Amol Bhatia  
Starterbook- Excel Challenge  
Homework 1  
1/30/2021

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?  
   1. Given the data, there is a high amount of Kickstarter campaigns for Theater (33.86% of all campaigns) of which 76.53% are Plays.
   2. However, Theater campaigns is not the most successful campaigns, but rather second. The honor of most success belongs to Music beating out Theater 77.14% to 60.23%. In general, campaigns will be successful (2185) than failures (1530)
   3. The month with the highest success-to-failure ratio is February (202-106) followed closely by April (192-102)
2. What are some limitations of this dataset?

We do not know the outcome of the Kickstarters. Just because a Kickstarter is “successful” or fully funded, does not mean it is successful for the backers. This data is not able to truly show if a Kickstarter was able to kickstart a business, rather that the idea was good enough to raise enough initial capital.

There are also data sets of campaigns with the goal of “$1” or “$2” or *“NOTE: THIS PROJECT IS ALREADY 100% FUNDED!!!*

*This is an "Extended Campaign Run" for anyone who wants a CD of my seventh solo album.”*

The data is not filtered or cleaned and includes repeats of previously successfully funded projects as well as campaigns started for fun. We also do not know if backers received any additional objects from the campaign such as tickets to concerts or were given physical objects in exchange for backing.

1. What are some other possible tables and/or graphs that we could create?

It would be interesting to see a breakdown by country (given there was more data from outside US to balance out the overwhelming majority) and see the amounts that are contributed as well as the backing power from each of those countries. It would also be interesting to see what other countries are more likely to back comparatively to summarize what cultures value differently.

Bonus Statistical Analysis

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

Due to the multiple outliers and in the successful data set, I would use the median to summarize the data more meaningfully. Having a campaign with 26,457 backers is a huge outlier that will be heavily affecting the mean.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

There is more variability with successful campaigns as each backer will contribute a different amount of money to the campaign and therefore will result in a varied number of backers for different goal amounts. Another reason is, if a campaign has failed, it is because it has not reached the needed goal amount. Usually due to lack of backers. This means that most of the numbers will be low and thus have lower variability.