Week 1 HW 1 YU

1. **What are three conclusions we can make about Kickstarter campaigns given the provided data?**

* From this data we can draw several conclusions. Most notable is the fact that, of all the 4064 completed campaigns (not counting the “live” campaigns), approximately 54% were successful, 38% were failures, and 8% were cancelled. This suggests that on average a campaign is more likely to succeed than not. This percentage does not imply a strong likelihood, but rather a moderately favorable likelihood.
* **By Sub-category**

In regards to campaigns based on sub-category, the data shows that “plays” have the largest number of campaigns. (with or without including “live” campaigns). If we then analyze the outcomes based on sub-category, the data shows that the “plays” sub-category also accounts for the largest portion of the total successes. The “plays” sub-category also has an approximate 66% success percentage (obtained from the total number of successful “plays” over the total number of “plays” …excluding live campaigns). This success percentage is well over the average success percentage. Together these facts suggest that “plays” sub-category is one of the most common and most promising campaign type.

* **By Month**

In regards to campaigns based on month, the data shows that number of campaigns for each month is fairly similar.

From the month chart we can also conclude that number of successful campaigns increases as we go from January campaigns to February campaigns. This trend reverses from February to March as the number of successful campaigns decreases back to an amount similar to the amount in January. From here we see a continuous increase from March to May, at which point we see a consistent decrease in the number of successful campaigns from month to month until September. From there we have a swift rise then a fall from September to December. This information, along with its 60% success percentage, suggest that month of May produces the most successful campaigns.

1. **What are some of the limitations of this dataset?**

* The size stands out as a limitation as the bigger the sample size, the better the representation of the actual population.
* The categories provided potentially influence the generalizability of the model as the inferences made would consequently be limited to the categories provided and would not be an entirely fair representation of the success of general Kickstarter campaigns types.
* The difference in currency could factor in as a limitation. Note, currencies have conversion rates. Thus, any analysis regarding monetary amounts would not be entirely accurate without first finding a specific unit of reference (as there would be a lot of room for misinterpretation). In that regard, many of the intuitive assumptions made could be flawed.
* Other limitations to this dataset could be the underlying elements that factor into people donating to a specific campaign. For example, how controversial (relative to the time) was the television project that was not funded in comparison to the one that was. Another example would be the amount of money requested vs what people were willing to donate to that project at the time. Some projects could potential succeed with lower goals. In other words, overestimation of value could influence a project’s result.

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

Another possible graph would be one that showed the relation between “goal amount” and “outcomes”. The graph could be a stacked column graph filtered by year to show how “goal amount” affects “outcomes” from year to year. This graph could potentially point to how economic, social, and political factors of a year might play a role in how much people were willing to fund specific projects.

You could also make a column chart comparing yearly success percentages. This chart would give you a visual representation of the trend of success over the years for the sample campaigns.