Week 1 Challenge

1. From our dataset we can conclude that 56.5% of all crowdfunding projects reached their target, 36.4% failed to do so, and 7.1% were either cancelled or still in progress. Based on the line chart generated to compare projects by month of the year, it is possible that crowdfunding projects are more likely to succeed during the summer months as we see a small spike in success rate between May and August approximately. Finally. We can conclude that a disproportionate number of crowdfunding projects can be categorized under “theater.” This is made even more obvious when the data is further broken down into sub-categories. In this case we see that “plays” represents 34.4% of total crowdfunding campaigns from our sample.
2. Limitations of this dataset include the sample size; without knowing how many total crowdfunding campaigns are out in the world it is impossible to know whether 1000 is a large enough sample to make strong conclusions from. Another possible limitation would be the diversity of the samples collected. For instance, if all these data points were collected from a similar geographic location, perhaps a city with a rich theater culture, one might expect there to be a disproportionate number of campaigns related to that category. This could potentially be avoided by sampling a variety of geographic locations and cultures.
3. In addition to the 3 graphs generated, we could compare the success rate by duration of the campaign to try to determine if campaigns that ran longer were more or less likely to be successful. We could also create a graph which relates the number of backers to success rate, along with average donation size to success rate to determine if having more backers, or larger donations on average is more likely to result in a successful campaign.

Bonus Statistical Analysis:

In the case of failed crowdfunding campaigns, the median gives a more meaningful summary of the entire dataset compared to the mean. This is because the data skews left due to there being a sizable amount of campaigns in which no donations were made at all. The opposite is true for successful campaigns. The data is skewed to the right due to a number of campaigns which exceeded their goals by a wide margin. The median ignores these outliers at both ends of the spectrum, which is why it gives us a more meaningful summary of the dataset.

The data shows that there is more variability in successful campaigns than those that failed. This makes sense when you think about how many of the successful campaigns surpassed their initial goals by varying margins. In contrast, many of the failed campaigns resulted in 0 donations, which brings down the overall variability of the dataset.