Samantha Seng

UCI-Data Analytics

7-20-22

Module 1 Challenge

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

The most successful crowdfunding campaigns were related to music, film/video, and theaters as noted on Sheet 1. Out of those plays were the most successful. However, despite the fact they had the largest number of crowdfunding campaigns overall, they had the highest amounts of successes and failures.

The summer months, specifically June and July, generated the most successful campaigns. This could be due to people having more money to spend during summertime related to children out of school.

Most number of campaigns started at the beginning of the calendar year in January and fiscal year, depending on company, in July. Some people may have been given raises and bonuses during this time and able to donate money to the crowdfunding. Some companies have budgets set aside to donate to a charity or fundraiser and can allocate funds to campaigns.

It’s very challenging to fund a campaign in a very short amount of time. Typically, to get enough funding to start a project there’s a lot of marketing and organizing. Investors want to look at numbers and the plan to create a project such as where the money is going, how will the plan work, etc.

**What are some limitations of this dataset?**

The dataset doesn’t tell the whole story such as the percent success and failure rate. An example would be plays, those had the most success, but almost as equal failures so their percent successes would be 54% whereas animation had 62% success rate. Another limitation would be sharing the demographics and who is donating the funds. The dataset shows how many people donated to a campaign, but not any information on their age, income, and location.

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

The chart on Sheet 3 only shows the data and trend over the past 10 years. To make it more relevant and pinpoint categories, it would be beneficial to use a bar chart. This allows the viewer to compare data within the last 3-5 years. There are many socioeconomic factors that affect crowdfunding especially after COVID and the recession.

An area chart would be good to compare different time frames of the year for the campaigns such as comparing the first quarter of year, January through March to last quarter of the year, October through December. This would allow people to see if there is more successful funding during different parts of the year. Some people might be more willing to donate to a campaign during summertime versus close to school year or holiday season when they need extra money.

Pie charts would be useful to visualize where which campaigns are the least and most successful. It would be a great chart to use for Sheet 1 and show the percentage of each campaign category.

Statistical Analysis Bonus Challenge

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

In this scenario, I believe the median would be the best tool to summarize the data because the mean is used to get ana average number it doesn’t account for outliers. The range with the min and maximum numbers is quite large. There might be more backers\_count for one successful or failed campaign and it can throw off the entire data. For example, for successful campaigns the min is 16 and the max is 7295, the max is an outlier, so the data becomes skewed.

**Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

There’s more variability with successful campaigns vs unsuccessful campaigns because each of the successful campaigns have different goals and backer\_count. Some campaigns succeeded with less backers whereas there were some that succeeded with more backers.