**Module 1 Challenge**

1. **Written Report**
2. **Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

First, the rate of success of crowdfunding campaigns was not high, with only 57% of campaigns ending up being successful, 36% failed and 6% cancelled.

Second, the category of crowdfunding campaigns with the highest success rate was journalism (100%). Categories with above-average success rate included technology (67%), photography (62%), publishing (60%), music (57%) and film & video (57%). Those with below-average success rate included theater (54%), food (48%) and games (44%).

Third, the month with the highest success rate for crowdfunding campaigns was June (64%), July (62%) and September (62%), although the difference among the months of the year was not significant.

1. **What are some limitations of this dataset?**

First, the donation amounts of the crowdfunding campaigns are in seven different currencies. A column could have been added to convert the “goal”, “pledged” and “average donation” to one single currency to allow for a more objective comparison.

Second, the blurb of the crowdfunding campaigns is long and uniquely named, while the category and sub-category show only the industry that the campaigns belong to. A column could have been added to describe the nature of the product (e.g. database, internet solution) to allow for an overview of products and a comparison of products with other variables (e.g. outcome).

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

First, a table showing the outcome of crowdfunding campaigns (i.e. successful, failed, canceled) by the duration of the campaigns (i.e. date ended minus date created) could be created to show how long it takes for campaigns to succeed. The same information could be presented in a stacked bar diagram (x-axis: years, y-axis: outcome). This could inform potential backers the expected time for their investment to yield results, if successful.

Second, a table showing the outcome of crowdfunding campaigns by staff pick and spotlight could also be created to compare if those picked by staff or highlighted were those that were successful. The same information could be presented in a stacked bar diagram (x-axis: staff pick and spotlight, y-axis: outcome). If those picked by staff or highlighted tend to be successful, it could serve as useful reference for potential backers when they decide which campaigns to invest in.

1. **Statistical Analysis**
2. **Use your data to determine whether the mean or the median better summarizes the data.**

The median can better summarize the data. With huge variance and standard deviation, the mean is skewed.

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

There was more variability with successful campaigns as they have a higher variance (1,603,374) and standard deviation (1,266). This does not seem to make sense as the variance is much higher than the difference between the maximum (7,295) and minimum (16) values of the dataset.