Assignment questions:

1. What are three(+) conclusions we can make about Kickstarter campaigns given the provided data?
   1. Arguably the least successful category on Kickstarter has been entries in the Journalism category, with a 0% success rate. However, the sample size is small, only 24 Kickstarters. If the client is interested in this category, it might worth looking for another avenue.
   2. By sub-category, plays have been the most popular individual sub-category, with a success chance of 66% excluding live Kickstarters. Sub-categories with high success rates include Documentary, Hardware, Rock, Tabletop Games, and Television. Sub-categories with no successes include Animation, Art Books, Audio, Children’s Books, Drama, Faith, Fiction, Food Trucks, Gadgets, Jazz, Mobile Games, Nature, People, Places, Restaurants, Science Fiction, Translations, Video Games, Web, and World Music. Using these categories, the client can help assess their chance of success relative to similar projects.
   3. In general, it appears that having a smaller goal increases your chance of success; 71.08% of goals less than $1000 were met, while only 19.37% of goals greater than or equal to$50000 were met. It also appears there’s a correlation between higher goals and having a project be cancelled, although it’s not as strong (e.g. 3.97% of goals less than 1000 are cancelled, while 22.52% of goals greater than or equal to 50000 were cancelled). The client can use these criteria to try and set as low of a goal as necessary to complete the project successfully, which might open up more traditional funding opportunities.
   4. It appears that there are fewer examples of successful Kickstarters created around December (count of 111 successes in Dec.) when comparing this dataset across all years (average of 182 successes). This might be because folks are already spending a lot of disposable income around the holidays (although on the other hand you might think that would have a positive effect). However, the rate projects fail in December is near the mean of failure rates by month (Dec. count: 119; 12-month average is 127.5). I would hesitate to pass this advice to a client without further research, but it was popped out at me so I’ll leave it here.
2. What are some of the limitations of this dataset?
   1. Some limitations of this dataset are that we don’t know how the sample was collected from the total dataset. The original dataset contained over 300k Kickstarter entries, while this only contained 4k. Not knowing the distribution of states for the original entry means any conclusions drawn here may be of unknown value to the client. For example, I know for a fact there have been successful Kickstarters for video games because I have participated in them.
   2. Another limitation is the changes to the Kickstarter platform that have occurred over time. These changes could skew the data in unpredictable ways due to poor sample size and bias for data when the platform was just starting, or in unknown ways due to rule changes for who and how you can apply, cancel a project, refund policy, etc. Some of these limitations may be less from the dataset itself and more on the domain knowledge required to understand Kickstarter as a service. It would behoove a client to know the rules for Kickstarters very well, and also any analyst using this data professionally.
3. What are some other possible tables/graphs that we could create?
   1. Create graphs for various ranges of “ambition” based on the monetary goal set, to see if more ambitious goals are more likely to fail (see bonus graph example)
   2. Create a line graph showing if Kickstarter success and/or fail rates have increased or decreased over time, and research any Kickstarter policy changes that might have influenced these results.
   3. Research the rules for cancellation of projects, and compute a distribution of how long it takes projects to be cancelled, to better ascertain what purpose cancellation is serving in practice.
   4. Create a table listing discrete status rates and status percentages for particular categories and subcategories within those categories, to show chance of success based on category/subcategory. The client may be looking to try a particular Kickstarter and finding more comparable results amongst the specific category might be more valuable.

My other notes:

My client wants to start a Kickstarter. **What kind of trends/info could be gained from existing information about Kickstarters to help gain success for this effort**? Keep in mind that only 1/3 of past Kickstarters have made it through funding with a positive outcome (not sure how that is defined, but have some ideas).

Some helpful info:

Getting funded requires meeting or exceeding the project’s initial goal

There have been over 300k projects launched on Kickstarter, with a wealth of metadata about the progress of each project, as represented in the spreadsheet (spreadsheet contains 4k past projects; not sure methodology used to choose selection).