Min Lee

UCI Data Analytics HW1 – Kickstarter Success Analysis

1. From the data given about the Kickstarter campaigns, we can draw the following three conclusions:
   1. First, with the inclusion of every data entry, the number of successful campaigns that reached their target goal was consistently higher than the number of failed and canceled campaigns, except for in December, when the number of failed campaigns was slightly higher. The number of canceled campaigns was always lower than both the number of successful and failed campaigns, meaning campaigns were most likely to reach their goal. If they didn’t, they were next most likely to fail to reach their goal. The least likely event would be for the campaign to be cancelled.
   2. Second, looking at sub-categories, we can see that campaigns in specific categories have higher chances of succeeding or failing than campaigns in others. For example, within the technology Parent category, we see a pretty even spread of campaigns that have met their goals, failed to meet their goals, or were canceled. Of the sub-categories within the Parent categories, there is definitely a bigger skew towards one result than the others. For example, within the technology Parent category, all 20 Kickstarter campaigns within the gadgets sub-category failed, while all 140 campaigns within the hardware sub-category succeeded in meeting their goal. These are extremes but generally the same is true for most of the other sub-categories.
   3. The number of Kickstarter campaigns that were recorded in this data set ranged from the years 2009 to 2017. The number of campaigns between these years show somewhat of a normal distribution with the peak number of campaigns in 2015 at 1255 campaigns. Since then, the number of campaigns has fallen to 948 in 2016 and 222 in 2017, the last recorded year. This shows that the number of Kickstarter campaigns in general has decreased dramatically.
2. One limitation of this dataset is that a large majority of the campaigns are from the United States. While that necessarily isn’t a glaringly big limitation at first glance, it certainly limits the scope of the dataset and analysis to one geographical location. Another limitation the data set has is in its lack of transparency with the length of each campaign. Although we do know the start and end date of each campaign, we don’t have a general idea of the length of each campaign, which would help us create a better story that explains the outcome of each campaign. This would help us answer questions like, “Were campaigns shorter in a certain sub-category to reflect a higher rate of success?”
3. One possible table that we could create that would help create a story about the successes and failures of the campaigns is a table with the methods of central tendency for each sub-category. We could then measure which sub-category produces the most successful campaigns and which ones produce the most failed campaigns using the mode. Another graph that would be helpful would be one that shows the average donation amount by month to see if there’s a relationship between seasonal changes and the amount of support given on Kickstarter.