**Kickstart Campaigns Report**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
2. Out of the 9 categories, “Theater” has the largest number of total projects (*n* = 1393). It also has the largest number of successful projects (*n* = 839). The category that has the highest successful rate is “Music” (77.14%), followed by “Theater” (60.23%) and “Film & Video” (57.69%). Thereby, current data suggests that projects in these categories are more likely to be successful comparing with other categories. However, “Theater” category also has the largest number of failed projects (*n* = 493). It also should be noted that “Technology” has the largest number and rate of canceled project (178 out of 600). It is hard to find out a reason without more details about the projects, so hope there will be additional information provided for a more in-depth analysis.
3. For the sub-categories, there are several of them have 100% successful rate (Rock, Documentary, Hardware, Tabletop games, Television, etc.) Although we cannot say these sub-categories are guaranteed to be successful without further information, they are still more likely to be succeed comparing with other sub-categories. “Plays” under the parent category “Theater” has the largest number of both total number of projects (*n* = 1066), as well as number of successful projects (*n* = 694). It is obvious that the current market has an interest on projects fall under this category.
4. When we look at the date each project were launched, the projects that were launched in May are more likely ended up being successful (*n =* 234) comparing with those were launched in December (*n =* 111). With that being said, date that a project was launched could be a factor that impact whether or not it get a positive outcome.
5. What are some limitations of this dataset?

The purpose of this analysis is to discover some trick for finding success in the Kickstart campaigns. Although the category, sub-category, and the launched data have some influence on whether or not each project will have a positive outcome, it is still hard to draw the conclusion that these variables have a causal relationship with being successful in the Kickstart campaigns. Even within the most successful category or sub-category, there are still many failed or canceled cases. It would be better if we could have more information, such as how the successful project are presented, or why the backers supported certain project. Thereby, we could have a better understanding of how to become successful in this Kickstart campaigns.

In addition, because these projects come from different countries across the world, the currency is different across each project. So we cannot directly compare the goal and pledged on different currency scales.

1. What are some other possible tables and/or graphs that we could create?
2. We could do statistical analysis on successful rate of each category/sub-category, and compare the percent funded within and/or across categories/sub-categories.
3. We could also calculate timeline (the time slot between data created and data ended) of each project, and compare the timeline between successful verse failed projects. We could see if there is a relationship between how much time were spent on a project and whether or not it would succeed.

**Backers count Statistical Analysis**

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

Given the current dataset, I believe the median summarizes the data more meaningful comparing with the mean. Because for both successful and failed groups, the variance and standard deviation are both vary large, indicating the present of outliners in the dataset. The mean will be affected by the outliners, but the median will not. So the median will be a better parameter to use to describe the central tendency in this case.

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

To compare the variability between successful and unsuccessful campaigns, I calculated the coefficient of variation (standard deviation/mean) for each group, which is a standardized measure of dispersion. The coefficient of variation for successful campaigns is 4.34, and 3.47 for unsuccessful campaigns. Thereby, there is more variability with successful campaigns. I believe it make sense because successful campaigns could have as many as possible backers to support in order to make them successful; but the unsuccessful campaigns usually have limited backers to support them, which make them didn't reach the goal.