Jordan Roessle

UCI Data Analytics Bootcamp

Excel Homework: Kickstart My Chart

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

a) Looking at the graph for outcome per month, there does not appear to be a glaring difference in chances of success by the month of the year. Increases in successes appear to mirror an increase in campaigns created during that time.

b) Looking at the graph for outcome by main category, music appears to have a significant higher chance of success compared to other categories. Of course, we should run a test to see if the difference in percentage is significant; however, visually it does look significant.

c) Looking at the graph for outcome by sub-category, plays have more selections than any other sub-category by a large margin. Theatre is the most popular category, and more than 75% of Kickstarter campaigns labelled category have the sub-category plays. Other categories appear to have far more spread among their sub-categories.

2. What are some limitations of this dataset?

a) This dataset cannot account for people who set unreasonable goals for their campaign. For example, if you look at ID 125, the creator asked for $500 to go see a movie with his sister. He received $70 and cancelled his campaign. For most situations, $70 would be considered enough to go see a movie so this could easily be labelled a success had the creator not set his initial goal so high. This is only one example and it would be tedious to read through every Kickstarter campaign to determine whether the goal was set reasonably.

One other thing to mention is that the chart does not have cancellation dates. Thus, if a campaign decided to cancel soon after launching, it may not have been given the time to succeed.

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

a) I think one interesting table to look at would be outcome vs the difference in time from the launch and deadline/cancellation date. This would help show if longevity of a campaign has any impact or if the blurb itself holds more weight.

Bonus Analysis

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

a) I included the above box and whisper plots only to show how many outliers there are in each set of data. Thus, I would suggest using the median rather than the mean.

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

b) The standard deviation of backers in successful campaigns is about 844, while the standard deviation of backers in failed campaigns is about 61. Thus, there is more variability in successful campaigns.

This makes sense to me because large, successful campaigns will likely have a much larger number of backers than those campaigns that were large but not successful. If the campaign failed than not enough people backed. This makes those extremely large, successful campaigns have a far more extreme number of backers, causing further variance. As you can see from the maximums, the highest number of backers for a successful campaign is 26,457. Whereas, for a failed campaign the highest number of backers is only 1,293. It also appears like they is 10+ more successful campaigns with higher backers than the highest number of backers in a failed campaign.