**Module 1 Challenge:**

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + Over half of the campaigns have succeeded. 565/1000 which is 56.5% of campaigns are successful. 364 campaigns have failed and only 57 campaigns have canceled. Comparing the ratio of successful campaigns to unsuccessful campaigns, we can see that crowdfunding is a viable method of raising funds and shows a viability of crowdfunding as a fundraising strategy for various projects.
  + July has the highest success rate. However, we are unable to deduce the reason for its success.
  + Similarly, August has the lowest success rate and highest failed rate. Although we are unable to determine the reason, the trend suggests that backers may have used more funds during the month of July and cannot provide funds for August campaigns. Again, this is just an opinion and we do not have concrete evidence to suggest this is the case.
* What are some limitations of this dataset?
  + No demographical information. We are unable to draw conclusions between creators and/or backers as well as any correlation with target audiences.
  + We are missing geographical information. Were these campaign launched in a specific region, continent, or even country. This could also include certain cities or even regional based campaigns. This could impact backers and creators as there may be some cultural backgrounds that is unaccounted for.
  + This dataset also lacks marketing strategies as well as external factors that could influence specific campaign success. Many modern day marketing strategies include hiring specific celebrities and internet influencers who have much more “influence” and can target a larger range of audience.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + Campaign Duration vs. Success rate: We can also provide a scatter plot that shows the relationship between campaign duration and success rate. This could help determine if shorter or longer campaigns tend to be more successful
  + Geographical Distribution of Campaigns: As we have stated in the limitations, we do not have geographical information but having that information could greatly highlight regions with higher or lower campaign activity
  + Backer engagement over time: A chart that could show the number of backers over the months. This could reveal trends in backer engagements and provide insight to indicate periods of increased interest

Use your data to determine whether the mean or the median better summarizes the data.

* Median would be a better to summarize the data. As there are many outliers, this could skew the mean. I created a box and whiskers plot to show that the outliers skew the mean. As the number of backers gets larger, it the mean will shift higher. There are some outliers that go over 6000 compared to most of the dataset between low hundreds to about 1400 for successful campaigns and 0 to under 1000 for failed campaigns.

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

* There is more variability in successful campaigns. We can determine this by the variance between both successful and failed campaigns. Variance in successful campaign is 1.6M where as failed is only 0.9M. The standard deviation for success is also larger than failed campaigns. This makes sense as failed campaigns tend to have lower number of backers. However, successful campaigns could have other factors that impact their success; smaller number of backers could have external factors that could increase their odds of success in their campaign such as marketing campaigns as well.