Sarah Ryan’s Challenge 1 Statistical Analysis:

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

I believe the median is the best measure of central tendency to summarize this data. When examining the successful campaigns, I notice that the minimum value is 16, while the median is 201. Since the median represents the 50th percentile, it indicates that the first half of the data falls between 16 and 201.

If the data were symmetrical, we would expect the maximum to be around 400. However, the actual maximum is 7,295, which is a significant outlier. This extreme value heavily skews the mean, which is 851.1. Given the impact of outliers on the mean, the median provides a more accurate summary of the central tendency for this dataset.

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

To determine whether there is more variability in successful or unsuccessful campaigns, I first examined the range of both groups. The range of successful campaigns is 7,279, while the range of unsuccessful campaigns (failed and canceled) is 6,080. Both ranges are relatively large compared to their minimum values, which shows a significant spread in both datasets. However, to further assess variability, I look at the variance and standard deviation.

For successful campaigns, the variance is 1,603,373.7, and the standard deviation is 1,266.2. For unsuccessful campaigns, the variance is 843,230.8, and the standard deviation is 918.3. These values suggest that successful campaigns have a higher variance and standard deviation, meaning the data points for these campaigns are more spread out from the mean.

This makes sense when considering the nature of the campaigns. Successful campaigns likely experienced a wider range of outcomes, with some campaigns achieving very high backer counts or goals, which would create significant outliers and increase the spread of data. On the other hand, unsuccessful campaigns, which failed or were canceled, likely had more consistent but lower outcomes, with fewer extreme values. This would result in less variability, as the data points are clustered closer to the mean.

In conclusion, unsuccessful campaigns show less variability because their outcomes are more uniformly low.