DATASET ANALYSIS

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

A screenshot of a graph

Description automatically generated

There were 9 categories (film & video, food, games, journalism, music, photography, publishing, technology, and theater) of projects in the Crowdfunding campaign. Theater showed the highest rate of being successful but also failed and canceled campaigns (187, 132, and 23 counts respectively), followed by film & video. On the other hand, journalism had the lowest number of successful campaign (4 counts) but it was worth noting that it had zero number of failed or canceled campaign.

The most favourite category for live campaign was film & video (5 counts) and the least were food, journalism, and music. Games (mobile games) is the only category that had a higher chance of having a failed campaign compared to its successful ones (23 and 21 counts respectively).

A screenshot of a computer

Description automatically generated

This resonated with the subcategory of theater, play, that had the highest number of successful, failed, and canceled campaigns. Science fiction under film & video category also showed a higher chance of failing the campaign (almost 50%) compared to the successful ones. Kickstarter who started their food and technology projects in CA had 100% rate of failure and 100% of journalism projects was successful only in the USA.

A screenshot of a graph

Description automatically generated

Campaigns done in Midyear (June, July) showed the highest chance of being successful. On the other hand, most failed campaigns happened in January, May, and August and the least in September. Lastly, April showed the least number of campaigns being cancelled.

In conclusion, Kickstarter will have a higher chance of having a successful campaign if the projects are within Theater category and carried out in the month of June or July. Moreover, those who started their campaign projects in Games (mobile games) and Science Fiction need be aware that there’s a higher chance of failure rather than success based on the data. Lastly, the target country where a campaign would be carried out should be carefully considered.

* + What are some limitations of this dataset?

There’s only a small sample being collected for certain categories, for example, only 4 journalism samples could be found and they’re all located in the US. Moreover, there’s lack of variables, for example, gender and the popularity of the Kickstarter could be included in the dataset as they could affect the result of the outcome.

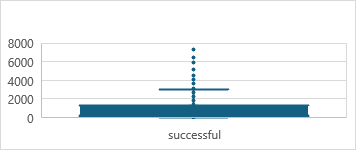
* + What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

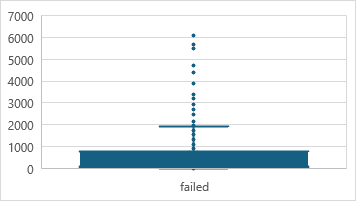
Tables/chart for the duration of the campaign to show comparison between the duration given for the successful, failed and cancelled projects to run.

Graphs to compare the number of backers and the outcome of campaign will provide a clear visualisation on how the number of backers can boost the success of a campaign.

STATISTICAL ANALYSIS

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





The variance values for the successful and failed campaigns were high ((1,600k and 921k respectively) which means that the dataset in the successful and failed campaigns varied far from the mean value. This could be affected by the number of outliers that both campaigns contained and shown by the range of values; successful campaign varied between 16 to 7k and failed campaign varied between 0 to 6k.

Hence, using median of both successful (201) and failed (115) campaigns is better than the mean (851 for successful campaign and 586 for the failed campaign) to summarize the data because mean is sensitive to extreme values and could skew the interpretation of our data. Median on the other hand, is not affected by the outliers and help us to identify a more typical success rates.

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

The value of the variance of the successful campaign almost doubled the value of the failed campaign (1,600k and 921k respectively) because of some reasons. Firstly, the backers were attracted to certain quality and appeal of the projects and hence, there was a significant difference in the success rate of certain projects. For example, projects in film & video (99 counts) were highly preferred compared to journalism (4 counts). In some countries, we could see significant project preferences which also contributed to the high variance value. For example, plays were generally favoured by majority of backers in European countries, Australia, and Canada, and the US. However, when it comes to projects in journalism, only backers in the US who supported it.