Homework Assignment 1

Questions

1. What are three conclusions we can make about Kickstarter campaigns given the provided data?
   * First level review there appears to be a few data points that influence success.
     + Program type may be an influencer in kickstart success.
       - 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
       - 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)
     + 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.
     + 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…
2. 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 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.
   * Nothing defines the columns “spotlight” or “staff pick”. Based on the value these seem to have it would be best to understand the actual column before it is safe to draw real conclusions from the data set. Based on the success rates spotlight appears to be used to highlight successful kickstarts and not something that influences success.
   * Taken at face value spotlight appears to be the most influential factor for success (based on available data)
     1. 100% of kickstarts Spotlighted are Successful
     2. 100% of kickstarts not spotlighted are either live, failed or canceled
     3. 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.
3. What are some other possible tables/graphs that we could create?

* As you can see from my spreadsheet I added %tile analysis to get some very high level % metrics to start my drill in analysis. Raw numbers are helpful, but percentiles do help when trying to review impact.
* Creating thresholds based on goals would be good. From here look at the success rates within specific tiers of the goal.
  + Oh….I see this idea is the bonus work….more to come when I finish that tab.
    - As expected….smaller kickstarts are more successful
* Charts that use the staff pick and spotlight flags to see how valuable these factors are to success.
* Show the success trends by year and month – perhaps kickstart is becoming less successful over time.
* Calculate the time between create and deadline. Perhaps failure is due to very short or very long deadlines.
  + As it turns out quick (<=15 day) quick starts have a high success rate (66%) though the 75-90 day are higher (92%) , but bother represent fairly small data sets so probably should be thrown out. It looks like 30-45 days may be the “sweet spot”.