A graph with different colored bars

Description automatically generatedA graph with orange and blue bars

Description automatically generatedA graph of different colored lines

Description automatically generated

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

Firstly, the stacked column pivot chart shows that Theatre, Music and Film&Video attract more backers around the world and have better chances of success compared to other categories.

Secondly, the highest number can be found in Plays which is a mainly subtype of Theatre. This indicates that plays can be the most popular campaigns among the crowdfunding projects. In addition, the subcategories of crowdfunding campaigns show diversity and competitive challenge as the rate of “successfully achieving a funding goal” is similar to the rate of “unsuccessfully achieving a funding goal”.

Finally, the line tendency demonstrates the number of successful campaign peaks around mid-year while the number of failed campaigns remaining relatively stable throughout the year.

1. What are some limitations of this dataset?

Firstly, some subcategories could overlap that makes it difficult to find out precise trends of success.

Secondly, the data could reflect seasonal biases, as some months show higher activity than others. It should cover more granular time data, such as weekly trends.

Additionally, there is a lack of contextual information on why certain campaigns succeed or fail. The data analyst could capture the factors like target audience engagement, campaign quality and marketing tendencies. The criteria of classification makes it difficult to understand data comprehensively.

Moreover, the financial information, such as the funding goals and the amount of money pledge, should be considered as it provides an assessment of scale of the campaign.

Finally, despite of simple numbers or amounts, we could think about the level of backer engagement, duration of the campaign, following-ups feedback that could influence the success or failure of a campaign.

1. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
2. A scatter plot of the funding goals of campaigns and their outcomes, which are categorized as successful, failed, canceled and live. This relationship could whether a specific funding goal range tends to become successful.
3. A map or bar chart of the number of campaigns by geographical region or country could highlight regional trends and show where crowdfunding is most popular or successful.
4. A list of top 3 most popular funded campaigns in each parent category which can offer insights of what made the campaign successful.
5. A table showing the relationship between the average pledge amount for each subcategory can offer insights of what kind of category is more likely to attract higher individual pledge. This can help the funder to decide the threshold for pledge amounts.
6. A bar chart showing the duration of campaign related to their successful rate can provide insights whether long-term or short-term campaign is more likely to be successful.

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

The Mean is highly sensitive to extreme values, which is the outliers, which can make it unrepresentative of the dataset, while the Median is not affected by the outliers. The scatter plots demonstrate most of the data clustered at lower backer counts. The Median is also better to show central tendency distribution.

Therefore, the Median is a more appropriate metric for summarizing the skewed distribution. It will provide a more accurate representation of the central tendency of the backer for both successful and unsuccessful campaigns.

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

I calculated the variance and standard deviation for successful and unsuccessful campaigns, in order to get more precise analysis. This is because the these measures can indicate how much variation exists from the average value. Higher values can show more spread out data, while low values show data is close to the mean.

Therefore, there is more variability in the number of backers for successful campaigns than that for unsuccessful campaigns. This makes sense because successful campaigns can keep attracting a lot of backers’ attention. Additionally, successful campaigns could have stretch goals to encourage contributions, which leads to further variability.