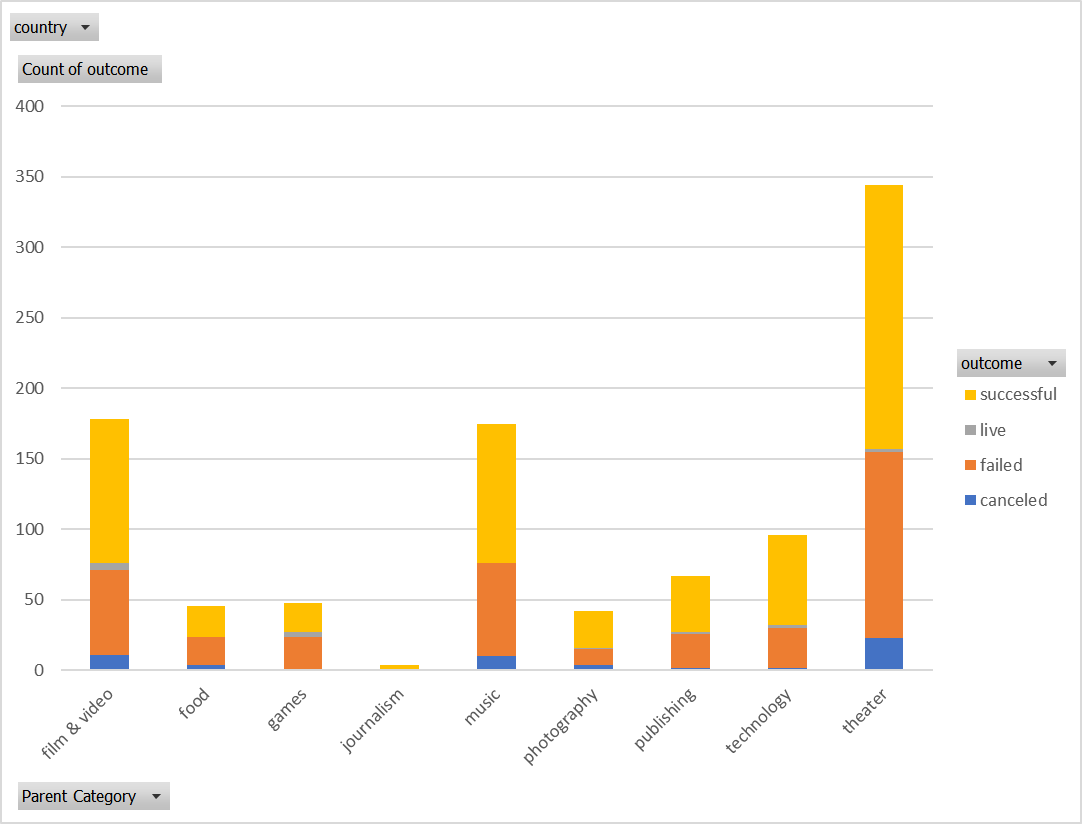
Question 1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. 69.7% of the projects are related to visual and audio entertainment, ie, Parent Category: film & video, music, theater. Theater is the most popular category, 34.4% projects are about theater/plays, astoundingly higher than other subcategories. Amoun all Parent Categories, Journalism is the least popular, with only 4 projects in the sample.



1. 56.5% of projects are successful, 36.4% are failed. Being the most popular projects also bring the highest fail count and successful count to category theater and theater/plays. However, the successful rate for theater is 54.36%, close to average. More analytic work is required to conclude if this is statistically significant. The launching date didn’t show an obvious impact on success rate.

| Count of outcome | Column Labels | |  |  |
| --- | --- | --- | --- | --- |
| Row Labels | canceled | failed | live | successful |
| film & video | 6.18% | 33.71% | 2.81% | 57.30% |
| food | 8.70% | 43.48% | 0.00% | 47.83% |
| games | 2.08% | 47.92% | 6.25% | 43.75% |
| journalism | 0.00% | 0.00% | 0.00% | 100.00% |
| music | 5.71% | 37.71% | 0.00% | 56.57% |
| photography | 9.52% | 26.19% | 2.38% | 61.90% |
| publishing | 2.99% | 35.82% | 1.49% | 59.70% |
| technology | 2.08% | 29.17% | 2.08% | 66.67% |
| theater | 6.69% | 38.37% | 0.58% | 54.36% |

1. Less popular categories - journalism, photography, publishing, technology - show higher success rate, as 64.11%, comparte to overall 56.5%. Cancelation and failure is below average.

Question 2. What are some limitations of this dataset?

1. There are 7 currencies in this dataset, it’s improper to compare value related content without specifying if the values are converted or not.
2. Based on the id, this sample is unlikely to be randomly selected from the population. It’s insufficient to draw any conclusion for the general population.
3. Category only reveals the format of the projects, subjects could also be a great influencer, which is not listed in the data.

Question 3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

1. Pivot table for Category and Average Donation shows less launched projects, such as games and publishing, for the group interested in those projects, they are likely to donate more. Cancelled projects showing higher average donations compared to successful projects is very interesting, but no conclusions for now.

| **Row Labels** | **canceled** | **failed** | **live** | **successful** | **Grand Total** |
| --- | --- | --- | --- | --- | --- |
| film & video | $ 74.11 | $ 61.33 | $ 89.29 | $ 69.27 | $ 67.46 |
| food | $ 77.06 | $ 63.82 |  | $ 70.77 | $ 68.29 |
| games | $ 25.01 | $ 70.38 | $ 66.92 | $ 73.81 | $ 70.72 |
| journalism |  |  |  | $ 30.49 | $ 30.49 |
| music | $ 60.98 | $ 59.99 |  | $ 67.59 | $ 64.35 |
| photography | $ 79.37 | $ 50.34 | $ 32.02 | $ 69.87 | $ 64.76 |
| publishing | $ 74.34 | $ 66.16 | $ 50.64 | $ 79.87 | $ 74.36 |
| technology | $ 41.88 | $ 52.89 | $ 59.70 | $ 70.41 | $ 64.48 |
| theater | $ 73.19 | $ 69.55 | $ 71.08 | $ 68.03 | $ 68.98 |
| **Grand Total** | **$ 70.03** | **$ 64.11** | **$ 70.82** | **$ 69.43** | **$ 67.55** |

1. 76.3% of the projects are launched in the US, showing the majority of users could be US based. But this conclusion highly depends on how sampling is done.

| **Count of outcome** | **Column Labels** |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Row Labels** | **canceled** | **failed** | **live** | **successful** | **Grand Total** |
| AU | 2 | 16 | 1 | 24 | 43 |
| CA | 2 | 19 | 1 | 22 | 44 |
| CH | 4 | 6 | 1 | 12 | 23 |
| DK | 1 | 12 | 1 | 17 | 31 |
| GB | 1 | 18 | 1 | 28 | 48 |
| IT | 3 | 19 |  | 26 | 48 |
| US | 44 | 274 | 9 | 436 | 763 |
| **Grand Total** | **57** | **364** | **14** | **565** | **1000** |

Statistical Analysis

Question 1. A brief and compelling justification of whether the mean or median better summarizes the data.

For outcount - successful and failed,

* Both groups show large variance between median and mean.
* Both group has maximum > mean + 5 \* standard deviation.
* Both group has minimum < mean - 1 \* standard deviation.

The data is highly skewed. In this case, median will be a better choice to summarize the data.