***Conclusions, limitations, and future direction***

Given the data, there are three conclusions that we can draw about crowdfunding:

* The theatre category had the highest number of successful campaigns relative to the other parent categories, but they also had the highest number of failed campaigns as they had the highest number of campaigns overall.
* The plays sub-category had the highest number of successful campaigns relative to the other sub-categories, but similarly to the aforementioned conclusion, this was because they had the greatest number of campaigns by a large amount when compared to the sub-categories as they also had the largest number of failed campaigns.
* The number of successful campaigns seemed to be highest in July for the average of all of the parent categories when compared to the performance in the other months. The highest number of failed campaigns happened at the start of the year, in January.

There are, however, a few limitations to this dataset. Without additional analysis and calculations, it is not possible to tell which category had the highest number of successful campaigns proportionally to their failed campaigns. Even though the theatre parent category had the highest number of successful campaigns, they also had the highest number of failed campaigns, so it is not possible to draw the conclusion that their category performs the best overall. Through further analysis, it was determined that proportionally, the technology category had the highest number of successful campaigns relative to their total number of campaigns. However, because their total number of campaigns was so small, their significance could have been missed given the provided data. Further analysis and pivot tables/charts comparing successful and failed campaigns of the various categories and sub-categories could have helped determine which category and sub-category was most successful relative to their failures. This could be done by dividing the number of successful campaigns by the total count of campaigns and multiplying by 100 to find the percentage of success for each category and sub-category, and then plotting this in a line graph.

It also would have been beneficial to compare how successful a category or sub-category was relative to how long the campaign was running for as some campaigns ran for longer than others, which is why they could have potentially gained more revenue. Plotting a graph of the total days of the campaign run against the number of successful campaigns could have furthered our analysis of which campaigns and categories performed better.

Furthermore, a last graph showing how the various categories performed through the years would be beneficial (in addition to the already existing graph showing performance over the months of any given year). Right now, there is no indication of long-term growth over the years so this would give a good indication of the long-term performance of each category as if there is a positive trend, that would show positive growth for any given category.

***Statistical analysis***

In this case, the median might be a better measure of central tendency compared to the mean as there are large differences between the minimum and maximum values of both data sets, which could pull the mean one way or another. The fact that the mean is so much higher than the median suggests that the data must be skewed and hence, the median is better in this case as it is not skewed by especially high or low values and is not as greatly affected by outliers.

The successful campaigns had a higher variance of the number of backers compared to the failed campaigns. This makes sense as it seems that there were comparatively not as many backers for the failed campaigns as it would make sense for a campaign to succeed with more backers. Therefore, although there was a low variance for the failed campaigns, there were not many backers in general comparatively. Furthermore, there seems to be a higher spread of values when comparing the minimum and maximum values for the successful category, and a large difference between the median and the mean, meaning that a high variance makes sense.