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

Given the provided data, it shows that most crowdfunding campaigns fall into the theater category (344 out of 1000), specifically plays which make up 342 of the 344 campaigns within theater.

You can also conclude that about 50% of crowdfunding campaigns are successful, given the metric used for this analysis.

You can also conclude that only a small percentage of crowdfunding campaigns end up being canceled (5.7%).

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

Some limitations include most data being from the US (763 of the 1000 campaigns). This is a limitation because you can’t really generalize your finding across the world unless you expand your data set to include a higher percentage of campaigns from other countries.

Another limitation is the wide spread of campaigns within each category. The largest category (theater) has 344 campaigns while the smallest category (journalism) only has 4 campaigns. This makes it harder to generalize the success/failure rate for campaigns in categories that are smaller.

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

Another possible graph that can be used is a pie chart, which would show the categories that have the most/least campaigns in the data set. You could also create a pie chart filtered by failed/successful to see the spread within each category.

Including the percent failed, percent successful and percent canceled in this table would also give you a better idea of the probability of the campaign’s outcome given their category/sub-category.

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

For both data sets, successful and failed, the median would better summarize the data because the data is skewed so the mean gets pulled towards the larger values. The median for successful is 201 which means half the data points are between 16 to 201 and half the data set is between 201 and 7295. The mean is much larger, 851.1469, which shows that you need on average 851 backers to have a successful campaign but that is not a good representation of the data because it is skewed. The median for failed is 114.5, meaning half the data set is found between 0 to 114.5 and the other half is between 114.5 and 6080. The mean for failed is also much larger than the median, 585.61538, so it is not a good representation of the data because it is skewed.

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

Both sets of data have a high variability because the data is skewed but successful campaigns have the largest variability with a value of 1606216.6. Variance is a much larger number so looking at the standard deviation gives a better representation of the spread of the data in terms of how spread out each value is in relation to the mean of the data set.