**Challenge 1 Homework: Crowdfunding**

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

1. Amongst the different categories, crowdfunding is particularly prevalent in Arts & Culture. Film & Video, Music, and Theater have the most recorded campaigns, with Theater leading.
2. The month the campaign is posted has little to no correlation in determining the outcome of the campaign. For example, posting in September vs posting in November likely won’t result in a significantly larger change in outcomes.
3. The goal amount has little to no correlation with the success rate. This means a campaign is not more or less successful based on the goal amount.

What are some limitations of this dataset?

* The number of campaigns per category is not a consistent number so it’s difficult to compare the outcomes of campaigns across different categories. For example, the conclusion that crowdfunding for Theater is much more successful than crowdfunding for Journalism is incorrect. Even though the stacked bar graph of “Outcomes by Category” shows a greater number of successes recorded for Theater campaigns, if we calculate the percent of successful outcomes, Theater has a 54.4% success rate whereas Journalism has a 100% success rate. This is an unfair evaluation because Theater has 344 campaign outcomes recorded, while Journalism has only 4.
* The data per year is not complete. When looking at the data over years, failed outcomes extend to 2020 but the number of successful and canceled outcomes stop at 2019. Without this data we cannot accurately determined trends from 2019-2020.

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

* Showcasing outcomes over time with years as a column rather than months, would be a helpful display of all the data over the years. This line graph tells us that crowdfunding platforms had the highest number of successes in 2017 and lowest number of success in 2012. Over time we also see that the number of successful campaigns remains greater than the number of failed campaigns, until the year 2020, where there are no records of successful campaigns.
* Calculating the length of time from the “Date Created” to “Date Ended” columns and comparing it to the outcomes could be an interesting data point because it would display correlation or lack of correlation between how long a campaign is and whether it is successful.
* Looking at “Spotlight” in comparison to outcomes could be useful in determining how effective of a feature it is on the platform.

**Statistical Analysis**

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

* The median better summarizes the data because the average is affected by outliers and large variability.

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 variability amongst successful campaigns. This makes sense because the range of backer counts is greater for successful campaigns and there is also a larger data set to compare against.