**Assignment #1**

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

* The Theater is the most supported crowdfunding campaign with plays being the most supported sub-category.
* Most campaigns are either successful or failed and only canceled less than 10% of the time.
* More Campaigns are successful in July than any other month of the year.

What are some limitations of this dataset?

* Not all categories are equally represented. For example: Journalism represents 4/1000 entries, where as Theater represents 344/1000 entries.
* The data is not all using the same currency of money.
* The countries are not represented equally with most of the data coming from the US.
* The years are not equally represented in the data set. For example: 2020 has 2 data points whereas the other years have between 81 and 107 data points.

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

* We could create tables/graphs that show the number of backers and the average donation for each category/sub-category. This would provide an added value of know how many people are offering support and how much on average each person is donating to each of the crowdfunding campaigns. This would help us to see whether a small number of people are donating a lot of money or if many people are donating a smaller amount to the campaigns.
* Another possible table/graph we could create would show the backers count by country. This would help to show which country has more people contributing to crowdfunding campaigns.
* We could also create a table/graph to show pledged/goal money based on category or sub-category. This would help to demonstrate which category or sub-category is getting pledged the most money regardless of outcome. For example: the ID 87 – Music category has a failed outcome but had 123,040 pledged whereas a lot of successful outcomes were for less pledged money. Perhaps the goal of each category plays a factor in the success rate.

Statistical Analysis

The Median would better summarize this set of data. This is because there are outliers within both the successful and failed backers\_ count outcomes. Therefore, the Mean is affected by the outliers and creates a bias to the outliers. To help better show this I have created boxplots for both the successful and failed outcomes by backers count which demonstrates that the medians are not in the centre of the interquartile ranges and the means are also pulled up from the median due to the outliers in the data set.

There is more variability in the successful campaigns. I believe that this makes sense because the standard deviation is larger for the successful outcomes than that the failed outcomes; this means that the data points in the successful outcomes are spread out more (further from the mean) than within the failed outcomes.