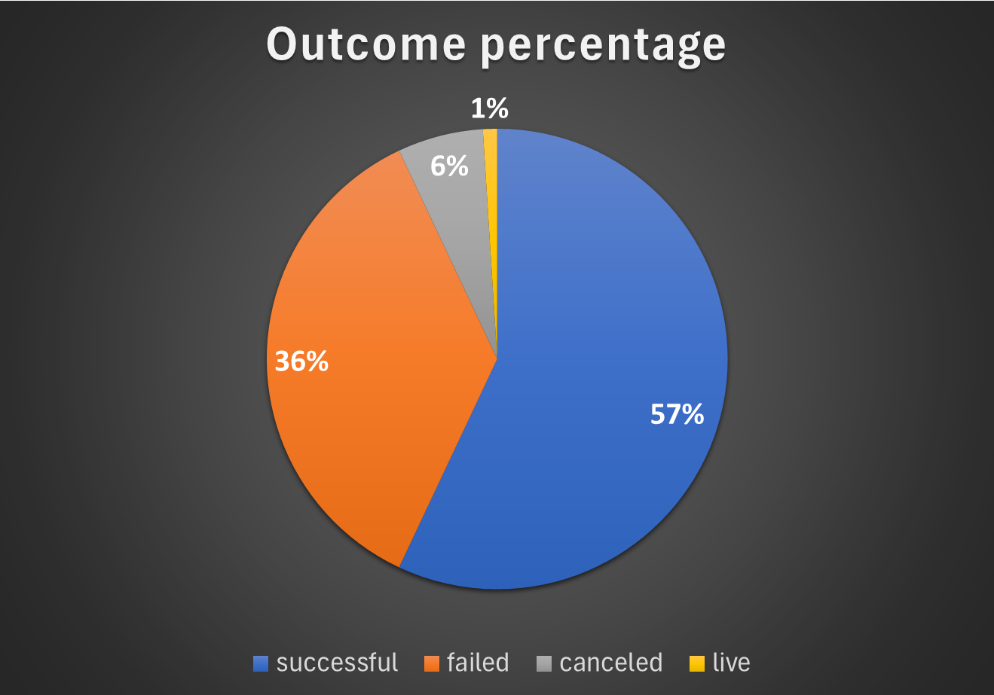
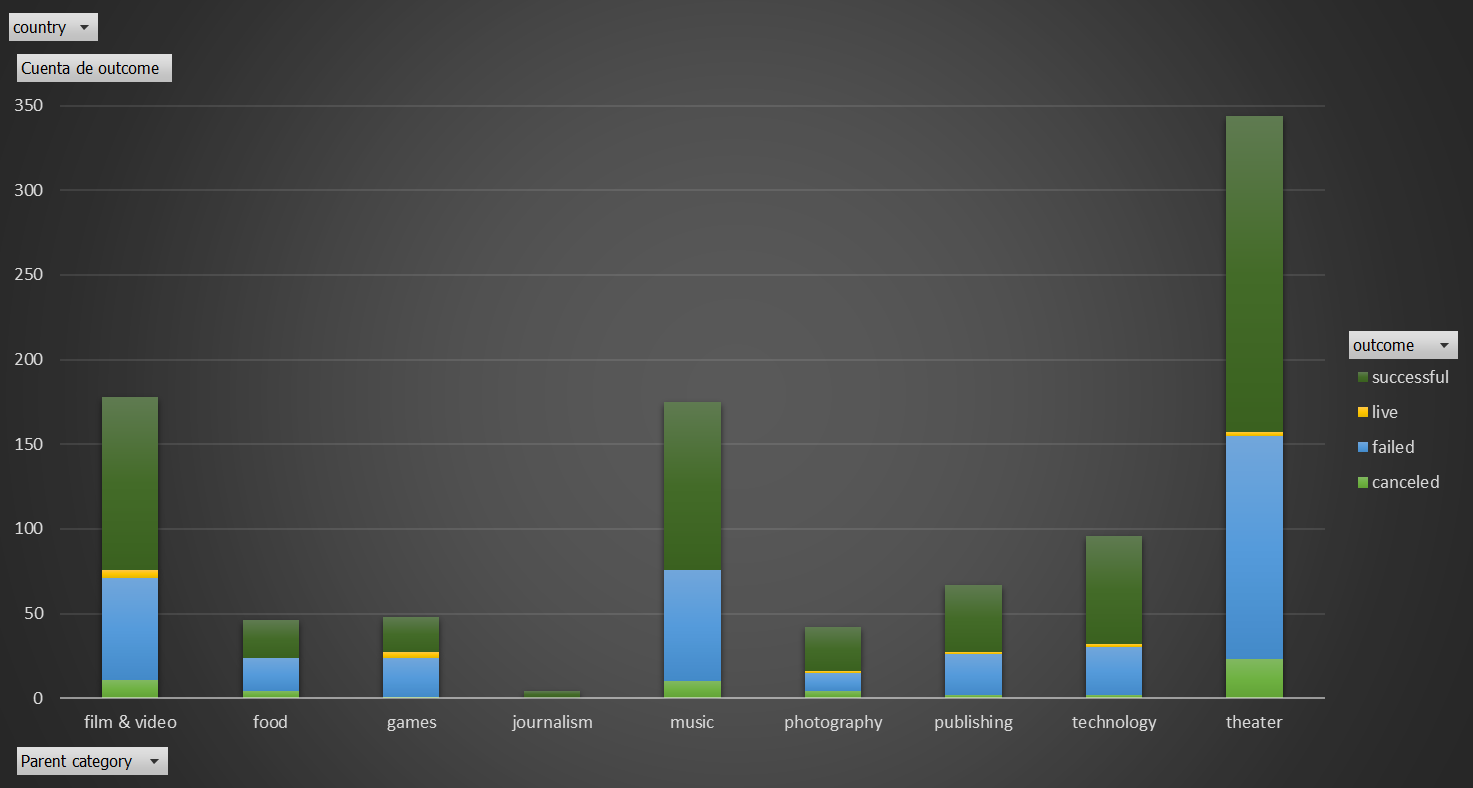
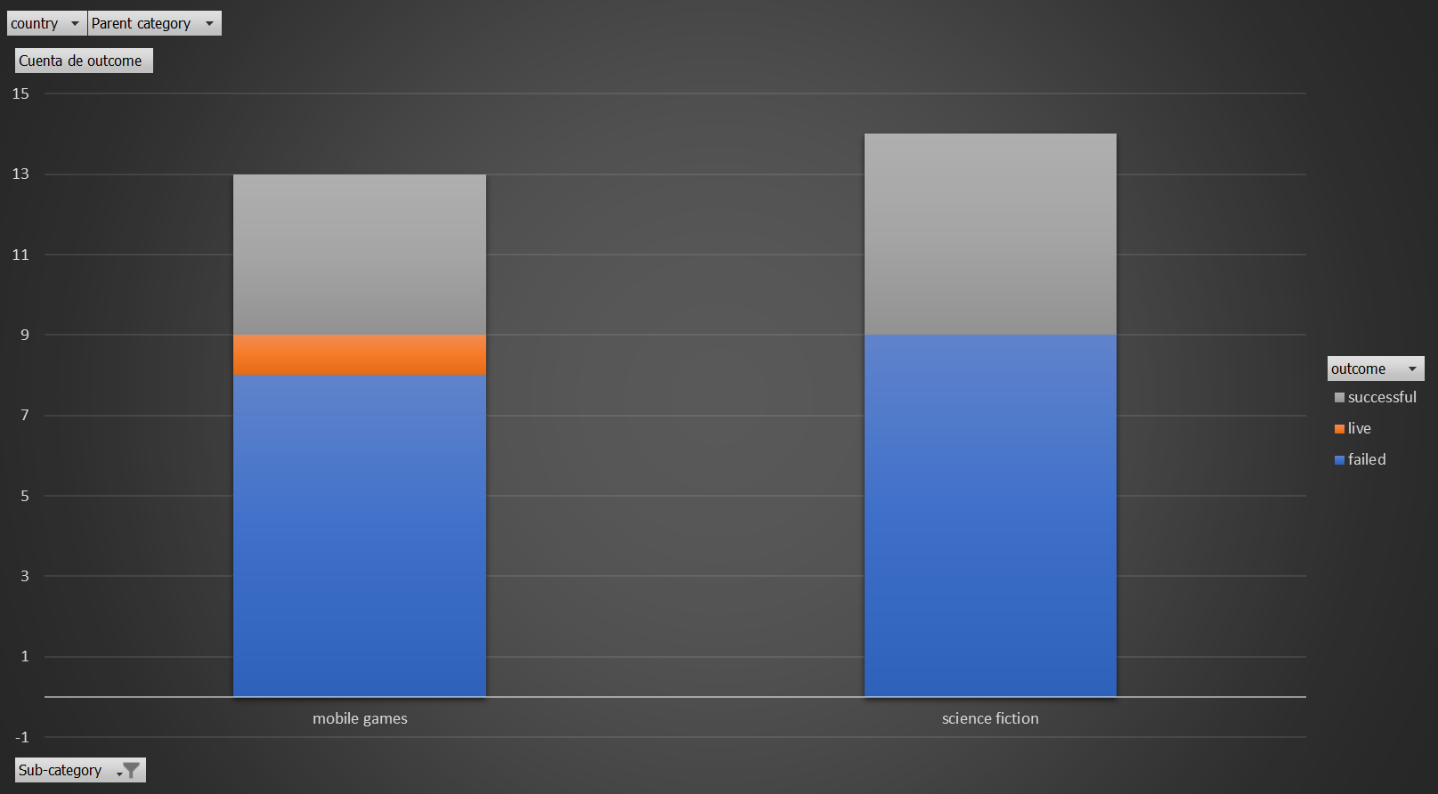
1. **Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**
2. Most of the campaigns (58%) have been successful. 36% of them have failed at some point and only 6% of the campaigns got cancelled.



1. Theater is the most popular category and has the most funding resources. However, the top 3 categories with the highest successful rate are: Journalism (100%), technology (67%) and photography (62%).



1. There are only two sub-categories with a higher failure rate than a successful rate: Mobile games and science fiction. These two sub-categories would be hard to fund in the future.



1. **What are some limitations of this dataset?**

Here are some limitations of this dataset:

1. Quality of data: Although there are no missing values, we are missing some relevant data than can provide additional value to our analysis.
2. Skewed sample: As we could see, the variance in this dataset is so much big. Therefore, we couldn’t get a normal distribution and perhaps our analysis will be affected by some bias or outliers in the dataset.
3. Lack of context: There are some columns that we did not get information about. Perhaps we can infer meaning or context, but that in data is a problem, because doing that means we are adding a bias on our analysis.
4. **What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

Some other charts that can provide additional value is the outcome percentage (the chart I am showing in the first question). We could see how many of them have been successful. And then we can go deep into the most successful/failed categories and start a deeper analysis on the root causes. A Pie chart is great for this.

Another one would be a chart outcome vs percent funded. I realized that a percentage funded less than 100% were all failed campaigns whereas the successful campaigns were all more than 100% funded. In other words, all the successful campaigns reached more money than the goal established.

I also would like to see a chart by category on the goals trend. Probably, some campaigns failed because the goal they were trying to reach was low for that specific category. Maybe the goal should have been higher and that’s a possibility of why they failed or succeed.

**Part II**

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

The median is the statistic that better summarizes the data. As data for both type of outcome (successful or failed) is very dispersed and we do not have a normal distribution, using the median is a better option for this type of cases.

Another point to reinforce this thought is through what a box chart shows us. There are too many outliers that affect the result of the average. Therefore, it is recommended not using average for this analysis.

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

There is more variability with successful campaigns. It makes sense because we can see the variance value, that is higher in the successful campaigns than in the unsuccessful campaigns. And visually, on a scatter plot we can see how much spread the values are. It also makes sense because we have more data on the successful campaigns, and the difference between max and min values is also higher than the unsuccessful campaigns.

Tabla

Descripción generada automáticamente