Lesson 1: Homework – Excel Challenge

Tiffany Harris

SMU-DAL-DATA-PT-12-2020-U-C

* Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
  + The theater, music and film/video categories are the most populate Kickstarter campaigns. The subcategory, Plays, make up over 24% of all campaigns.
  + Theater has the lowest number of backers for all their successful campaigns, yet they were the most popular.
  + Projects with less than $1,000 goal are the most successful campaigns (71%).
  + Projects with more than $50,000 goal have the greatest failure rate (58%) and largest number of cancelations (23%).
  + May, June and July have the largest number of campaigns, while December is the lowest.
  + The number of successful, failed and canceled campaigns in the technology category is almost equally distributed (failed 40% and the difference split between successful and failed).
* What are some limitations of this dataset?
  + With over 300,000 actual Kickstarter campaigns, this dataset samples only 1.33% of the entire population of campaigns. There are challenges and broad assumptions that have to be made when analyzing a small sample of such a large population. There is a higher degree of variance that is included in the data to predict the population.
  + Comparing various currency for different goal levels is inconsistent. The exchange rate and value of money is not adjusted for in the calculations.
* What are some other possible tables and/or graphs that we could create?
  + Having a chart to compare percent funded to goal would normalize the smaller goal campaigns with the larger goal campaigns.
  + A chart to review the number of backers per campaign state would give insight into the popularity of the campaign.
  + Knowing which categories were most successful in different countries would provide awareness to the national cultures that influence different trends.
* Use your data to determine whether the mean or the median summarizes the data more meaningfully.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?