Homework 1

Jill Peloquin

We were asked to analyze Kickstarter data and the following are the results:

1. The data from the sample of 4,000 Kickstarter Campaigns have roughly a 53 % chance of success, with 2185 out of 4114 being successful efforts. When the number of the campaigns based on the state of the campaign (Successful, Live, Failed, Canceled) was compared to the Primary Category of the type of campaign, it was found that the highest total number of successful campaigns were related to theater, 839. By the same token, the most failed campaigns of the dataset were also related to theater, 493. The most cancelled campaigns tended to be related to technology, 178. Music is the second most successful subcategory with fewer failed and cancelled campaigns.
2. When the data was sliced further into respective subcategories, it’s clear that Kickstarter campaigns for plays are not only the greatest contributor, but they are also, consistently the most successful. There were several sub-categories that had 100% success, such as classical music, documentary, electronic music, and hardware, among others.
3. There appears to be a temporal trend with when Kickstarter campaigns are most successful. Campaigns are more successful than failed in January with peak success in the months just before the summer. Success begins to decline as the summer progresses with the number of failed and successful campaigns becoming roughly equal in December. While there are variations in success through the year, there doesn’t seem to be a strong monthly trend for cancelled campaigns or failed campaigns.

Limitations:

1. The Data that was analyzed is from one dataset of 4,000 outcomes, not the entire body of evidence. Without knowing how these data were selected and if they were truly random, overarching conclusions may not be able to be drawn. These data go only up 2017, so any trends that have appeared in recent years wouldn’t be present in this analysis.

Other tables/graphs to analyze would be around stacking percentages so the user can understand and compare the likelihood of success of a category. Several categories have more than 50% chance of success, such as theater, music, and film & video.

Another analysis that would be interesting is how successful each campaign was based on the met goal of the campaign, rather than just looking a particular overall state. This would also give us more information on campaigns that were injected with well over 100% of the ask. This information could be interesting in entrepreneurs looking to invest or for campaign initiators with several ideas, for example. This next figure illustrates that two subcategories, hardware and tabletop games seem to exhibit large over investment compared to other sectors.

Bonus questions for Bonus2:

1. Usually when data has a large amount of variability, which these data seem to have, it makes sense to use the median value, which is less influenced by outliers. In this case, you can see the mean plots in the upper quartiles; indicating it’s being influenced by those values, making the median more meaningful.
2. There appears to be more variance from successful campaigns, I think this makes sense without slicing the data more, the number of bakers could vary widely since it’s not a normalized value to the original ask/Kickstarter request.