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

1. Theater is by far the most prevalent category, which leads there to be a peak in Kickstarter from May to July. However, if the Theater category is removed, January has the highest number of Kickstarter.
2. Later years tend to have many more Kickstarter campaigns than earlier years do in this sample. This may be partly due to the financial crisis. This may also partly be due to the fact that Kickstarter is a relatively new company (founded in 2009).
3. While music tends to be the most successful parent category (with 77% of the sample’s Kickstarter campaigns succeeding), music also has the lowest average

2/ What are some limitations of this dataset?

There is evidence that this sample does not fully represent the population of Kickstarter campaigns. For example, the assignment’s instructions mention that only a third of Kickstarter campaigns make it through funding with positive outcomes, but this database shows that more than half of the programs selected are successful. Additionally, this sample includes quite a few Theater Kickstarter campaigns, but Kickstarter’s website does not identify Theater as the most common category.

There are not any metrics that describe the success of each industry/category/sub-category as a whole during the covered time period. It would be helpful to have data that shows whether or not each industry/category/subcategory does well or poorly during the relevant time periods. This might help us account for external factors while analyzing the data.

Kickstarter was founded in 2009, making it a relatively new company. It might be useful to have data on metrics for consumer preferences and/or trust in Kickstarter over time. For example, maybe certain categories/subcategories are more successful during Kickstarter’s infancy while others are not due to lack of consumer trust.

Additionally, there are several outliers in the funding goal/target amounts of money for several categories/subcategories. While we have ‘blurbs’ for each Kickstarter campaign, it would be helpful to have more detailed information into what each Kickstarter hopes to accomplish (in terms of actionable items). It might not be worth including both campaigns with ambitious/expensive goals and campaigns with smaller/cheap goals in the same study.

Lastly, the dataset does not have much information on what donors expect to receive in exchange for their donations (e.g. a product prototype, tickets to a production, t-shirts, etc.). It could impact a donor’s decision to donate if she/he believes she/he will receive something significant in exchange for her/his donation.

The data could be incomplete. Missing values, even the lack of a section or a substantial part of the data, could limit its usability.

* If you're using data from surveys, keep in mind that people don't always provide accurate information.
* Data collected from different sources can vary in quality and format

a limiting condition; restrictive weakness; lack of capacity; inability or handicap: He knows his limitations as a writer. something that limits; a limit or bound; restriction: an arms limitation; a limitation on imports. The act of limiting.

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

There are several tables/graphs we could create. Below is a list of some additional ideas for data displays:

1. Graphs for all the metrics covered with and without “plays” (the most popular category).
2. Tables showing the most successful categories/subcategories versus the most unsuccessful categories/subcategories and detailed data into each category/subcategory’s level of success. Some possible metrics include (1) success rate per category/subcategory, (2) median, max, and min funding goal/target per category/subcategory, (3) median, max, and min of average donation per category/subcategory, (4) median, max, and min of percent funded per category/subcategory, (5) median, max, and min length of Kickstarter campaign timelines per category/subcategory etc.
3. Displays showing the average timelines (how much time companies/organizations have to complete the Kickstarter campaigns) for the most successful and least successful types of programs (perhaps having more time to achieve funding targets would lead to greater rates of success).
4. Displays and tables that exclude significant funding goal outliers.
5. Graphs and displays that break up each category into its specific subcategories to uncover what makes different types of organizations successful. Several categories are relatively broad so it might be that some subcategories within each category have significantly different funding goals, average donations, lengths of Kickstarter timelines, and take place during different periods of time (years and seasons). Breaking up each category into subcategories would help us better understand the data.

Data visualization is the presentation of data (both qualitative and quantitative data) in graphical format. Through data visualization you can easily: make sense of data (especially big data), classify and categorize data, find relationships among data, understand the composition of data, understand the distribution of data, understand the overlapping of data, determine patterns and trends, detect outliers and other anomalies in data, predict future trends and tell meaningful and engaging stories to decision-makers