**Shannon Hoffman**

**Module 1 Challenge**

**Crowdfunding:**

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

Based on subcategory: Plays had, by far, the most campaigns at 344. About 55% were successful, 38% failed, 7% failed, and less than 1% are live.

Based on outcomes: success rates increased across May and June then dramatically decreased in July for all Parent Categories.

Based on goals: The success and failure lines on the graph were mirror images, typically indicating a strong negative correlation (meaning that as the goal increases, the percentage of success decreases consistently).

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

* Data quality:
  + Titles: As an analyst, I don’t know what the data in the columns actually is (most evident is the “blurb” column).
  + Currency: The currency column includes different types of columns, making it difficult to make direct comparison (we could filter based on currency, but that limits the sample size.
  + Outcomes: “live” have end dates
  + Outcome: how was “failure,” “success,” etc. determined?
* Limited sample size: we’re only looking at 1000 projects (of the millions in existence).
* Data span: The collection dates cover nearly a decade; trends (depending on definition, and maybe even by definition) change over that period.

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

* We could look at a pivot table to compare the number of backers to how many campaigns were successful to analyze if there’s a connection between the two.
* We could look at a pivot table to count how many campaigns were successful, failed, canceled, or are currently live per currency. This has the potential to tell us more about how crowdfunding trends across countries.
* We could create a pivot table to count how many campaigns were successful, failed, canceled, or are currently live by staff pick and spotlight, with filters for year. This could tell us how much influence staff picks and spotlighting influences campaign success and failure.

**Statistical Analysis:**

1. **Does the mean or median better summarize the data?**

In both cases, the median better summarizes the data because the outliers in the set (determined by plotting on a Box and Whisker) skew the mean.

Successful: has 20 outliers

Unsuccessful: has 17 outliers

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

More variance with the successful campaigns. Yes, this makes sense because there are more outliers spread further across the data in the “successful” data set. Put another way, the data stretches further from the measures of central tendency.