**Excel Challenge**

* Given the provided data, what are three conclusions we can draw about Kickstarter campaigns? Explain the reasoning behind your answers.
  + 1) From the category stats analysis, we can see that the number of successful campaigns > failed > cancelled > live. The campaigns under the category film and video as well as music are mostly successful. The campaigns under the games, the publishing or the food category mostly failed. Lastly, all the campaigns of journalism (which were only 24) were canceled.
  + 2) Looking at the subcategory stats analysis, it is clear that the plays subcategory is the most popular one, with 2/3 of the campaigns being successful and 1/3 being unsuccessful. It is also important to note almost all the campaigns under the documentary, hardware, rock, indie rock and shorts were successful. It is also important to know that the subcategorys video games, animation, children’s books, fiction and people tend to fail, although some of them have a reduced sample size so it would be difficult to generalize.
  + 3) Looking at the launch date statistics, the total number of campaigns is greater during the months of May through July. Looking at the number of canceled campaigns, it looks like it is pretty consistent across all the months of the year. As a highlight, the number of successful campaigns tends to decrease during both November and December.
* What are some limitations of this dataset?
  + One of the biggest limitations of the dataset if that we are only analyzed 4,000 out of the 300,000 projects launched at Kickstarter. This is only 1.33% of the population, meaning it might not be sufficiently big to represent the whole population. If choosing other 4,000 projects, results may vary greatly
  + We should have more information about the type of promotion that each of the campaigns receives to see if there is a correlation between the marketing of a campaign and its success.
  + We should study other factors that affect the campaigns such as where are they conducted, how long it lasted, how much each baker contributed etc
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + Duration of campaign – We have the raw data about the launch date and the deadline of each campaign. Therefore, a histogram comparison of duration versus status of a campaign could be useful to determine if there is a correlation between these two.
  + Geographic Location – By collecting information of the geographic location of the campaigns, we may see a correlation of some areas where most campaigns are successful versus other areas where most campaigns failed
  + Staff pick – Study with a pivot table and and a column chart if the campaigns with “staff pick” are more successful than those that don’t.

**Bonus Statistical Analysis**

* Use your data to determine whether the mean or the median summarizes the data more meaningfully
  + The median summarizes the data more accurately because the mean is prone to be affected by outliers. This is definitively the case in this scenario since the median and the mean are really different from each other.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  + The variance and the standard deviation are the statistics used to determine the variability of the data. Since the Standard deviation and variance is much greater in the “Successful Campaigns”, there is more variability in the successful campaigns. This makes sense because you will expect the unsuccessful campaigns to have in general a really low number of bakers (definitively under the goal).