Scott Duncan

Homework 1: Excel

02/21/2020

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

There are a number of conclusions that we may draw from the data. One of the conclusions we can draw is that in the data pulled from Kickstarter, theatre projects were the most sought-after funding and also the second highest percentage of successful projects. Music Kickstarter accounts have the highest percentage of projects being funded. This information can instruct a potential fund starter to the general probability of success in a given field.

To further illustrate this, the subcategory of animation did not see a single successful campaign. This should issue a warning to future individuals hoping to fund their animation endeavor that they are in store for a potential failure. Additionally, rock music Kickstarter campaigns were fully funded during the selected time period of the data. Whereas, jazz music projects were all failed in getting to their funding goals.

While there are some outliers, there is a positive relationship between the number of backers and the success of a project. As backers continue to increase in a project the likelihood of a successful funding campaign increases.

Additionally, campaigns showed a decline in success between the months of May through September through all years where information was gathered. Further information as to the reason for the declined could not be ascertained from the data pulled.

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

Limitations occur when trying to understand why certain campaigns succeed for fail. For example, did the content on the campaigns website not fully implore potential backers to contribute funds toward the goal? This question as to why certain backers did not back certain campaigns would need further information to make a significant determination.

As stated earlier, narrowing down a reason as to why there was a decline in the middle months of the year is difficult to determine without additional information. Also, without additional clarification as to some of the column descriptions it is only through inference that some of the data sets can be defined, i.e. spotlight. There does seem to be a positive relationship between a campaign succeeding and having a “True” value in the spotlight field.

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

Other potential graphs that we could make that would tell the story of the data in regard to category and subcategory would be a pie chart. The advantage to this graph would show the comparison between each of the categories to the total number of funding campaigns. Also, scatter point graphs with trendlines could also show more information and illustrate the relationship between multiple data sets, i.e. backers vs percent funded.

Bonus Statistical Analysis

1. *User your data to determine whether the mean or the median summarizes the data more meaningfully.*

When discussing mean and median in regard to the backers, in this instance I believe that the mean shows a more accurate picture between the two. Given the large number of successful campaigns (2185) and where the maximum and minimum number of backers being 26457 and 1 respectively, the median shows 62 backers where the mean shows about 194. When taking the average amount of funds raised per campaign ($18,579) the resulting average pledge per backer comes to $95.77 for 194 backers and $290.30 for 62 backers on a campaign. These results when applied to a middle-income household averaging $75,000 per year, the $95.77 figure seems more likely than the approximately $300 price tag. Additionally, more campaigns would have more perks at the $100 range than the $300 range to entice more contributions.

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

When looking at variability in these data sets, there is a significant difference in the variability in the successful campaigns in relation to the failed. There is a great deal more variability in the successful campaigns when looking at the number of backers, 713,167 variance opposed to 3,776. This does make sense when looking at popularity of certain campaigns as a parallel to social media “viral” videos. Campaigns that obtain a viral media following can see their positive deviations spike while modest campaigns that may not need a large amount of backers to obtain their goal will draw the mean down/lower allowing those viral campaigns to increase the variance and standard deviation.