Alice Liu

Module 1 Challenge

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. For all countries, the percentage of successful campaign is higher than a failed campaign.
   2. Out of all the subcategories, science fiction and mobile games were the only two fields that had a higher failure rate compared to their respective success rates.
   3. Just by looking at the crowdfunding goal analysis graph, the more you pledge, the more likely your campaign will succeed.
2. What are some limitations of this dataset?
   1. The amount of data isn’t the same from each country. For example, Italy has 48 total crowdfunding campaigns, whereas America has 763. This causes bias towards America when looking at the “world” crowdfunding analysis.
   2. When looking at the parent categories, journalism only has 4 data points. The data is biased towards journalism (100% success rate). If I were to run a model on this data, I would need to drop the journalism data points unless I get more data.
   3. Not that many companies who invested between 10,000 upwards. Even if we increased bin size, the number of projects within the 1,000-10,000 range would still be skewed. I would either get more data or report that the data ranging from 10,000 onwards is not statistically significant.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. I would adjust the currencies to equal the same standardized amount (ex: USD in 2020), then compare summary statistics in multiple tables between each country’s goal and pledged amounts. Since different countries use different currencies, it would provide better insight into the ‘Outcome Based on Goal’ graph.
   2. I would make a pie chart with all the countries’ contribution to the dataset. As I mentioned in the limitations, it would give me insight into how the data might be skewed.
4. Use your data to determine whether the mean or median better summarizes the data.
   1. Since the difference between mean & median is very high, and there is high variance, the data is clearly skewed. Therefore, median is the best measure of central tendency.
5. Use your data to 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. Purely from a data sense, it doesn’t make sense because more data points for “successful”, but in terms of looking at what the variables mean, it makes sense because more backers would increase the likelihood of a successful outcome.