**Challenge 1 Questions**

1. Given the provided data, what are 3 conclusions that we can draw about crowdfunding campaigns?

Conclusions:

* Theater is the most popular project category on crowdfunding.
* Journalism is the least popular project category on crowdfunding.
* Although journalism is the least popular project, 100% of journalism projects were successful.

2. What are some limitations of this dataset?

This dataset limits the analysis towards the genres of projects that are popular/unpopular and successful/unsuccessful. Nevertheless, it does not give as much insight on how and why certain projects became successful. For example, how much time was put into each project or how much money each creator invested into their project?

There are no descriptions on the unit of measures for the goal and pledged variables. Therefore, when analyzing these columns, it is best to look at the rate of change in the project goals and the amount pledged rather than the face value of each goal/pledge.

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

Other possible tables:

* Staff picks and outcomes: used to gauge the significance of crowdfunding staff’s opinions towards the success of projects
  + Graph: Bar plot
* Project duration and outcomes: Can calculate the project duration by taking the difference in launch and deadline times. Could be used to see how long on average successful/unsuccessful projects last
  + Graph: Bar plot to show the average project duration for successful, failed, and canceled projects
* Project duration and percent funded: Can look into the correlation between these two values to see if there is a relationship between the length of time a project lasts vs. the percentage of funds
  + Graph: Scatterplot with trend line
* Location and outcomes: look to see where most projects are being developed and which locations have the highest amount successful/failed projects. Which locations have the highest/lowest rate of success?
  + Graph: Map graph

Bonus Statistical Analysis:

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

The median would be a better to summarize the data as the data is not normally distributed. The data is skewed towards the right and has the majority of the values within the left most bin. Thus, the mean would not be a good indicator for the average of the data as there are large values that skews the mean to be larger than the true average.

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

There is more variability with successful campaigns compared to unsuccessful. This makes sense because successful campaigns could have a large range of backers. Backers could be contributing $0.01 to any large amount they want. Therefore, some campaigns can be labeled as successful with less backers if the backers supporting them are contributing more money. While other successful campaigns may have a lot of backers who are contributing a smaller amount of money. With unsuccessful campaigns, there is less variability because the amount of backers would generally be a less with very low contributions.