Anthony Giusti

Homework #1

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

* Theater is the most popular category for a Kickstarter campaign and even more specifically Plays are the most popular subcategory by a large margin.
* Kickstarter projects specific to the category “Tech” have a high chance of being cancelled. Within other categories a project being cancelled seems to be fairly rare except in Tech.
* In April/May projects are more likely to be successful. The number of failed projects increases as well but since a large portion of successful Kickstarter projects are plays, music, concert and arts related; one theory for the high success rate is these activities being more popular as summer approaches.

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

* One flaw of this dataset is not knowing the origins or marketing prior to it landing on Kickstarter. There may be certain categories that are well advertised or go viral on social media that can dramatically change whether projects are successful or not. This is not reflected in this dataset.
* The subcategories can be divided into even more specific subcategories. For example, the subcategory Rock has both musicians who are trying to make an album, scheduled tours, or plan concerts. If this were broken up into Concert vs. Tour vs. Musical Project (Album or EP), we may be able to dig deeper in this data. Another example could be television which could have even more subcategories.
* Another aspect that is missing is understanding when projects receive the most amount of money during their campaign. As I was viewing the data, it interested me if there was a breaking point where people received a lot of money at once. For example, maybe when a project hits 80% funded people are inclined to donate to it because they think the project is close to becoming successful. Maybe these donations are larger once they cross a threshold.

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

- We could do a regression analysis to compare the amount of backers to the state of the project or category.

- We could do a scatter plot to see what percentage of unsuccessful or canceled projects vs percent funded. Maybe there is a specific percentage range that is difficult to overcome.

- A line graph would also be a useful way to view the length of a campaign vs amount raised or state.