to use them to quantify the variables that they represent.

Other possible analyses can be done using this dataset. For example, it would be interesting to understand the data per country, since most of the projects are related to culture, it could help to check if there are any other trends taking culture into consideration.

Checking the data against the total is very important also, this way it is possible to weigh the numbers in a better way. For example, when we analyze the outcomes per category, I would create a table calculating the % of successful projects per category and the % of failed projects per category so we can compare which categories would be the most successful and the most unsuccessful, like the following table:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **canceled** | **failed** | **live** | **successful** | **Grand Total** | **%Category** | **%Successful/Category** | **%Failed/Category** |
| film & video | 11 | 60 | 5 | 102 | **178** | 18% | 57.3% | 33.7% |
| food | 4 | 20 |  | 22 | **46** | 5% | 47.8% | 43.5% |
| games | 1 | 23 | 3 | 21 | **48** | 5% | 43.8% | 47.9% |
| journalism |  |  |  | 4 | **4** | 0% | 100.0% | 0.0% |
| music | 10 | 66 |  | 99 | **175** | 18% | 56.6% | 37.7% |
| photography | 4 | 11 | 1 | 26 | **42** | 4% | 61.9% | 26.2% |
| publishing | 2 | 24 | 1 | 40 | **67** | 7% | 59.7% | 35.8% |
| technology | 2 | 28 | 2 | 64 | **96** | 10% | 66.7% | 29.2% |
| theater | 23 | 132 | 2 | 187 | **344** | 34% | 54.4% | 38.4% |
| **Grand Total** | **57** | **364** | **14** | **565** | **1000** |  | **56.5%** | **36.4%** |

For a better understanding of the statistics, I would create boxplot graphs to visualize the variation of the data, like the following:

