Responses to Excel HW Questions

1. What are three conclusions we can make about Kickstarter campaigns given the provided data?

There are a number of conclusions we can draw from the data that follow from observations, but are pretty obvious and not too interesting. For instance, the first conclusion I can draw from the data is that theatre-related Kickstarter campaigns appeared as the most popular type of campaign, while no journalism-based Kickstarter campaigns succeeded (all 24 failed out of the 4000+ campaigns tracked in our data). The above observation leads to the conclusion that people are generally interested in theatre and have little to no interest in journalism on Kickstarter, if we presume donations are an approximate signal of interest in some subject or campaign. These observations come from the first pivot table.

The second pivot table, which shows a sub-categorical breakdown, gives us more insight into exactly what type of theatre performances people were interested in funding. In this case, it would appear that plays dominate with 694 successes against 353 failures (so, about 66% of all plays achieved their target) in receiving successful amounts of funding, where other fields usually associated with theatre, such as drama, received a shocking 80 failures (100% rate).

Following the same approximate analysis of these charts and data, it appears that campaigns starting out in the middle of the year (June) are more likely to be successful than campaigns starting towards the end of the year (December). The higher failure rate in January is an expected consequence in the general decrease in successful campaigns towards the end of the year, though it is important to keep in mind that the data describes campaigns that are all unique, so no direct comparison about the “inverted” trend or relationship between successful campaigns and failed ones can be made. However, such a trend is suggestive and would be worth investigating. For instance, the conclusion that less campaigns are successful in December may prompt a further question such as “Are people less likely to donate to Kickstarter campaigns during December due to holiday spending?”

1. What are some of the limitations of this dataset?

* The dataset only captures donations made to the Kickstarter platform, and we are not sure if Kickstarter holds occupies enough of a market share in the category of crowdsourcing platforms for us to identify any statistically significant trends (due to issues related to representation). Representation becomes even more of a problem when you start considering the drastic variance between sample means of smaller countries and the overall mean (the average of the set of 4115 observations)”. You can see this effect if you filter by country.
* Do we really think that 4115 observations is anywhere enough data to allow us to make any meaningful conclusions if the population is closer to 300,000 according to the homework’s description: “Of the over 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.” 4,000 observations is about 1.1% of 300,000. So you are trying to make conclusions about Kickstarter using a set that captures only 1.1% of all projects on that site. Also, we don’t know how these 4000 projects where sampled when the excel sheet was generated, leaving huge room for bias in the statistical results.
* The homework specifies that the goal is to uncover hidden trends that may help inform someone interested in starting a Kickstarter campaign on what to do to have a better chance at receiving full funding for their project. The only metric here is whether or not enough funding was obtained, and that is the only information that this set of data has the potential to reveal. There is not even the beginning of a basis to explaining if a causal relationship exists between the funding and any other factor: category, timing of the campaign’s launch, number of donors, country, and so forth. If you can’t prove a robust causal relationship between funding and timing, for instance, then any conclusions you’re drawing from this data are more or less guesswork and speculation.
* The data looks like it hasn’t been updated in two years. So, it may not be reflective of the current funding trends.

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

Something that might be interesting might be to find create a new column that has the length (word count) of each of Kickstarter campaign blurb displayed, and then create a pivot table of this in relation to the outcomes (i.e. success, failure, cancelled, and live). So, the idea there is that you might be investigating if the length of a blurb has any impact on the funding outcomes of Kickstarter campaigns. There are a lot of columns to play around with in this excel sheet, so you could conceivably just pair the outcome (success, fail, live, cancel) with any thing that looks like it might be a plausible independent variable (this assumes you are creating tables and graphs for the purpose of trying to see what impacts outcomes, if that wasn’t a natural conclusion).