Data Analysis Module 1

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

* Based on the data set, there were more successful crowdfunding campaigns than failed, cancelled, and live campaigns. Roughly half of all crowdfunding campaigns are successful, whereas the other half fall into the other categories (failed, cancelled, live).
* Each campaign type seems to have a month to month fluctuation of number of campaigns
* Parent Category Theater had the most crowdfunding campaigns
* Sub-Cateogory Plays had the most crowdfunding campaigns

What are some limitations of this dataset?

* The currencies are split across multiple types (USD, CAD, DKK, etc). It is hard to determine if the currencies impacted the monetary values in comparison.
* The Sub-Categories could be broken down further. For example, Sub-Category “Food Trucks” – are there specific food types that may have been more successful than others. Or Sub-Category “Television” – Does television aimed at children vs. adults see a higher success rate?

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

* A table to show how success rate of crowdfunding campaigns based on country – this might help to determine which
* A table to show the success rate of campaigns based on years – were some campaigns more successful during specific years, could there be outside factors that were impacting this during that year
* A table to show success rate based on the length that campaigns were live from. Do campaigns have a higher success rate if they are live longer?
* A table to show goal and pledged amount based on Parent Category and Sub-Category – Are certain Parent Categories or Sub-Categories start off with too high of a pledge amount, which could be impacting the success rate.
* A table to show the average donation per Parent-Category and Sub-Category – do some Parent-Categories and/or Sub-Categories receive a higher on-average donation per backer? Similarly, a table to show total amount of backers per category – are there some categories which see a high amount of backers vs. categories that show a low amount of backers. This could help to identify popularity trend in particular categories

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

* The mean better summarizes the 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 more variance with successful campaigns. This does make sense, and is likely due to that fact that there are many more backers of successful campaigns, therefore allowing for more variance in data.