Andrea Aurrecoechea Diaz.

**Week 1 – Challenge 1**

**Background**

Crowdfunding platforms like Kickstarter and Indiegogo have been growing in success and popularity since the late 2000s. From independent content creators to famous celebrities, more and more people are using crowdfunding to launch new products and generate buzz, but not every project has found success.

To receive funding, the project must meet or exceed an initial goal, so many organisations dedicate considerable resources looking through old projects in an attempt to discover “the trick” to finding success. For this week's Challenge, you will organise and analyse a database of 1,000 sample projects to uncover any hidden trends.

**Module 1 - Challenge**

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

**Conclusion 1**

Theatre-related campaigns (i.e plays) are overrepresented in the data set, accounting for 34.4% of all campaigns. This suggests that the theatre industry actively seeks crowdfunding platforms for funding, potentially indicating high competition within the sector.

It is worth noting that their success rate in relation to the total number of campaigns is 54.68%, which is somewhat lower than the average of this data set at 57.30%. This indicates that while theatre campaigns are prevalent, they may face slightly lower success rates compared to other campaign categories in the dataset.

It is important to note that the success rates vary significantly across different project categories. Some categories such as journalism and world music showed 100% success rates, however they both had very limited number of projects and further analysis may be required to determine any significant trends.

On the other hand, the Technology category, particularly projects related to the Web, demonstrated a total of 94 projects with a success rate of 68.09%. Specifically, the subcategory of Web projects had an even higher success rate of 72%. These numbers indicate a strong market demand and significant interest in this field, making them promising areas of focus for innovative projects.

**Conclusion 2**

Fundraising campaigns with lower funding goals, specifically between $1,000 and $4,999, had the highest success rate at 83%. This suggests that setting a more attainable fundraising target may lead to a higher likelihood of success. Projects with goals of less than $1,000 also achieved a relatively good success rate of 59%. This indicates that even smaller fundraising campaigns can be successful if well-executed and targeted to the right audience.

Projects with funding goals of $50,000 or more faced a lower success rate of 37%, suggesting that raising larger amounts can be challenging and may require more strategic planning and effort.

While campaigns looking to raise between $15,000 and $50,000 had an impressive average success rate of 85%, it's crucial to note that this conclusion is based on a relatively small sample of 75 projects, representing only 7.6% of the total data. Therefore, further analysis and consideration are needed when interpreting results from this small sample size.

Other areas of interest are:

* Across all parent categories, most successful campaigns exceeded their initial fundraising goals. On average, they raised approximately 1.97 times the amount of their original target. This trend should be strategically considered when setting campaign goals to maximize fundraising potential.

**Conclusion 3**

The data reveals a substantial increase in the number of successful campaigns during the middle months of the year. In June, the number of successful campaigns is 1.6 standard deviation points above the mean, and in July, it is 2.2 points above the mean. This observation suggests a potential seasonal pattern or influential factors contributing to the higher number of successful campaigns during these months. Understanding and leveraging these trends can provide valuable insights for optimizing future campaign strategies and resource allocation to maximize overall success.

It is worth noting that when the monthly success rate is calculated in relation to the number of campaigns per month, the following months have the highest success rates:

* June (63.95%)
* July (62.37%)
* September (61.64%)

These rates are all somewhat higher than the overall success rate of this data set at 57.30% (not including live campaigns).

Understanding the success rate in relation to the number of campaigns provides valuable context, showcasing the efficiency and effectiveness of marketing efforts during these peak months. Leveraging this information can guide future campaign strategies and resource allocation to capitalize on the successful patterns observed in June, July, and September, potentially leading to further improvements in overall campaign performance.

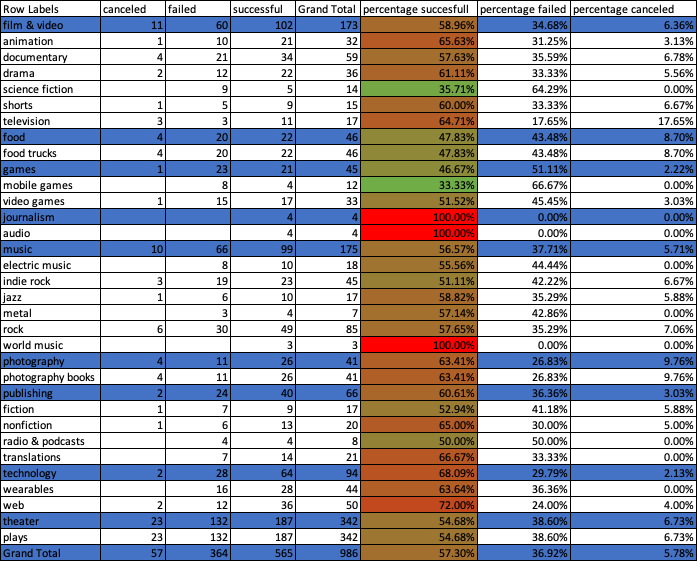
1. **What are some limitations of this dataset?**

* The high representation of Theatre/plays (34.4%) can introduce bias and skew the overall analysis. Conclusions drawn from this data may not be applicable or representative of the broader project landscape, as the dataset is heavily weighted towards this category.
* There is a wide range of goals, from $100 to $199,200. This results in an uneven distribution of projects across different goal amounts. With such broad goal range, there is a higher chance of extreme outliers that can impact the overall analysis. This also brings challenges when trying to group this data.
* The available data is limited to observations until 2020, with only 2 entries for this year. This lack of recent data may not accurately reflect the current crowdfunding landscape and behaviour, particularly in the context of events like COVID, which could have impacted crowdfunding trends and outcomes in more recent times.
* It is important to mention that the dataset explicitly states different currencies for goals and pledged amounts. However, for the purpose of this exercise, I have disregarded this distinction and treated all amounts in the "goal" and "pledged" columns as if they were in the same currency, specifically AUD (Australian Dollars). It's essential to recognize that if we were to consider the currency column, the analysis pertaining to or depending on these two variables would yield significantly different results.

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

**Additional Tables:**

Success and Failure Rates per Parent and Sub Category



**Other useful graphs:**

* + Bar chart showing the success Vs the failure rate by category and/or subcategory:

A graph of a graph of a graph

Description automatically generated

* + Box whiskers charts or histograms showing the skewed distribution of number of backers as well as campaign’s goals, for example:

A graph with blue lines and dots

Description automatically generated

**Bonus Statistical Analysis**

Use Excel to evaluate the following values for successful campaigns, and then do the same for unsuccessful campaigns:

|  |  |  |  |
| --- | --- | --- | --- |
| Successful Backers | | Unsuccessful Backers | |
| Mean | 851.146903 | Mean | 585.615385 |
| Median | 201 | Median | 114.5 |
| Minimum | 16 | Minimum | 0 |
| Maximum | 7,295 | Maximum | 6,080 |
| Variance | 1606216.59 | Variance | 924113.455 |
| Standard Deviation | 1267.36601 | Standard Deviation | 961.3082 |

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

Based on the provided data, there is a notable difference between the mean and median values for both successful and unsuccessful backers. This discrepancy suggests that the data may be skewed, and when visualized in a box whisker chart, it reveals the presence of several extreme or outlier values, further contributing to the skewness of the data.

Given the skewness and the presence of outliers, the median is likely to be a better summary measure of the data compared to the mean as the median is less influenced by extreme values.

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

Based on the data provided, the variability appears to be higher in successful campaigns compared to unsuccessful campaigns.

This suggests that the number of backers in successful campaigns tends to vary more widely from the average compared to unsuccessful campaigns.

Successful campaigns often attract a broad audience, leading to a wider range of backer participation. Some campaigns generate significant engagement, with backers potentially sharing the project's potential with friends/family, creating hype and contributing to high variability. The variability in campaign goals may also influence the number of backers for successful campaigns.

Conversely, unsuccessful campaigns tend to have more consistent or limited interest, resulting in less variability. However, it's worth noting that even unsuccessful campaigns exhibit significant variability, likely due to diverse factors influencing backer engagement, project types, and funding goals.