Brittany Chin

Data Visualization Bootcamp

Homework Assignment #1 Excel Challenge

September 5, 2020

Homework Assignment #1 Written Analysis

* Create a report in Microsoft Word and answer the following questions.

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
2. What are some limitations of this dataset?
3. What are some other possible tables and/or graphs that we could create?
4. **Based on the provided data, three conclusions that can be drawn about the Kickstarter Campaigns are:**
   1. The campaigns with the highest number of entries that were successful fell under the category of Theater, Music, and Film & Video. Further, the subcategory Plays was by far the most popular with 65% success. The most successful campaigns in terms of backing fell where the initial goal for the campaign was less than $1000 (71%).
   2. The likelihood of a campaign failing increases dramatically (19%) when the project’s initial goal is greater than $50,000. The number of successful campaigns with an initial goal of $50,000 only decreases by 2%. Contrarily, the number of canceled campaigns decreases significantly by 16%. Although successful campaigns only decreased by 2%, when a project’s initial goal is greater than $50,000, there is a much higher likelihood that the campaign will fail versus being canceled. The decrease in canceled projects may indicate that at this goal level, a project that is not on track may not be salvageable by cancellation, thus causing it to fail altogether. This may also indicate that project’s with higher initial goals may intimidate potential backers.
   3. The time of year campaigns are seeking funding seems to play a large part in the success of the campaign. The number of successful campaigns seems to peak around the month of May. However, after May, there is a steady decline in the number of successful campaigns until the month of September. Then, again, there is a sharp decline in the months of November and December where the number of failed campaigns surpasses the number of successful campaigns. This may be indicative of money being allocated elsewhere during the summer months (school summer vacation, summer travel, relocation, etc.) and during the holidays. When narrowed down by category, the trend further illustrates that no industry shows success in the month of December.
5. **What are some of the limitations of this data set?**

One of the obvious limitations is in the sample size of the some of the categories and subcategories. It is hard to say that the smaller categories would be relevant in statistical analysis. The solution might be to combine data from other crowdfunding sites or resources.

1. **What are some other possible tables and/or graphs that we could create?**

Although this is not a field provided in this data set, I would also like to analyze by State. There are certain categories that I believe would be more successful based on their geographic location. For Example, one might assume that Theater/Plays would be more successful in the northern hemisphere of the US, in a state like New York, but maybe less likely to be successful in a State such as Kansas.

We could also analyze the amount of time a project was given to be determined “successful” and which categories reached or surpassed their goals the fastest.

**Bonus Statistical Analysis**

In this data set, the median number of backers is more relevant to the data. The mean number of backers indicates on average how much support of the project existed. Using the mean in this instance would become an issue because the maximum number of backers on a project was 26,457 and the minimum was 1. This range would lead the average to become irrelevant in statistical analysis. The median is a better picture for the number of backers, splitting the data directly in half providing an unbiased center point of the data.

In this data set, there is more variability with successful campaigns. This means that the number of backers of successful campaigns varies greatly and that there may not be any correlation to the number of backers and the success of the campaign. Looking further, the standard deviation is quite large compared to the standard deviation of failed campaigns – pointing to the wide range in the number of backers in each campaign. The campaign could have been just as successful with one backer as it was with 100 or 1000. This makes sense because definition of “success” is not multifaceted, and the campaigns were not measured by different levels of success.