# Module 1 Challenge

## Dataset Analysis

### Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

* There were more successful campaigns than failed ones. There were more successful campaigns than failed campaigns in all categories other than the Food category.
* The theater category has the highest number of total campaigns, as well as the highest number of successful campaigns.
* Out of all the months, July has the highest overall number of successful campaigns.

### What are some limitations of this dataset?

* Outliers: The data has many outliers for the number of backers which could lead to skewed results with certain calculations.
* Sample size: Some categories have more campaigns, so the overall results are not necessarily the same as the results by category.

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

* A pivot table and line graph using years to sort the funding outcomes. This would provide us with a display of the campaign’s success each year, allowing us to determine if there was a specific year that was the most successful or least successful.
* Funding outcomes by country. Pivot table and bar chart.
* A pivot table and bar chart showing funding outcomes by country. This would allow us to see the number of campaigns in each country and if certain countries had more successful or unsuccessful campaigns than others.

## Statistical Analysis

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

* The median is better suited if a data set is significantly skewed because it is less affected by extreme values. The box and whisker plots created with this data show many outliers for both successful and unsuccessful outcomes. The mean for each data set is much higher than the median, which shows that it is more affected by these high outliers.

### 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 with the successful data campaigns. This makes sense because there is a larger range for the successful data set and more extreme values, all of which can lead to more variability.