Juan Moncaleano  
Data Analytics Bootcamp

**Report**

Three conclusions that can be drawn from the provided data regarding crowdfunding campaigns are:

* Theater campaigns had the highest success rate, while journalism campaigns had the lowest success rate.
* Campaigns for plays were 114 times more likely to be successful than campaigns for world music.
* The months of June and July had the highest success rates for campaigns, while December and January had the lowest success rates.

Some limitations of this dataset include:

* The popularity of the company/ business/ individual organizing the campaign is not accounted for.
* Demographic information about the backers is not included.
* The context in which the campaigns were launched is not provided.

Additional tables and/or graphs that could be created to extract more value from this dataset are:

* A comparison of campaign goals and pledges by category to determine if certain categories tend to have higher or lower amounts pledged compared to their goals.
* A breakdown of average campaign goals by currency to see if currency type affects the amount of money that is pledged.
* A chart showing the number of spotlights and staff picks by subcategory to identify which subcategories are more likely to receive these forms of promotion.

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

Based on the given data, it seems that the median would be a better measure of central tendency than the mean for both the "Successful" and "Failed" columns. This is because the mean is heavily influenced by outliers and extreme values, which are present in the data set (as seen by the large difference between the mean and median values). In contrast, the median represents the middle value of the data set and is not affected by outliers to the same extent.

For example, the maximum value in the "Successful" column is 7295, which is much higher than the other values in the column. This extreme value will pull the mean higher, making it less representative of the typical value in the data set. The same is true for the "Failed" column, where the maximum value of 6080 is much higher than the other values.

Therefore, in this case, using the median would provide a more accurate summary of the central tendency of 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?

Based on the given data, the variance for the "Successful" campaigns is 1,603,373.73 and the variance for the "Failed" campaigns is 921,574.68. This suggests that there is more variability in the "Successful" campaigns, as the variance is larger than that of the "Failed" campaigns.

However, it's important to note that variance is heavily influenced by outliers and extreme values, which are present in both data sets. Therefore, while the variance provides some information about the spread of the data, it may not be the most reliable measure of variability.