

## Module 1: Challenge Report

Henry White

Data Analytics Bootcamp

April 19, 2024

# Crowdfunding Campaign Data Analysis

The following explores insights on the data provided around crowdfunding campaigns from edX Bootcamps LLC.

## Conclusions Drawn from the Data

Three conclusions drawn from the analysis of the data are:

1. **Campaigns for plays are common.** The largest parent category of campaigns was theater, with all of the campaigns in this category being plays. Plays made up over a third of all campaigns in the dataset.
2. **Technology has a high success rate.** While it was not the most common campaign category, two-thirds of technology campaigns proved successful. The only category that had a higher success rate was journalism, but given that journalism campaigns account for less than one percent of the dataset, this could also be insufficient data.
3. **Summer is a good time to launch.** Campaigns started in June and July had some of the highest success rates, with a “crash” in August. It is not immediately clear why the data indicates this and further testing of this hypothesis should be explored.

## Limitations

There are a few limitations to the data available that should be considered. This data only spans through 2020, meaning it is missing context from the last 3-4 years. It is possible that a post-pandemic crowd funding scene would exist given shifts in participants experiences, demand from markets, and economic conditions. The data is also limited in explaining certain industries with fewer campaigns, such as journalism, as there is not sufficient data to determine if it is representative.

## Further Aggregations

With the data available, it would also be prudent to create pivot tables/charts aggregated by parent category and sub-category showing the average donation and the goal for each campaign. As further statistical analysis is performed, and campaigns are compared to one another, it would be helpful to benchmark whether these campaigns were comprised of large donations or if they had loftier targets. This could further help segment potential

donor bases and determine if any particular segments are oversaturated with too many campaigns.

## Shape of Data

The median is a much better measure of central tendency than the mean. The data has significant positive outliers and a right-skew. The median resists the pull of these outliers, which the mean does not.

There is more variability among successful campaigns than unsuccessful campaigns, signified by a higher variance and standard deviation, as well as larger difference in the mean and median. This makes a lot of sense, as there is not positive ceiling on how much a campaign can exceed its goal, but there is a floor of zero on all failed campaigns.