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

2. What are some limitations of this dataset?

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

1. Three conclusions to be made about Kickstarter campaigns include that ‘Theatre - Plays’ are overall the most abundant (1066) Sub-Category with about a 51% success-rate with ‘Music – Rock’ being the most successful (260) Sub-Category with a 100% success-rate. The overall most successful month for a Kickstarter Campaign appears to be May (54%) with a total of 386 Campaigns, whereas the worst month to throw a Kickstarter Campaign would be December with nearly a 101% failure-to-success rate. The overall most successful Parent-Category would have to be ‘Music’ with nearly a 3.75x success-to-failure rate in 660 total shows that are not live nor have been canceled.
2. Some limitations of this dataset include if the venues of these shows are more biased toward certain Categories over others. Geography may play a role in Kickstarters; for some areas may be wealthier and may be able to contribute more to a Kickstarter campaign over others in poorer areas. Another limitation would be how long a campaign took to reach their actual goal, whereas we could compare the amount of time that each campaign took to reach its goal.
3. We could create a Geographical table/graph to show what areas showed more interest into what Category/Sub-Categories were more popular in region of the World/United States.

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

Although the Mean will be close to the center of the data being the ‘average’, it will still be off a hair as far as what a true representation of a set middle of what the data is; in comparison to the median. The ‘median’ gives us a true representation of the middle, which is more meaningful than just getting “close” par say like the ‘mean’ gives. “It's best to use the mean when the distribution of the data values is symmetrical and there are no clear outliers. It's best to use the median when the the distribution of data values is skewed or when there are clear outliers”- https://www.statology.org/when-to-use-mean-vs-median/

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 campaigns because of the sheer numbers. In the sample vs population Variances of the successful campaigns there is a ~327 backer difference where as the failed campaigns sample vs population Variance only has ~2 backer difference. It makes sense that in lower numbers outcomes do not vary as much as they do in larger numbers.