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

* We can infer that the rate of success is solely dependent on the project goal; the lower the funding goal, the higher the chances of success. Although theater is the most prevalent category, music had the highest rate of success with the usage of Kickstarter, and food having the highest rate of failure. The average goal of music projects is about $7,000, indicating that there is approximately a 53% chance of success. We can further see this when looking at sub-categories. We can conclude that plays are the most prevalent, however rock had the greatest number of projects with 100% success rate.
* We can infer that the rate of success is solely dependent on the project goal; the lower the funding goal, the higher the chances of success. However, we can also note a correlation between category and the time of the year (season). For instance, when looking at theater, we see a peak in Kickstarter from months May to July. Because theater represents a big chunk of our data, when filtered out, the peak shifts to January. We see a similar peak when filtering for music.
* We can infer that although there was a steady increase of Kickstarter campaigns from years 2009-2012. In 2013 is when we see a tremendous rise in usage of Kickstarter. This can be due to many things such as credibility of the service itself.

1. What are some limitations of this dataset?

* First and foremost, we must look at the data as a whole and ask, are we looking at the entire population of companies using Kickstarter or are we looking at a sample size? That could greatly affect the dynamics of the trends, as we can be looking at data that may or may not be fundamentally biased. For instance, we are given a background which states that more than 300,000 projects were launched and only and third of them had a successful outcome. However, looking back at the given data, we have a total of 4,114 projects represented with a 53% success rate. Therefore, we can conclude that this sample contains bias based on the selection of the projects to represent Kickstarter.
* Correlation does not imply causation; sometimes we may see correlations in data however it may not directly imply causation, there may be other cofounding factors that may have direct impact on causation. For instance, not having enough measures to analyze the data. Based on this data, we may need more metrics in regards to what makes a projects successful/failed/canceled during different time periods from Kickstarter launching to present day.
* Limitations can also rise when we don’t have sufficient information about the consumer. We may want to include a more informative description of each project so that we know what basis backers made their decision to support a project or not. Additionally, more information on who the backers to see if there is a proposed bias on their selection based on their interests.
* We may want to know how the data was collected to begin with. What sources of data collection were used to gather the data. (ie surveys, interviews, applications etc). This can also lead to limitations with data breach, were all data entered manually? Who has access to this data and alterations to the data provided?

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

* Looking at percentage of pledged per backer.
* looking at the timeframe of each project campaign to see if that has any correlation with success/failure rate.
* Looking at the distribution of projects based on location.
* Looking at the size of the project vs its success rate
* Looking at success and failure rates more in detail per category/subcategory by adding metrics that include values such as min/max/average/range of the funding goal and amount pledged.