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

Based on the data, plays are by far the most popular sub-category to run a crowdfunding campaign for. They represented 34.4% of all campaigns from 2010 to 2020. The next closest sub-category was Rock music which comprised only 8.5% of all campaigns. This information leads me to conclude that crowdfunding campaigns are a significant financial component for theatrical companies to perform plays.

Most of the crowdfunding campaigns in this dataset were successful. 56.5% succeeded by meeting their stated goal, while 36.4% failed, and only 5.7% were canceled before the campaign finished. Based on this success rate, crowdfunding campaigns are a worthwhile form of fundraising as they succeeded 14.4% more often than they failed.

The vast majority of crowdfunding campaigns took place in the United States. 763 campaigns took place in the US, while only 237 campaigns occurred in the other six countries represented by the data. Crowdfunding campaigns were also slightly more successful in the US at a 57.1% success rate compared to 54.4% in the other six countries. This data leads me to conclude that the US is a desirable location for crowdfunding campaigns of this type.

* What are some limitations of this dataset?

The dataset has a limited sample size. The data is also limited to 2010 through 2020, leaving me wondering what the data prior to and after that period might show. The donations are only broken down by the average donation amount and number of backers, but there could be telling information if given other types of donation data.

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

We could include a bar graph showing how close each campaign came to reaching its goal at the time it ended. This could be filtered by parent category and/or sub-category to give us more insight as to how successful campaigns were by category.

We could also create a table and accompanying line graph with the date created conversion data to see trends of when campaigns are most likely to launch throughout each year (are more campaigns launched in the summer, less in the winter? etc.). We could also see trends of the number of campaigns launched year over year, to see if crowdfunding campaigns are becoming more or less popular from a long term point of view.

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

I believe the median better summarizes the data, because the data is heavily skewed by outliers. The mean is more influenced by these outliers, which makes the median more representative of the data as a whole.

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

The successful campaigns had more variability. This makes sense to me, because the unsuccessful campaigns had fewer backers, and thus the numbers were grouped more closely together. The successful campaigns had a lot more outliers on the high end, which created more variance.