Data Analytics Boot Camp, Rana Mousavi

Challenge 1

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

* The majority of projects and successful projects are centered in the arts. Plays, music, film.
* The success rate for projects is highest in the mid-tier funding areas. The low end below $10k and about $50k have the highest failure rates. Projects between $10k and $40k had the best chance of success.
* The rate of success, failure, and canceled projects is fairly steady throughout the year. This is impactful as there is no right time to launch a campaign.

1. What are some limitations of this dataset?

* With the different currencies, it difficult to see how the different projects compared without a conversion of funds to a common currency. If two projects show similar results but one is in US Dollars and the other is in Euros, they may not be so similar if they were both compared in US dollars.

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

* A comparison of Staff pick/spotlight vs. outcome would be interesting. How impactful is it to have a project picked to succeed have on the outcome of the crowdfunding effort? Does this extra attention bring success?

Bonus Statistical Analysis

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

* I believe the median is a more accurate measure of the data set. For successful projects, with a median of 201 and a mean of 851, there is a large difference between the two. With half the data falling below 201 but the average number of backers being over 850, there are a few large backed projects that push the average very high. This is also seen in the standard deviation of 1266. This puts 1 standard deviation below the mean at 0. You have to go above the mean to get to 2 or 3 standard deviations from the mean. Again, this is due to the extreme outliers above the mean.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

* The variability is higher on the success side then the failure side. This makes sense as a highly successful project can greatly push the number of backers greatly over the mean. A failed project on the other hand will not vary as greatly, because as it starts to push away from the mean it will more likely meet its goal and become a successful project. This place an upper bound on failed projects on how much it can vary away from its mean.