Statistical analysis – summary statistics of crowdfunding

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

The median better summarizes the data because the data shows great dispersion, based on the vast range (difference between the max and minimum number of backers), the large variance and standard deviation, and the fact that in both successful and unsuccessful crowdfunding campaigns, the median is far lower than the mean. The average is certainly driven by the presence of outliers in the data, whereas the median can offer more practical insights for those who are involved with crowdfunding campaigns, as it is where the 50th percentile of the data is. I expect that the 75th percentile also be rather low compared to the mean number of backers.

There is less variability based on the lower standard deviation and variance numbers for crowdfunding campaigns that failed versus those that succeeded. This makes sense because there are overall a fewer number of backers, and the distribution of the number of backers for failed campaigns is more tightly distributed about its average value. If we had more information about how sophisticated or experienced these backers were, then we could potentially make the claim that backers can distinguish bad or fundamentally lacking campaigns, but not necessarily understand whether a campaign is worth backing at all.