**Summary Analysis of Kickstarter Dataset**

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

Conclusions:

1. The successful projects have more pledged funding than their goals; and the successful projects make up about 53% of the 4,114 projects. However, about 37% of the projects do fail and the pledged funding is less than the goals. Additionally, about another 8.48% of the campaigns are canceled; and of the 349 canceled projects there were 5 projects that had more pledged funding than their goals but were interestingly cancel.

These projects are:

* + 1. Roman Dead – id 139
    2. LED Sports clothing for running cycling and walking, we make – id 1001
    3. HALLAM new york SMART JACKET 2.0 for TRAVEL with 29 FUNCTION – id 1012,
    4. CORE: Roam – id 1309
    5. Sleepman: The First Sleep Enhancement & Fatigue Alert Device – id 1343

Lastly, there were 50 live projects about 1.2% of all projects. The majority are in the theater, music and food categories. And 4 projects in the music and 4 projects in the theater categories have more pledge funding than their goals.

1. Theater plays projects are the most popular making up 25.91% of all projects, and they are the most volatile. They have the highest success rate of 31.76%, have the highest failed rate of 23.07%, have no failed projects and have the second highest go live rate of 38%. While technology web projects have the highest cancel rate of 28.65% and music faith projects go live 40% of the time.
2. The highest successful projects were accounted for in Feb., 202, May 234 and Jun. 211, these make up about 29% of all successful projects. And from May to Sep. there is a decline in successful projects. Inversely, there is an increase in failed projects between the same months. This may suggest a correlation between successful and failed projects during the summer and fall months. However, there is a bigger gap between these two states during the winter and spring months. Perhaps people are more productive in the winter and less motivated in the summer. Lastly, the canceled and live project counts are flat throughout the year, under 50, and remain relatively close together.
3. What are some limitations of this dataset?

Limitations:

* Have factors and/or indicators/metrics to determine the legitimacy/accuracy of each state
* Is the 4114 count a sample of the population or the complete population because the data goes up to 2017 not to present 2020. If not, would the results of the project count be the same or different if we take the complete population. If the results are similar than this dataset is accurately represented
* Were there any outliers that were taken out of this dataset?
* Create some variables to predict outcomes, i.e. use specific metrics that determine each state of the project

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

Other possible tables and/or graphs:

* Look at the R-square to determine the correlation between the states, countries and project date creations, and between states and backers
* Use percentage of total pivot for easier visualization of the counts
* Use pivot tables to filter data and create charts to help with visualization
* Determine mean, median, variance, standard deviations, z-score to see if the numbers are normally distributed
* or determine the minimum, 1st quartile, 2nd quartile (median), 3rd quartile, 4th quartile (maximum) and Inter Quartile Range (IQR)) to help describe, analyze and validate the data.
* Use scattered plots or box and whiskers plot to view any outliers