Heather Tzou

HW1

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   * More successful projects compared to failed and canceled projects
     + Number of successful projects: 2185
     + Number of failed projects: 1530
     + Number of canceled projects: 349
   * Majority of successful projects is in theater and plays dominate in this category
   * It seems that the month in which a project is launched has no relation to the project’s state (ie. there’s no season or pattern where you see more successful/failed/canceled project compared to others)
2. What are some limitations of this dataset?
   * Variability in the dataset (number of projects within each category and subcategory differs) which could skew the results
     + In this dataset, it is clear that theater has the most number of projects compared to other categories. Hence, it could be the reason why that there are more successful outcomes (more project submissions 🡪 more chances of success). Also noticed that theater, although it has more successful projects, it also has a lot of failed projects. If you look at the big picture, theater has the most number of successful projects but it also has the most number of projects than any other category.
     + The launch and deadline date differs for each project, which could affect the projects’ state. For example, the ones that got more time has more opportunity to raise money to reach/exceed their initial goal.
   * Sampling of the data: There are project submissions from different countries. This could also affect how we view overall successful/failed/canceled projects in the campaign. 74% of the overall projects come from the U.S.
   * We’re only evaluating 4000 out of more than 300,000 projects launched on Kickstarter, which is roughly 1.3% of the entire projects. Results here cannot be generalized and be deduced that one category is better or more likely to get funded than the other.
3. What are some other possible tables and/or graphs that we could create?
   * Scatter plot between time (how long from creation date to end date) and number of successful and failed projects to see if there’s any correlation between time to get funded and number of successful projects/failed projects/canceled (if there’s a positive correlation between time and the number of successful projects as well as time and the number of failed projects then we can say that the project that gets is more likely to reach its initial goal)
   * Line graph that looks at backers count and number of successful, failed, canceled projects to see if there’s any relationship between the two and if the number of backers affect whether the project becomes successful or not
   * Individual bar graphs of the different categories that show how many successful, failed, live, and canceled projects there are for the subcategories within the category just to see each category separately