**Report**

Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

* There seems to be less projects in between the $10,000 and $49,999 goal. Between these 8 sets of ranges, there are only 85 campaigns. That’s 8.5% of the 1000 campaigns.
* Plays definitely get more attention than any other category. That does not mean that category is successful, while they have the most successes they also have the most fails.
* It seems like success, fails and canceled are correlated together rather than contrast. I expected whenever fails rise successes would fall but that is not the case in the chart. That means these product categories seem to follow a ratio of fails and successes that is consistent over the years.

What are some limitations of this dataset?

* One limitation that I saw was how sparce the middle ranges are that I mentioned in the first question. When I look at the stat breakdown chart, it looks like the $10,000 to $49,999 ranges are performing way better than the rest, almost all of the sections have a 0% fail/canceled rate. These ranges have such a small sample size though that the comparison between the two should be taken with a grain of salt.
* The data set is also using multiple currencies which means the average donations columns don’t accurately represent how much each organization raised relative to the others.

What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

* I’d like to see how long projects took to get funded. I’d compare the goal and check how long it took to fund. I would assume the higher the goal is the longer it would take to fund but I’m open to a surprise.
* I’d also love to see company performance after these projects were funded. Here I would also compare the percent funded and see if there’s correlations between the company’s performance and how much they raised proportionally. I assume the more they raise the better they perform but I also expect diminishing returns at some point.

**Statistical Analysis**

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

* Whether we use median or mean depends of the situation. Median should be used for non-normally distributed data and mean should be used for normally distributed data. It seems like the failed backers should be using median. Since the median and mean are so far apart it’s obvious that the data is skewed. The successful row seems to be normal though so both mean and median would be a good way to represent that data.

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 much more variability in the failures. This is shown by the large difference between the median and mean.