1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns? Explain the reasoning behind your answers.

Based on the parent category data and looking at the fraction of successes, music is the most successful category, while food seems to have the lowest chance of success. Journalism has yet to see any success but currently has a small sample size.

Looking at the sub-categories, plays are by far the most popular campaigns and have the highest number of successes, but do not have the highest chance of success. Interestingly, many of the other sub-categories appear to be quite binary in terms of chance of success. For example, rock music has an impressive 260/260 success rate, while all 100 video game campaigns have failed. Cleary there is/are some strong hidden feature(s) at play here.

Third, a more thorough statistical analysis is needed to possibly conclude any trends in the launch date data. However, it is worth noting that December is the least popular month to launch, and the only month that sees more failures than successes. We can speculate about the holiday season playing some negative role here, but that is just speculation.

2. What are some limitations of this dataset?

To make better sense of any trends in the sub-categories, more data is needed. As mentioned, many of the sub-categories appear to either mostly fail or mostly succeed. Are the donors on Kickstarter part of a close-knit community that loves rock music and has no interest in video games? It would probably be interesting to see data on the donors themselves, how many support multiple campaigns, how much they donate, and to which categories. Also, perhaps the staff of Kickstarter are playing a role here and somehow causing some bias.

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

One other possibly useful table is one that shows the goal amount per category. Since goal amount is a likely factor in determining success, it would be interesting to note any disparities in goals between the categories and see how that compares with the outcomes. Perhaps lower outcomes of certain categories are due to higher average goal amounts, but perhaps their success may be higher than expected when compared to the overall data for success based on goal amount. This will provide insight into how much donors may value certain projects.

Another possible table may be to look at how long a campaign was active and compare it to its number of backers and outcome. I noticed the length of time between launch and deadline can vary and is therefore a possibly relevant feature when determining the chance of success.

Lastly, I would also consider lumping the campaigns that were canceled with the failed campaigns. This would reduce the number of data splits while likely not losing much insight since both statuses can be viewed as effectively the same.

Bonus Questions:

4. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

Since the means are much larger than the medians, the data is heavily skewed towards the larger numbers. With such a high skew, the median is a more meaningful number when describing central tendency.

5. 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 a much larger variability with the successful campaigns. This makes sense to me based off a quick glance at the raw data and some intuition. The number of backers for unsuccessful campaigns is probably clustered around 0, and has a lower maximum compared to successful campaigns. For successful campaigns, there are some that were backed by very few, but generous, donors (as little as 1 in some cases) while some were immensely popular and received thousands of petty donors – generally much more than the unsuccessful campaigns. Considering the two extremes and everything in between, there are many more permutations of donor types and therefore greater variability in the number of backers of the successful campaigns.