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

There is a count of 2,185 of successes in all 4000+ projects, and among all the categories, the category “theater” has the most count of states (1393) and number of successes (839), which the sub-category “plays” has the most successes (694). The second category with the most count of states and number of success is music. The third category with the most count of states is technology, but the third most successes is film & video.

Among all the sub-categories, anyone who is interested in investing would have to pay close attention to animation, children’s books, drama, fiction, food trucks gadgets, jazz, mobile games, nature, people, places, restaurants, translations, science fiction, video games, web, world music. These categories have no success but only failure and cancellation. On the other hand, the following sub-categories all have an 100% successful rate: classical music, documentary, electronic music, hardware, metal, nonfiction, pop, radio & podcasts, rock, shorts, tabletop games, television.

For each month, there are approximately same total count of states. For the trend of successes, it reached its peak at May, but then starting to decrease to its minimum value. The count of canceled project is very consistent, and the count of failed project is oscillating between 100 and 150.

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

It is unclear how these data are selected and there might be a bias when in selection. In other word, we cannot know whether these projects are a good sample to estimate the whole population of the Kickstarter projects.

It would be better if the reason why the project decided to cancel their project. There are some canceled projects that have well-exceeded the goal or almost achieving the goal, which is a little bit strange.

**What are some other possible tables and/or graphs that we could create?**

It would be helpful if we can have a separate table created to figure out the time duration each project has reached their current state, whether it is successfully reached their goal, or decided to cancel, or failed to reach their goal. This way we would have another indication of popularity of the specific category.

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

For the successful project, it would be more meaningful to use median. From the statistics, we see that the mean is much higher than the median, which means that the sample is right skewed. In other words, we have most of the data smaller than the mean. If using the mean, it would give the audience a misinterpretation of the typical backers count.

For the unsuccessful project, it would not really matter as much, but it would also be better to use median to summarize the data for the same reasoning as above.

**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 projects than unsuccessful. This make sense because we can see that the maximum of successful projects is much higher compared to the unsuccessful projects. There are certain projects that has a very high bankers count, which will increase the means by a great amount, which will eventually increase the variances.