Module 1 Challenge: Crowdfunding Analysis

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
2. 57% of the crowdfunding projects that were started were successful while about 37% of crowdfunding projects fail and only 5% of projects that are posted are cancelled before their funding end date. This suggests that if a project can last through to its deadline, it has a decent chance of succeeding in the long run.
3. Film & video, music, and theater projects make up 70% of crowdfunding projects and journalism by contrast only comprises 0.4% of recorded projects insinuating that most money in crowdfunding is going primarily to entertainment over any other venture. Plays make up 34% of all crowdfunding projects and 33% of the successfully funded projects that were observed suggesting that most of the money that can be put towards funding is coming from backers who prioritize the theater and performing arts over other projects.
4. Campaigns with the most backers we also the ones that had some of the most affordable average donations with backer counts over 2000 especially having average donations of less than 50 suggesting that it is easier to fund a campaign with larger amounts of smaller donations rather than smaller amounts of larger donations.
5. What are some limitations of this dataset?

Not everyone who enjoys the media that is produced may have the money to back the project so the projects that are successful have more to do with the interests of the funders rather than the interests or needs of the population at large. Seeing as a majority of backed projects were in the parent category of theater and subcategory of plays, it shows that most backers are interested in plays rather than other categories such as documentaries, journalism, or technologies.

Another limitation is that the currencies are not unified therefore not all contributions or goals are equally balanced against each other. What may be affordable in one country may be unaffordable in another, which may limit international backing and success. This can also result in limiting the types of projects that are available internationally. This also ties into another limitation of there being differing amounts of subcategories depending on country of origin. This can skew the data especially for graphs such as “Outcome vs Subcategory for instance. Having many projects in a subcategory that is only represented in one country can skew the data when looking at all countries at once.

Campaign lengths were also another variable that were not unified which can affect the amount of time backers are exposed to and able to fund projects. While less realistic practically for the sake of development and implementation, each project having a similar amount of time to fund could allow some projects that may have failed to succeed instead.

There was also the matter of campaigns that were on the staff pick or spotlight lists. Campaigns that are promoted may have a higher likelihood of success based on accessibility and advertising. This can be better analyzed using another table as will be mentioned below.

1. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

A graph that measures average donation vs backer count can help better represent which donation amounts may be more affordable when developing donation tiers for campaigns.

Staff Pick vs success rate and spotlight vs success rate graphs can help display to developers how much visibility from promotion may help ensure that their campaigns are funded before the deadline.

A length of campaign vs success rate graph can help set pacing for development of future projects to ensure that there is enough time for backers to view and fund the projects before the deadline.

A graph showing average donation vs category would also be interesting to see to infer the variability of interests among socioeconomic levels. Being able to see which types of entertainment higher average donation backers prefer over lower ones or which category receives the highest average donation can help speculate priority level based on socioeconomic class.

Statistical Analysis:

The median better represents the data because of how high the variability is due to the outliers present. The mean is heavily affected by the outliers of the backer count leading to a right skewed distribution. Therefore, using the median would provide a more accurate middle value that is less affected by outliers.

There is more variability among successful campaigns than failed campaigns which matches the disproportionate number of backers for certain types of campaigns over others as shown in the graphs comparing outcomes vs categories.