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

Overall, there are more successful outcomes in each category and sub-category.

1. Based on pivot table of category, theater has the highest outcome, both failed and successful outcome. Film & video category is the second highest outcome with more successful than failed outcome. Music category is the third highest outcome with more successful than failed outcome.
2. Based on pivot table of sub-category, people are interested to attend plays sub-category the most in theatre. Rock and documentary are the other sub-categories people are interested in attending.
3. Based on the count of outcome from pivot table, summer months (Jun- Aug) are predicted to have the most successful outcome and July is the peak month. One can say people are more outdoors in summers and take interest in participating various events.

* What are some limitations of this dataset?

1. The crowdfunding campaigns data doesn’t include any information how many people or supporters were involved in raising the money and how were they raised and what were the sources for raising the money.
2. When pivot table of category and sub-category is performed, we can see that there are missing data in journalism, food, and music in category and several sub-categories have missing data. These missing data could provide valuable information which we could compare with other categories and conclude for a trend.

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

1. We could plot one graph between goal and pledged to find out how much money is remaining and needed to be raised. If there is more money raised, then is it enough to start a new campaign.
2. We could plot a graph for the number of campaign supporters involved in each category and sub-category every month and year to see the pattern of donation involved by supporters. This could answer questions like how many supporters are required to raise funds in a specific category.

**Crowdfunding goal analysis**

* Create a line chart that graphs the relationship between a goal amount and its chances of success, failure, or cancellation.

We can draw conclusions like:

Goal amount range from 15000 to 19999, 20000 to 24999 and 30000 to 34999 were 100% successful in achieving the fund. Higher the goal amount, more likely the chances of successfully achieving the goal.

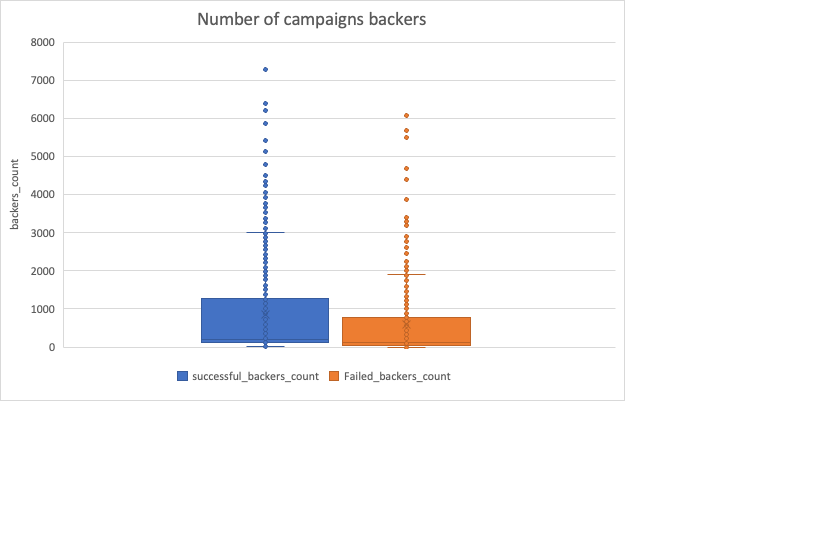
**Statistical analysis**

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

Median is better in summarizing the data as this data is not following normal gaussian distribution for both successful and failed backers count. Boxplot is distorted by the number of outliers in the upper quartile and hence its skewed to right. Mean is sensitive to the outliers and is impacted by them. Mean is pulled towards the outliers resulting in a distortion. Median considers the middle value of the ordered data set regardless of the number of outliers. Hence, Median is better in summarizing the data here. Median in successful count is higher than the failed count.

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

1. Based on the calculated summary statistics, both variance and standard deviation are high in successful campaigns than unsuccessful one.
2. There is more variability in successful campaigns than unsuccessful one because of the large number of outliers. Outliers contribute to the variability of the dataset and inflate the variance.
3. The min and max values of successful are higher than unsuccessful campaigns. Hence, based on summary statistics, it makes sense to see more variation in successful campaigns.



Mean affected by outliers

Median unaffected by outliers

Median unaffected by outliers

Mean affected by outliers