Homework Assignment 1

Questions

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

* In summary there appear to be four factors that have the most significant influence on success; Project type, Staff Pick, Length of the Kickstarter and Target Goal.
  + Project Category - The following categories appear to have a good following in kickstarter
    - Rock, documentaries, hardware, tabletop games, television, shorts, nonfiction, electronic music, classical music, pop, metal and “radio & podcasts” have a 100% success rate
    - Plays are the most common Kickstarter project type and have a high success rate (65.10%)
  + Staff Pick - Projects that are flagged staff pick have an 87.25% success rate
  + Length of the Kickstarter - meaning the time between start and deadline, is a moderate influencer. This data tells us what should be avoided versus the best option.
    - Based on the date the timespan to avoid is 45 days <= Timespan < 60 days (36.7% success).
  + Target Goal has a significant impact. Recommendation is to keep your kickstarter under $5k

Details and analysis below.

* + Program type can be a significant influencer in a kickstarter success. It is not a 100% indicator of success but it does provide some guidance in specific scenarios.
    - Plays are the most common kick start program. Their success rate is better than the average kickstart program.
      * Plays represent 25.91% (1066) of all kickstarts within the program. The next closest are Rock kickstarts at 6.32% (260)
      * Plays have high success rate. Average without plays included is 48.92% (1491 of 3048). Plays have a success rate of 65.10% (694 of 1066)
    - While plays do seem to have a high success rate there are many other subcategories with 100% success rates.
      * Rock, documentaries, hardware, tabletop games, television, shorts, nonfiction, electronic music, classical music, pop, metal and “radio & podcasts”
    - Many other categories have a 100% fail rate which leads one to believe kickstart is not a good fundraising program for these subcategories
      * Video Games, People, animation, gadgets, places, drama, mobile games, jazz, fiction, restaurants, nature, children’s books
  + Time of year a kickstart is initiated appears to play a role in the success of kickstart programs
    - Based on the Data December is the worst month to start a kickstart
    - Feb – Jun appear to have the best success rates
    - More review is required. I would look at this based on annual rates as well. Perhaps the failures are recent, and the successes were when kickstart first began.
  + Staff Pick appears to be a significant influencer for success
    - 87.25% (486 of 557) of Staff picks succeed while only 47.76% (1699 of 3557) without staff pick succeed
    - Those without Staff Pick have a higher cancelation rate 9.7% (345 of 3557)
    - Staff Pick does appear to impact. If there is a way for the kickstart owner to get their project be staff picked then this data shows it provides value. Thankfully they do provide guidance.
    - <https://www.kickstarter.com/blog/how-to-get-featured-on-kickstarter>
    - Based on their guidance it would seem just following their suggestions on what draws a staff pick selection would be good advice regardless of being selected.
  + At first glance Spotlight looks like an influencer but the data is far to consistent to be a real influence. This appears to be a flag that kickstart uses to show what programs were a success – without needing to use the state column.
    - 100% of kickstarts Spotlighted are Successful
    - 100% of kickstarts not spotlighted are either live, failed or canceled
    - This level of data consistency means this flag represents something other than an influence for success. It is more likely a flag set once a kickstart has Succeeded. Validation is required.
    - <https://www.kickstarter.com/spotlight>
    - The website confirms this analysis. Spotlight is used to flag successful projects and allows the project owner to talk about their success.
  + Targeted end time for the kickstart program
    - I created buckets to look at start to deadline to determine if the number of days for the deadline would be an impactor.
    - 15<x<=30 and 30<x<=45 are the only real data points with enough data to be relevant and they have very similar success rates (53.2% and 57.6% respectively)
    - Oddly enough the 45<x<=60 has the highest cancelation rate and a very low success rate (36.7%).
    - <=15 and 75-90 day kickstarters have a high success rate (66%, 92%) and 60-75 is slightly better than 15-30, all three represent small data sets, not statistically relevant. I would not use this data to support using these data ranges even though they appear to have higher success rates.
    - It looks like 15 to 45 days may be the “sweet spot” but it is not a huge factor.
    - Really this factor points to a question why does the 45<x<=60 have a high cancel low success rate.
  + Goal appears to have a strong impact (bonus work)
    - <1000 has the highest success (71%) and lowest fail rate (25%)
    - There is a steady trend of success rate dropping through every Goal bucket defined.
  + Country
    - This shows that three countries are the primary submission point for kickstarts; United States (74%), Great Britain (15%), Canada (4%)
    - Data is so heavy in one country it is hard to say this is much of a factor
    - We do not really know what factors are used to define country
      * Nationality, IP location for submission etc…

1. What are some of the limitations of this dataset?
   * It is interesting data, but it really does not tell you the techniques used that lead to success or failure. Not all kickstart programs are the same. You are forced to look at the data as though the only factors impacting success are the category/subcategory, the financial goal, staff pick, spotlight or the description of the kickstart program.
   * This data does not tell you the demographics of the people that donate to Kickstarter. Perhaps you have a great project but it will fail due to the people that are interested in your project have no awareness of kickstarter. The project type helps identify some areas that appear to be challenged in this way.
   * This data does not give you any idea of the success of other fundraising techniques. If you have already decided to do a kickstart then this could help. But if you are looking to do fund raising this data is only part of the picture. You would need other sources of data to decide the best fundraising option.
2. What are some other possible tables/graphs that we could create?

* For my review I added % to the raw number analysis requested in the HW. Raw numbers are helpful, especially when trying to determine if a data set is statistically relevant but the two combined help make the data more actionable.
* Country analysis
* Length of the Kickstarter tiers
* Staff Pick influence
* Spotlight Influence
* Goal tiers (this is the bonus work)
* Show the success by year and month. This would help expose the success of kickstarter as a fundraising technique. Perhaps its peek was in 2015 and people are starting to not use it anymore. The recommended monthly analysis does not expose this type of trending.
* More analysis that compares multiple factors at once.
  + Since goal and deadline are factors that the creator can control combining these factors to see if there is sweet spot for their specific Kickstarter category would be recommended.
    - Ideally you should merge all four major influence factors to see what is your best option instead of looking at these factors individually.