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?
  + Artistic categories such as theater, music and film/video are the most popular Kickstarter campaigns. The theater category has the lowest number of backers for all their successful campaigns, yet the theater category has the highest number of campaigns. Additionally, the subcategory plays make up over 24% of all campaigns.
  + Projects extreme low and high goals represent the greatest success and greatest failure. Those with less than $1,000 goals are the most successful campaigns (71%). Those with more than $50,000 goals have the greatest failure rate (58%) and largest number of cancelations (23%).
  + There is a seasonality when Kickstarter campaigns are created. May, June and July have the largest number of campaigns, while December is the lowest.
* What are some limitations of this dataset?
  + With over 300,000 actual Kickstarter campaigns, this dataset of 4,114 campaigns represents a sample of only 1.33% of the entire population of campaigns. There are challenges and broad assumptions that are made when analyzing a small sample of a large population. There is a higher degree of variability to predict the population when analyzing small sample datasets.
  + Comparing various currencies for different goal levels is also 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.
  + There are many outliers in the dataset. Because the successful campaigns have backers ranging from 1 to 26,457, the mean of 194 does not represent most of the backer count. Instead, the median of 62 is more statistically significant. The same conclusion applies to failed campaigns. The spread between minimum and maximum number of backers is not as large as successful campaigns; however, the spread between the minimum of 0 backers to 1,293 backers is large enough that the mean of 18 is not as relevant as the median of 4. The extremities cause the median to be more useful than the mean.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  + The standard deviation in number of backers with successful campaigns is 844. This standard deviation is larger than the standard deviation of 61 in number of backers with failed campaigns. This statistic can be normalized using the coefficient of variance (CV) which is 4.3 for successful campaigns and 3.5 for failed campaigns. The larger CV in successful campaigns proves successful campaigns have higher variability over failed campaigns. The higher variability seems practical given the number of factors that contribute to a successful campaign.