**Report on Crowdfunding campaigns.**

**Part 1 Analysis**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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** |

1)The "Grand Total" row and column give the overall count of each outcome across all categories and the overall count of each category across all outcomes. This table provides a clear overview of the distribution of project outcomes across different categories in the dataset.

1. The success rate relative to number of projects.

The below table gives an overview of the success, cancelled, live and failure rates of crowdfunding campaigns across different categories. The data shows that, the "Theatre" category has the highest number of projects with successful campaigns (187) and failed campaigns (132). Which is indicate that high in popularity but around 50% of chance of success. The Journalism had lowest number with 4 projects with all success, translated to 100% success rate. The technology, film, photography, publishing category has a relatively high success rate with technology having highest success rate.

1. The success and failed rate highest among the “play” subcategories

Based on the table, it is evident that the "plays" category has a significant contribution with a total of 344 projects, out of which 187 projects were successful, representing an impressive success rate of nearly 55%. However, there were also 132 projects that did not meet their funding goals.

Among the subcategories with 50 to 60 projects, the "web" category stands out as the most successful, achieving a remarkable success rate of over 70% compared to other projects in this range.

1. The monthly trend does not show any significant indicators.

After analysing the monthly data based on past years, no significant patterns or trends are evident. The number of successful, failed, and cancelled projects shows fluctuations across the months, but there is no consistent or remarkable pattern that stands out. The distribution of projects remains relatively consistent throughout the year, with no specific month showing a significant deviation from others in terms of project outcomes. However, during the month of September to November there are slight fall in funding activities but very minimal. This suggests that the monthly performance of projects does not exhibit any notable seasonality or consistent changes over the long-term period.

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

Based on the observation of the diverse dataset, it appears that some parent categories have a significantly lower number of projects compared to others. To improve the analysis and reduce the variance in the data, it is suggested to categorize the projects into lower, middle, and higher range of project categories.

For example, categories like journalism, which only had 4 projects out of 1000, could be classified as "lower range" due to their limited representation. On the other hand, categories such as theatre, film, and music, which seem to have a higher number of projects, can be grouped into the "higher range" category for better analysis.

By adopting this approach, the data analysis will be more balanced, and it will allow for a more accurate assessment of trends and patterns within specific project categories. This categorization will help focus the analysis on the categories with a significant number of projects, ensuring that the findings are more representative and meaningful.

Using the country of origin to find comparable data on crowdfunding success rates can indeed provide valuable insights into the countries where people are willing to invest their money in various projects. Since the data includes only a few countries (US, Canada, and Europe), we can analyse the success rates for crowdfunding projects from these regions.

To perform this analysis, we would need to:

* Filter the data to include only projects from the US, Canada, and Europe.
* Calculate the success rate for each country by dividing the number of successful projects by the total number of projects from that country.
* Compare the success rates across these countries to identify which ones have a high crowdfunding success rate and where people are more willing to invest.

By focusing on these specific regions, the analysis will be more manageable, and it will allow us to draw meaningful conclusions about the crowdfunding success rates in these countries. This information can be valuable for understanding the crowdfunding landscape and identifying potential areas of interest for investors and project creators.

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

Including the "Spotlight" feature and analysing its correlation with the number of projects and their success rate can provide valuable insights.

The "Spotlight" feature in crowdfunding platforms typically highlights certain projects or gives them special visibility to attract more attention and backers. By examining the correlation between the "Spotlight" status, the number of projects, and their success rates, we can determine if there is a meaningful relationship.

If there is a positive correlation between being in the "Spotlight" and higher success rates, it could indicate that the special visibility provided by the "Spotlight" feature attracts more backers and leads to greater crowdfunding success.

By incorporating the "Spotlight" status into the analysis, we can uncover additional insights and potentially identify factors that contribute to higher crowdfunding success rates. This information can be valuable for both project creators and backers in understanding how project visibility and promotion impact their chances of success in the crowdfunding platform.

Part 2 Analysis.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Campaigns | | |
| Backer | Successful |  | Unsuccessful |
|  |  |  |  |
| Mean | 851 |  | 586 |
| Median | 201 |  | 115 |
| Minimum | 16 |  | 0 |
| Maximum | 7295 |  | 6080 |
|  |  |  |  |
| Variance | 1606217 |  | 924113 |
| Std Deviation | 1267 |  | 961 |

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

The Mean represents the average number of backers for each group. For Successful campaigns, the average number of backers is approximately 851, while for Unsuccessful campaigns, the average number of backers is approximately 586. This suggests that, on average, successful campaigns tend to attract more backers compared to unsuccessful ones.

The Median represents the middle value when the data is arranged in ascending order. For Successful campaigns, the median number of backers is 201, while for Unsuccessful campaigns, the median number of backers is 115. The median provides a measure of central tendency that is less influenced by extreme values. In this case, the median for successful campaigns is significantly higher than for unsuccessful campaigns, indicating a notable difference in the distribution of backers.

The data analysis provides insights into the distribution and spread of the number of backers for successful and unsuccessful projects. The mean and median provide central tendency measures, while the variance and standard deviation give an idea of the data's variability. This information is valuable for understanding the typical and extreme values of backers in each group and how much the data points deviate from the mean.

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

The variance of successful campaigns is higher than unsuccessful campaigns. Consequently, the number of backers in successful campaigns exhibits greater variability than that of unsuccessful campaigns. This higher variance suggests that the data points in the successful campaign group are more dispersed from the mean value, indicating a broader range of backer counts for successful campaigns.

The standard deviation for successful campaigns is 1,267, while for unsuccessful campaigns, it is 961. Like variance, the standard deviation confirms that the backer counts for successful campaigns have more variability compared to unsuccessful campaigns.

Overall, based on the variance and standard deviation measures, it becomes evident that the number of backers in successful campaigns displays a higher level of variability compared to unsuccessful ones. This observation is logical as successful campaigns tend to draw a more diverse and varied pool of backers, resulting in a broader distribution of backer counts. In contrast, unsuccessful campaigns typically have a more concentrated distribution of backers, leading to less variability in the dataset.

The data analysis highlights a crucial point successful campaigns exhibit a greater degree of variability in backer counts. For crowdfunding platforms and project creators, this information is vital to comprehend the potential range of support they may receive for their campaigns. Understanding the variability in backer engagement can aid in making informed decisions and tailoring strategies to achieve better results in crowdfunding endeavours.