Data Analytics Bootcamp

Lorelei Legg

Module One Questions:

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

A) Overall, the campaigns launched between June and July were more successful than other months, based on their launch date and outcome. Looking at the total outcomes of projects launched over all the years included in the data set, the line graph in the “outcome based on Launch dates” tab shows a positive trend in the success of a campaign between the months of June and July.

B) The theater parent category had the highest number of attempted crowdfunding campaigns, although they did have a relatively high number of canceled or failed crowdfunding campaigns. They made up 344/1000 campaigns in the data set, of the 344 the combined total of failed or canceled campaigns was 155/344 (45%). Therefore, despite the Theater campaigns being popular they have a relatively high failure and cancelation rate.

C) If the campaign was more than 100% funded, meaning the amount pledged exceeded the campaign’s goal, the campaign succeeded. Although it may not be the only contributing factor, receiving more funding than the campaign’s goal contributed to a campaigns success.

1. What are some limitations of this dataset?

The data does not represent a wide variety of countries. Most of the data collected represents the United States therefore there is a chance of bias.

The data may not accurately represent the current trends as the Crowdfunding data set does not include data from the past 4 years and only includes data for one month in December of 2020.

There is also the chance of incomplete or inaccurate data, whether it be in the collection method or missing values. Data may have been reported or recorded incorrectly.

It would additionally be beneficial to convert the currencies into one currency. Calculating the exchange rate would be beneficial for analysis, as the value of currency varies per country. For example 1 USD and 1 CAD do not have the same value and therefore different purchasing power.

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

A table that includes the amount of successful, canceled, and failed crowdfunding projects per country. The other columns would calculate the percentage that were successful, failed or were canceled per country. A bar chart would be a useful visualization and analysis tool to compare the results. This would show which countries out of the data set have the most projects, as well as which has seen the most success, failure or cancellations.

A table that calculates the average funding for each category with a bar chart that shows the comparison. This would be helpful to determine what categories are gaining the most funding.

Questions for Statistical analysis:

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

The median best summarizes this data because there are a high number of outliers in both campaigns which causes the mean to skew right. The mean is more affected by outliers and therefore it would be less representative 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?

There is more variance in successful campaigns. The high level of variance is logical as it shows that there are possibly other factors contributing to the success of a campaign other than the number of backers.