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

Given the data, some conclusions are

* 1. After creating the pivot table from the crowdfunding sheet, in the pivot table field I put the parent category as the row, and the outcome in the column to get the relationship between each category and the number of outcomes based on canceled, live, failed, and successful. Based on the category statistic, the theater has the largest portion of the grand total with 33.4% and journalism only takes the smallest part of the grand total with 0.4%. We can see the theater category consists the highest success with a rate of 33% compared to other categories also it has the highest percentage of failure with 36%.
  2. After creating the pivot table from the crowdfunding sheet, in the pivot table field I put the subcategory as the row, and the outcome in the column to get the relationship between each category and the number of outcomes based on canceled, live, failed, and successful. Based on the subcategory statistic, Play has the most number from the grand total. Based on the subcategory statistic, we can see that the Plays subcategory consists of the highest success rate of 34%.
  3. Based on the launch date statistic, we can see that the highest success rate is during July.
* What are some limitations of this dataset?

Some limitations of this dataset are the recency and the accuracy of the data. The dataset does not contain the latest year or it is not up to date. We also don’t know how experienced the data creator is and how trustworthy their data are.

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

Other tables or graphs that we could create be the successful rate vs the campion goal. This way we can see the correlation between them. Another graph could be the duration time of the campion vs the campion goal. This way we can see how long it takes to reach the campion goal. Additional values we can provide are sorting the successful rate by country.

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

I would use the median to better summarize the data because this data consist of some outliers and it will increase the mean more than it would because of the outlier. The median is where most “normal” are.

* 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 in successful campaigns because the variance and the standard deviation is higher in this case. It makes sense because the variance is a measure of the difference between the sample and the actual. A large variance indicates that there is a large variability