**Crowdfunding**

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

1. Based on the number of outcomes, the theater industry is the leading category in crowdfunding campaigns.
2. Plays, documentaries, and rock are the three sub-categories that lead the crowdfunding campaigns (based on total number of campaigns).
3. The summer months (especially July) are the times where a lot of crowdfunding campaigns are created.

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

One limitation is that the size of the dataset is not big enough. There are many categories that are yet to be listed in this crowdfunding data sheet. The current crowdfunding data sheet only lists 9 parent categories. Another limitation for this dataset could be that we do not up-to-date data sets. The current data set only has information in the past 10 years from 2010 to 2020. We should also include data from the years 2021 and 2022 to count for any economic factors that would affect the amount of raised funds and backers count.

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

Some other possible tables and/or graphs that we could create other than bar and line graphs are pie charts. By creating a pie chart, we can categorize each industry of crowdfunding and see the percentage of the amount funded. It would allow us to see which industry is the best leading in the crowdfunding industry based on percent funded rather than number of crowdfunding campaigns.

**Statistical Analysis Questions**

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

Both successful and failed campaigns have right skewed data. Data that is skewed to the right show mode as the lowest value, median as the next higher value, and mean with the largest value. The mean backers of successful and failed campaigns hold a greater value than the mode and median of the data set. For successful campaigns, the mean is 851 backers, whereas the median and mode are 201 and 85 respectively. For failed campaigns, the mean is 586 backers, whereas the median and mode are 115 and 1 respectively. For this reason, the median better summarizes the data because the mean is more susceptible to any extreme outliers than the median of the backers count.

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

Based on the data, it seems that having more backers leads to a higher probability of achieving the goal amount or exceeding the goal funds amount. As a result, after analyzing the data set, there is more variability (where variance for successful campaigns is 1603374 and variance for failed campaigns is 921575) with successful campaigns because successful campaigns tend to have more backers than failed campaigns.  To showcase this outcome, the maximum value for the successful campaigns was 7295 backers compared to 6080 backers for the failed campaigns. In addition, there are more data records for successful campaigns (565 data inputs) than there are failed campaigns (364 data inputs). When dealing with larger data sets, there tends to be more variance, which makes sense that successful campaigns would have a bigger variance value.