Excel-Challenge-Jeannaej

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   1. One conclusion that can be drawn from the Kickstarter campaigns dataset – Category Pivot – is that campaigns that are categorized as theater related have the most campaigns and the greatest number of successful campaigns.
   2. A second conclusion that is drawn from the dataset – sub-category pivot – there is 13 sub-categories – Animation, Children’s books, Drama, Fiction, Gadgets, Jazz, Mobile games, Nature, People, Places, Restaurants, and Video Games – do not have any successful campaigns.
   3. A third conclusion that can be made from the dataset – date created pivot – is that May is the month that the greatest number of campaigns were created between 2009 -2017.
2. What are some limitations of this dataset?
   1. One limitation of the data is the lack of information on cancelled campaigns. The lack of this data excludes the information on whether the campaigns were being canceled due to the lack of interest or the lack of funds.
   2. Another limitation is the “staff picks” and “spotlight” data, we do not know whether a campaign becomes a ‘staff pick’ or ‘spotlight’ campaign because it is already gaining large interest. And if being a selected campaign increases it chances of being fully funded.
   3. Another limitation is we don’t know what causes the campaign to be successful vs failed vs canceled. Are campaigns successful if it is fully funded and fail if it doesn’t? Are they successful if the project is created with the funding or if the project created was successfully sold to the public? Are campaigns considered failures even if the project was created but then no one brought the project? Or the success/fail state is determined on collecting pledged for a certain amount of time and if the campaign is not funded in that time period it failed.
3. What are some other possible tables and/or graphs that we could create?
   1. Percent of funds in each state by the goal
   2. We could also figure out how long each campaign is, in relation to percent funded and the state of the campaign
   3. The length of campaign, with its state and number of backers

Bonus Statistical Analysis

1. Determine whether the mean or the median summarizes the data more meaningfully?
   1. In both the successful and unsuccessful data, the median summarizes the data more meaningfully because in both cases there are several outlier campaigns that pulled the mean/average to a greater amount than the median and away from the majority of the data.
2. Determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   1. There is more variability with successful campaigns, as seen with a larger standard deviation of 844 campaigns oppose to a standard deviation of 61 campaigns for unsuccessful campaigns. This makes sense because according to the data the minimum backers for successful campaigns is 1 and the maximum is 26,457 backers, this means some campaigns take the funding of only 1 backer to fulfill their goal or 26,457 backers. While unsuccessful campaigns had a minimum of 0 and maximum of 1,293, the smaller variance make sense because if a campaign is not gaining enough interest and backers to pledge and meet their goal, they will fail.