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

When looking at the entire set of data, we can see that more than half of projects submitted result in a success at 53%, with only 37% ending in failure and 8.5% get cancelled altogether, while 1% is still live. But if we break it down into smaller groupings, we can analyze the data to see which categories, subcategories, and which month would give us the best chance of success.   
  
Looking at the data through each categories, we can see that there are a few categories that lead the pack in terms of success, music at 77%, theatre at 60% and film & video at 58%. On the other hand, we have lower success rates in journalism at 0%, food at 17%, as well as games, publishing, and technology all have sub 40% success rates. Another observation worth noting is that 100% of journalism projects were canceled, as well as 30% of technology projects are canceled. It also appears that most successful projects have high number of staff pick. Out of 4144 staff pick, film & video has 520 selected, 700 for music, 1393 for theater and 600 for technology. Even though technology as a whole was not a successful project, its hardware category is the only successful project with the most staff pick of 100% successful rate.

Breaking it down further into subcategory for the top successful projects. Documents, shorts and television have 100% success rate within film & video. Similar with classical, electronic, metal, pop & rock for music. However, the success rate for theater’s subcategories are not skewed compare to music, film & video. Plays have a success rate of 65%, following by musical and spaces successful rate is around 45%.

Based on the monthly graph, we can see that May have the highest successful projects of 61%. Majority of the success project comes from Plays in the US and Great Britain. Season takes a large contribution to the success. May is the beginning of summer which gives people more free time to go see plays. On the contrary side, December has the highest failure rate of 56%. This could be due to it being the end of the year and holidays, people have less disposable income to pledge to projects. Besides May having a higher success rate, it also had more than double the number of successes compared to December. May had 234 successful campaigns compared to December, which only had 111.

2. What are some limitations of this dataset?

Given that the current dataset is only 4144 projects from Kickstarter, it accounts less than 1% of the total projects ever started on the site (which is currently greater than 445,000). Due to having such a small sample size, we are unable to grasp the exact trends of success/failures. We are only able to conclude the trends of the given data. Another issue that comes up with such a limited, seemingly random set of data is that it may show that there is more success with one particular subcategory that might not be true when looked at the entire picture. Based on the dataset that was given, we would conclude that many subcategories would have no chance of success on Kickstarter, but that would most likely not be true.

Since we are only given the dataset broken down into the most basic forms of success or failure based on whether a project met its goal, we do not know any other information on what led to said success/failure. We are unsure on if the projects that met success have had a prototype/preview of their item. Kickstarter also allows for the backers to receive perks based on the amount pledged, which could also entice others into backing the product.

Based on given data, I would recommend Theater, specifically Plays and would launch in May. Plays that launched in May had the most successful rate.

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

Another possible graph to use with this dataset would be the use of a pie chart. It would allow for us to view the data in a way to see how much each of our subcategory account for in terms the entire set. We could also use scatter plot in place of line chart to identify the trend.