Homework 1 Report

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

First, the category of theater spearheaded in terms of successful campaigns. However,

the number of successful theater campaigns dropped between the months of September to December among all years. In addition, the total number of campaigns decreased between 2015 to 2017. Contextually, there may have been an economic situation that could have influenced consumer buying and spending practices during these following years. Lastly, based on the total counts of campaigns per country, the USA came in the lead in terms of the number of successful campaigns. This lead could be attributed to the fact that the US exceeds in the amount of pledge funding compared to other countries.

1. What are some limitations of this data set?

According to Kickstarter’s background information, it is the goal of Kickstarter to collect funding to meet or exceed the company’s goals. However, there are no detailed metrics or descriptions of what is considered a “completed” goal. In other words, the company has not identified what variables and outcomes it expects for a goal to be achieved. For example, would a campaign that had sustained funding for a stable period meet the company’s goals? Or, would a peak number in pledges with a significant drop in pledges over time be considered a desired outcome? Identifying measures and metrics can help data scientists and data analyzers understand what variables or trends that should be followed to meet the needs of the company.

Although there is a variable called ‘state’, which defines the outcome of a project, another variable to consider is the total cost spent on each project. It is important to identify whether a project went above or below budget. This helps us understand whether the company is making any profit or losses overall. In addition, the company can identify where money is being directed or where it should be better directed.

Lastly, it would be beneficial to add measures of central tendency (mean, median, mode) and measures of dispersion (variance, standard deviation) in order to see which data points, lie within the sets in data. This can help the company understand what categories and which locations resources could be diverted to. In contrast, it could also identity why certain data points are outliers from the mean and why they are less popular.

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

* Box Plots: These structures are excellent visualizations of identifying outliers and data points centered around the mean. Particularly, visualizing the number of successes per country using a box plot would be a better of way of showing aggregations of data by geographical location.
* Scatter Plots: Scatter plots would be a beneficial way of representing relationships between different sets of variables that are independent of each other. For example, it would be interesting to see a scatter plot detailing a correlation between ‘staff picks’ and ‘state.’ There could be observations identified that are not typically seen on a simple spread sheet.
* A pivot table that displays a break-down of each sub-category under its parent category. Parent categories are too broad. Data in this set can be more understandable when it can be broken down into sub-categories.
* A pivot table that displays the length of time a campaign began and ended, average donations, and goals of each campaign to highlight whether objectives and outcomes were met.