Alyssa Ma

StarterBook HW Questions

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

   - From Kickstarter's start to 2017, we see the most Kickstarter campaigns started in 2015, 1225 campaigns to be exact.
   - Overall, the number of successful campaigns falls considerably in December, only accounting for about 5% of successful campaigns in total. December is also the only month where the total number of failed campaigns exceeds the number of successful campaigns.
   - Campaigns under the parent category "theater" make up the majority of total campaigns at 34%, around 60% of which end up successful. The most popular subcategory is "plays" which makes up 26% of total campaigns.

2. What are some limitations of this dataset?

The most recent data in this dataset is from 2017, so we would not be able to determine if there were any radical changes to Kickstarter campaigns during the past three years. Moreover, the data only tells us the outcome of a campaign but nothing about whether other promotional efforts were taken or how well-designed campaign pages were (design, photos, etc. aside from description) in promoting their campaigns. Also, we do not have any data regarding the demographic of campaign backers.

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

We can break down the average donation by category, look at the average campaign duration from launch to goal reach/deadline, campaign status by country, and the relationship between Kickstarter goal and campaign status.

Bonus Statistical Analysis

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

The mean and median values are 194 and 62 for successful campaigns and 18 and 4 for unsuccessful campaigns, respectively. To determine which value summarizes the data more meaningfully, I began by looking at histograms of both data sets. It is evident that the data is skewed to the right in both cases. To be more exact, I calculated the lower and upper bounds and determined that out of 2185 total successful entries, 244 of them are statistically considered to be outliers, while out of a total of 1530 unsuccessful entries, 192 of them are outliers. With the large number of outliers in both groups skewing the mean, I determined that the median would summarize the data more meaningfully.

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

Successful campaigns have a standard deviation of 844 while unsuccessful campaigns have a standard deviation of 61. Based on these calculations, successful campaigns have evidently more variability. The wide difference in backer count standard deviations makes sense as successful campaigns have backer

counts that range from 1 to 26457 with an average of 194, while unsuccessful campaigns have backer counts that range from 0 to 1293 with an average of 18. In other words, as unsuccessful campaigns are ones that were not able to reach their goals in time, they have overall lower backer counts in comparison to successful campaigns. To elaborate, a campaign's success is determined based on its individual monetary goal, so the backer count can vary greatly from campaign to campaign. On the other hand, an unsuccessful campaign can receive no backing whatsoever, and in this case, 287 out of 1530 unsuccessful campaigns fall under this category. Thus, it makes sense to see more variability with successful campaigns in this data set.