March 9th, 2023

report - Module 1 challenge

Boot Camp – Data Analysis & visualization

# Three conclusions, given this data, we can draw about crowdfunding campaigns:

## Disparity in the popularity of crowdfunding

### There is a significant disparity in the number of crowdfunding campaigns between ‘*film & video*’ (178 campaigns), ‘*music*’ (175 campaigns), ‘*theater*’ (344 campaigns) and ‘*food*’ (46), ‘*games*’ (48), ‘*journalism*’ (4), ‘*photography*’ (42), ‘*publishing*’ (67), ‘*technology*’ (96), but the success of the campaigns does not look significantly better in either of these two categories with an approximately 56% and 58% success rate, respectively.

## 76% of campaigns are USA-affiliated.

### It is difficult to accurately compare crowdfunding campaigns across countries, so a more useful analysis would silo the countries for a more in-depth analysis.

## Small or very large funding goals

### 86.31% of all crowdfunding projects had either a funding goal of 1,000 – 10,000 or 50,000+.

# What are some limitations of this dataset:

## Datasets are limited by their biases. Common statistical biases include ‘confirmation bias’, ‘selection bias’, ‘outlier bias’, ‘observer bias’, ‘funding bias’, ‘omitted variable bias’, and ‘survivorship bias’ according to [Intuit Mailchimp](https://mailchimp.com/resources/data-bias-causes-effects/#:~:text=There%20are%20several%20types%20of%20bias%20in%20statistics%2C%20including%20confirmation,they%20can%20affect%20your%20business.). In this dataset, there appears to be the following biases:

## The Selection Bias

### We are not sure which crowdfunding platforms were used to gather the data in this data set. The assignment referred to two, Kickstarter and Indiegogo, out of the 1,478 crowdfunding platforms in the United States, as of 2021, according to [Zippia](https://www.zippia.com/advice/crowdfunding-statistics/#:~:text=As%20of%202021%2C%20there%20are,%2C%20Indiegogo%2C%20and%20Crowd%20Supply.).

## The Omitted Variable Bias

### This sample dataset was artificially generated with a total of 20 variables/features to analyze. It is unlikely this artificial dataset and small selection of variables gives a legitimate dataset to draw practical conclusions from. An example of a useful missing variable is the timestamp of when each donation was given within the time period of the campaign. Another missing variable is standardizing the currency. There were several distinct currencies in the currency variable and standardizing them was not included in this activity.

## The Confirmation Bias

### This artificially created dataset was based on preconceived notions of a realistic distribution of the outcome of crowdfunding campaigns.

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

## A Scatterplot conveying popularity of funding goals and campaign success.

### A scatterplot showing the distribution of campaigns applying for different funding goals and how successful they were. This can give insight into which funding goals are most popular and successful, not just which percent are successful.

## A ribbon graph visualizing rank change.

### Ribbon graphs can show rank change over time for the crowdfunding campaigns, with the highest rank campaign shown at the top of the ribbon.

## Use Pie charts for conveying percentage data.

### Pie charts are useful for showing percentage of a whole. The percentage values in the Crowdfunding Goal Analysis would be better understood visually in a pie chart rather than a line graph.

# Does the mean or median better summarize the data:

## Median is better.

### Looking at the summary statistics, the median is a better statistic. The dataset is skewed by some large outliers.

# Is there more variability with successful or unsuccessful campaigns – does this make sense – why:

## There is more variability with successful campaigns.

### It makes sense that there is more variability with successful campaigns because there is a greater range between minimum and maximum pledged funds and a greater distance between the mean and maximum and minimum pledged funds than the unsuccessful campaigns.