Data Science

# Homework 1

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

One conclusion we can draw from Kickstarter campaigns is that the subcategory “plays” is far more successful than any other subcategory. There are less successful campaigns from June to September. The most successful parent category is theater, even though they have more failed campaigns than every other category. Technology and food truck the least successful campaigns. The data also showed that when you set lower targets, you were more likely to succeed in the campaign.

1. What are some limitations of this dataset?

Some limitations of this dataset are that there have been 300,000 campaigns but we only got 4000 of those campaigns, which comes out to about 1% of the available data. There could be other trends that would have shown if the entire dataset was available. The data is from 2017, so it is not the most recent set of data. We do not know why the successful campaigns meet their goals because we don’t know how much advertisement budget was allotted. We also do not know how many of the donors were large versus small, so we don’t know how that affected the success rate of each campaign. We don’t know the demographics of the people running the campaigns and the people who donate, which provide insight on how to spend advertisement money in future campaigns.

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

Another possible table and bar graph is one that depicts percent funded pledged money (over 100%) for categories or subcategories, to only display who received the most funding out of the campaigns that were successful. These graphs can be broken up between countries as well. There can also be a table and graph depicting how fast it took specific category or subcategories to finish their campaigns but finding out the days between the start and end of the campaign. You could also make a graph that shows what categories are live at any given time.

Bonus Questions

1. As shown on the Outcome Based on Goal Graph, most of the successful campaigns (70%) are the ones with a smaller goal than $1000. Therefore, the median of $62 is a much better representation of what an average backing will look like for a successful campaign, versus the $194 mean of the same data.
2. There is a lot more variability with successful campaigns versus failed campaigns because there is a much higher standard deviation in the successful campaigns. This does make sense because there are always going to be more monetary pledges to a successful campaign, so that even if the average is $62, there were some donations to successful campaigns closer to $1000. Versus a failed campaign will get smaller amounts and it will die off before the goal is met, meaning the spread of donations is smaller. The higher monetary donation spread in successful campaigns supports the higher variability.