UCI Data Bootcamp

HW 1 Excel Write-Up

Vishal Patel

1. One conclusion that can be made with this Excel file, based on the second pivot chart created, is that *plays* on Kickstarter are wildly successful and are by far more successful than any other subcategory. For example, based on the pivot table created in Part 2 of the homework, we found that there were 694 successful from the Kickstarter Data, while the next successful subcategory was *rock* with 260, which is an enormous difference of 434. This disparity is also seen in the bar chart created in the first part of the homework. Another conclusion is that journalism Kickstarters are not successful at all. This is seen clearly in the first Pivot Table/Chart we made, where there were no reported successful or even failed journalism kick starters. There were only 24 recorded cancelled. If we filter the pivot table created in Part 2 of the homework, filtering the chart by Category “journalism”, we see that the only subcategory of journalism recorded is audio, with the 24 recorded cancellations. There are a variety of reasons why journalism doesn’t perform well on Kickstarter and always gets cancelled, but in terms of the dataset provided, there is no doubt that it is the most cancelled and least successful category. Finally, a third conclusion can be made about third chart we made. This conclusion is that it seems Kickstarters peak at being successful at around the month of May. In fact, this peak is very noticeable in the chart create, as it creates a sharp point in the graph. There could be a variety of reasons for this, including that this might imply that people create Kickstarters that last half a year more likely than any other time frame, but the only conclusion we can make here is that successful backings peak in May.
2. One limitation of this dataset is the size of it. This is because we know Kickstarter is such a large site, and there is doubt that around 4000 items cover all the ideas posted on the site. Looking at Kickstarter.com’s stats page, they state that there are 183,093 total successful backings on the site (<https://www.kickstarter.com/help/stats>). This results in a less accurate output considering we only analyzed about 2% at the very most, as this statistic does not even count those that have failed or been cancelled. Observing such a small percentage of the total backings could give us a narrow look at the whole picture, and mislead us. Another limitation of the dataset is the potentially small number of sub-categories. For example, when talking about journalism in the previous question, I mentioned how audio is the only sub-category available to analyze, which made me say that journalism is the least successful category. But if there is only audio to analyze, does that conclusion hold weight?
3. Some other possible graphs we could see is with a pie chart for categories, making it easier to see which categories made up more of the site compared to others. A possible pivot table could be one that has rows represent a category, and each column represent each sub-category and its statistics. This could be useful if we want to see what specifically makes up Kickstarter, and how many sub-categories make up the majority of the larger category. Another chart is also one that sees a box and whiskers for successful projects for a category. We might be able to see outliers of a successful category that could skew the data. For example, if a category has only 1 sub-category or 1 listing, and that failed, that shouldn’t affect a percentage of total failed that much.