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

* Based on the given data, out of the 1000 crowdfunding campaigns in the provided data, 565 of them are categorized as "successful." This indicates that a significant proportion of campaigns can meet their funding goals and achieve success. Theater campaigns have a significantly higher number of successful outcomes compared to other categories. Theater campaigns have 187 successful campaigns out of 344. This suggests that this category may have a strong and supportive fanbase willing to fund projects.
* The months of June and July show a peak in the number of successful campaigns. This could be attributed to people having more free time during the summer, being in a more generous or giving mood, or being more active on crowdfunding platforms during vacation periods.
* The mean (average) number of successful backers is approximately 851.15, while the mean number of unsuccessful (failed) backers is approximately 587.51. This indicates that, on average, successful campaigns tend to attract more backers compared to failed campaigns.

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

* If the data comes from multiple crowdfunding platforms, each platform may have its own rules, audience, and campaign dynamics. Treating them as a single group may oversimplify the analysis.
* The absence of detailed information on the specific factors contributing to campaign success or failure in the dataset limits the ability to perform a comprehensive analysis of the underlying causes and hinders the development of actionable insights for crowdfunding decision-makers.

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

* Country-wise Summary Table

Provide a summary of campaign outcomes by country. List countries, the number of campaigns originating from each country, and the success rate for campaigns from each country. A "Country-wise Summary Table" would be a useful addition value to the analysis. It can provide insights into how campaigns from different countries perform. This table can help to identify countries with a high success rate, which might be interesting for further marketing or expansion strategies. It also provides an overview of the geographical distribution of campaigns in the dataset. A pie chart can be used to graph the proportion of pledged amount from each country.

* Top 10 Categories by Average Pledge Amount:

This table would list the top 10 parent categories (e.g., "technology," "music") and their corresponding average pledge amounts. The "Top 10 Categories by Average Pledge Amount" table is a valuable addition to the analysis, as it helps identify which categories tend to receive higher average pledges. This table allows us to see which categories tend to attract backers who are willing to make larger average pledges. It can help creators and marketers in those categories understand the potential for higher funding. A bar chart can be used to graph the top 10 categories based on the average pledged amount.

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

* The mean number of successful backers is significantly higher (approximately 851.15).

The median number of successful backers is lower (201).

The skewness value for successful backers is 2.170415479 which is significantly higher. In this case, the data for the number of successful backers is likely right-skewed, with some campaigns having a very high number of backers, which is pulling the mean higher. In such right-skewed distributions, the median is often a better summary statistic because it is less affected by extreme outliers. Therefore, the median (201) better summarizes the data for successful backers.

* The mean number of unsuccessful backers is higher (approximately 587.51).

The median number of unsuccessful backers is lower (114.5).

The skewness value for unsuccessful backers is 2.684034624 which is significantly higher. Similar to the successful backers, the data for the number of unsuccessful backers is likely right-skewed. Again, in such cases, the median (114.5) is a better summary statistic because it provides a measure of central tendency that is not heavily influenced by extreme values.

In summary, for both successful and unsuccessful backers, the median is a better choice for summarizing the data because it is more robust to outliers in right-skewed distributions.

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

* For successful campaigns, the variance is higher (1,603,373.732) compared to unsuccessful campaigns (925,919.3218).
* For successful campaigns, the standard deviation is also higher (1,267.366006) compared to unsuccessful campaigns (963.5788474).

This suggests that the number of backers for successful campaigns tends to be more spread out from the mean (average) compared to unsuccessful campaigns. This makes sense in the context of crowdfunding. Successful campaigns may vary widely in terms of popularity and the number of backers they attract. Some highly successful campaigns may have many backers, leading to higher variability, while less successful campaigns may have fewer backers, resulting in lower variability.