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

* Parent categories theatre has the highest absolute number of campaigns followed by music
  + Within the theatre category, Plays has the highest absolute campaigns and across all sub-categories as well
  + Parent category journalism has the lowest number of absolute campaigns
* The most successful campaigns are in the music parent category
* A campaign on average is more likely to be successful then fail given it has not been cancelled

2. What are some limitations of this dataset?

* The data does not account for the popularity of Kickstarter over time, for example a campaign that is successful in the early years of Kickstarter say when they had 1K users would not be the same as a campaign that is successful when users have increased to 1M
* The cancelled campaigns can be understating the number of failed campaigns, we don’t have any data on why campaigns were cancelled, they could of have been cancelled due to lack of traction

3. What are some other possible tables and/or graphs that we could create?

* We can create a graph and table that shows the state of campaigns by duration of the campaigns this can help determine optimal duration of a campaign
* We can create a graph and table to show state of campaigns by goal to see if there is a relationship
* Also, we can graph by duration vs goal and filter by state to look for correlation between these variables

**Bonus**

Use your data to determine whether the mean or the median summarizes the data more meaningfully.

* Median is more meaningful in summarizing the data as the data is sewed, there are a few outliers that are inflating the mean, we can see that the max backers for successful campaign is 26,457 which is significantly greater then the mean of 194, thus we can concluded the mean is being inflated by outliers

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

* There is more variability in successful campaigns as standard deviation (SD)/mean and max/SD is greater in successful vs unsuccessful
* This makes sense as successful campaigns can have a lot more backers and exceed the goal by a wide margin vs failed campaign