Given the provided data, the first conclusion that can be drawn is that theater was the largest category of Kickstarter campaigns, with the highest number of total projects. Additionally, this category had the highest number of successful projects. Within the theater category, plays were the most successful subcategory. Music was the second most successful category, overall, and had the highest percentage of successful projects when compared to live, canceled, and failed projects within the same category. Another conclusion is that the least successful category was journalism, as there were zero successful projects in this category. Considering the information on the bonus worksheet, there is a higher percentage of successful projects than failed projects, for campaigns with funding goals of up to $19,999. For projects with funding goals of $20,000 to $34,999, the percentage of successful projects was lower than the percentage of failed projects.

One potential limitation of the data set could be that the data spans from 2009 to 2017. For an organization that wishes to study past projects, it may be more relevant to examine more recent projects, as keys to success could change over time. Additionally, the background of the assignment states “of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.” Our data set of 4,000 projects is a relatively small sample size when compared to the overall number of projects. In our data set, 2,185 of the 4,114 projects were successful (about 53%), so the representation of successful projects might be higher in our sample than in the number of projects launched, overall. Lastly, I would want to determine if any countries have not been represented in the data set, and if so, how these projects might alter the overall pattern of the data.

Another possible table/graph we could create would be one that demonstrates the percentage of successful projects within each country, compared to the number of failed projects. This could then be filtered by category of project, as well, to determine the success rate based on categories. Secondly, of all the projects in our data set that were 200% funded or greater, only one was cancelled. While this is not surprising, I think it could be beneficial to analyze the goal range, categories, subcategories, and countries within this “percentage funded” subgroup.

**Bonus:**

The median summarizes the data more meaningfully, due to outliers skewing the data set. For example, for number of backers of successful campaigns, the first quartile = 29 and the third quartile = 141. A median of 62 is a better representation than a mean of 195. The maximum number of backers (26,457) is significantly higher than the upper bound (309) of the data. For the number of backers of unsuccessful projects, the median (3) also falls within the first quartile (1) and the third quartile (12), whereas the mean (19) would be above the third quartile.

There is higher variance in the number of backers of successful campaigns, which makes sense given the number of backers for successful projects ranges from 1 to 26,457. The data set for the number of backers of unsuccessful campaigns is not quite as skewed as the number of backers of successful campaigns.