Conclusions on Kickstarter Data

Question 1:

From the data we can conclude that certain categories are far more successful than others. Taking journalism as one situation, all proposed Journalism projects were cancelled indicating that this is not necessarily a popular Kickstarter choice. We can conclude that companies who are interested in seeking funding may wish to choose a different forum to propose/fund initiatives. We caveat this by noting that this is only a small number of projects (4,000 projects vs. 300,000+ projects).

Due to the small nature of the data set, we also conclude that there is a very weak correlation between the number of backers and the average donation. What this means is that the increasing number of backers does not result increased funding for donors. Though it is prudent to note that 7 out of the Top 10 high average donations (of successful projects) were US based which could lead to an interesting hypothesis that people are willing to donate more to US based companies.

Another interesting aspect of the data set is that games that were selected by on the Kickstarter Spotlight and were picked by the staff were always successful (486 cases). Furthermore, when you filter further, ~400 of these games were based out of the US. We can assume from this that Kickstarter prefers selecting US origin projects.

Question 2:

The current data set is quite limited afterall, it is only limited to 4,000 projects which is a small sum compared to the rest of the Kickstarter data set available. Furthermore, the data does not necessarily tell us whether the project has actually launched what was promised (currently we only if the project has been funded, cancelled or is in progress). Another limitation is that there projects seem to be all in different currencies. So doing a range evaluation is not necessarily useful if projects are in USD, CAD, Euro, etc. because they are carry different values. The data should be standardized to a single currency. We also need to consider the date of the dataset. Most of the dataset is 2017 and before. There are 2 years of data that is currently missing that is far more useful for us.

Question 3:

Another useful set of graphs would be a historical chart (histogram) that allows us track when donations peaked over project time. Pie charts would be useful in showing how much funding was raised from country, category and sub-category.

Bonus Question:

Question 1:

The median as the point of central tendency is more useful for data that is skewed towards or is being heavily impacted by outliers. While the mean can tell us the average number of backers there are also a number of projects that essentially failed and have no backers assigned and some that have very few backers but were successful.

Question 2:

Our data indicates that there is more variability with successful projects and this is reasonable. Donors are able to share any amount of money that they want. The variance for the failed project is smaller indicating that these projects tend to fail much faster and have smaller donor sets.