Module 1

1st part of questions

* The first conclusion that can be drawn is that on average, there are more successful campaigns per month than failed and cancelled. The most popular sub-category by far is plays. Plays highly out surpasses the rest of the sub-categories by far. Theater was the most popular category, but it only had a 54.36 successful campaign rate, while technology had 66.67 successful rate, but it was the 4th most popular category. July saw the highest number of successful campaigns, while January had the most failures.
* Limitations of the dataset include not considering outliers if any are involved, why some pledging stayed at 0, how weather could affect pledging, taking account inflation or deflation during certain months, how holidays effect pledging, and age of people that are pledging/donating.
* A scatterplot could be created to look for how many outliers are involved. A scatterplot to compare two sets of data, such as average donations and backer’s account.

2nd part of questions

* The Median is a better measure of the for the data because it’s more in the middle of the data. When drawing a line graph of the data, the data is skewed in one direction. Also, when looking at the means, only around 4/10ths of the data go over the mean, whereas the median is in the very middle.
* There’s more variability with the successful campaign because the variance is greater than the failed campaign’s variance. The successful campaign also has a higher range of numbers.