Crowdfunding Analysis

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

1. Theater, Music and Film & Video are the categories that have the largest amount of crowdfunding projects. These three categories represent 69.7% of total projects and 68.6% of total successful projects.
2. Plays has the highest amount of crowdfunding projects at a subcategory level. With 34.4% of total projects. The US is the largest country to crowdfund plays with 273 total projects (or 79% of the total) with a 55% success rate.
3. For the majority of the years, 2010-2019, there were about 100 crowdfunding projects that were created. For years 2010-2019 there was a success rate of 56.6%.

What are some limitations of this dataset?

1. The data dramatically decreases in 2020, presumably from COVID. Category success based off of this data being pre-COVID could no longer be relevant given the changes that COVID brought about.
2. Different currencies are used in the data set, limiting the analysis we can do with the amount of money that was raised. We could always apply a conversion rate to make the money comparable, but it is a current limitation.

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

1. A stacked-column pivot chart that can be filtered by parent category and subcategory with country and outcome of campaigns as the variable. This chart can be used to help identify if a country is excelling in a category vs another category. It can help paint the picture if a certain category is already very saturated or has room for more opportunity.
2. A line chart with year and category as the variables, country and subcategory as additional filters. This chart could allow us to see if certain categories increased or decreased over years and allow us to draw conclusions about growing and exiting trends.

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

* In this instance the median better summarizes the data is skewed on the high side. There max values, interquartile range and high differences between mean and median are all telling me that some of the data set points are largely skewing this data set to run high in calculating the mean, so the median would be a better option to summarize the data set.

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

* Both successful and unsuccessful data sets have high variability. The successful data set has a slightly higher variability than the unsuccessful data set. This would make more sense because the mean, max and standard deviation are all higher in the successful data set. I gather that top crowdfunding ideas gain the attention of more backers, further creating greater variability in that data set.