* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. The categories with the most successful campaigns would be theatre, followed by film & video with music close behind. The success may be due the majority of Spotlight and Staff Picks coming from these categories, so these categories are being promoted more to people and would mean that these projects get viewed more often.
2. The data shows that there is a higher total number of successful campaigns during June and July which then drops off in August which is one of the lowest month for successful campaigns. This may be due to this period being close to the start of the summer holidays, with students finishing exams around this time and with the holiday season approaching there may be higher traffic going the website, but which then drops off during August as people go away on holiday which means there would be fewer people to support campaigns.
3. The higher the campaign goal is the more likely it is to fail, as it means that more backers are required, or the backers need to donate more money to make a campaign successful. This data is shown in the Outcomes by Goal line chart, with the line for successful campaigns getting lower at the higher values, while the failed line rises. This ends in the failed line going above the success line.

* What are some limitations of this dataset?

The instructions on the assignment show that the database was a sample of 1000 campaigns, so the average numbers in each category may not be a true representation of the total number that Kickstarter and Indiegogo would offer. The data also only shows campaigns between 2010 and 2020, but Kickstarter started in 2009 and Indiegogo in 2008. Data from these years could show a different trend in successful, failed, and cancelled campaigns, with people being more favourable to alternate categories in the earlier years.

Kickstarter offers a tiered reward system which can influence how much a backer would donate, and how many backers a campaign gets. Although people would want to support the creator, they would also want something tangible in return for what they are putting in. It would be more beneficial to have a detailed breakdown of what tiers each campaign has and would make it easier to compare the campaigns that offer greater rewards compared to the ones which offer less.

The database uses Unix timestamps, to which a formula is applied to get the date for each project, but this doesn’t show which day in the week the project was launched, or at what time. A further breakdown of this data would be able to show if there is an optimal day or time to start the campaign, as there may be certain times when there is more traffic on the Kickstarter or Indiegogo websites.

* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

A box and whisker chart could have been used to show the data in the Summary Statistics table more clearly, as it would highlight the number of outliers in each section of data and would also make it easier to visually compare unsuccessful backers to successful backers.

A new table and graph could be included to compare the length of time for each campaign and the respective outcome. This would be able to show if it would be better to have a shorter window of time to create pressure and fervour to generate the goal level, or if it is better to have a longer window to get more time to have more backers.

Further tables and graphs could be created for the Outcome based on Goal, so that each category and/or subcategory can be represented. This would then specify the average for each category, so a company can decide based on their campaign what Goal they should use for it to be more likely to be successful. This information could also be represented through a scatter plot to see if there is a trend line towards a higher or lower goal for each respective category/sub-category.