Stefan Sampaleanu

NU Data Science Boot Camp – MWS

2/3/2018

Excel Homework Questions

1. **What are three conclusions we can make about Kickstarter campaigns given the provided data?**
   1. The three most successful categories of projects that were observed were film & video (300/520 = 57.7% success rate), music (540/700 = 77.1%) and theater (839/1393 = 60.2%).
   2. Certain sub-categories speak to the success of their parent categories. Documentaries were 100% successful (180/180) and were all in the film & video category. Hardware is a strong sub-category of the technology category, with 140 of 140 projects successful. In the music category, rock really shined as a sub-category with 260 of 260 projects successful.
   3. Although successful projects formed the largest share of Kickstarters in the set (2185/4114 projects = 53.1% of total), this does not mean that a project had a consistently higher likelihood of success than failure. In July 2014, for instance, cancellations and failures spiked, with failures affecting the largest share of projects (82 failures vs. 59 successes).
2. **What are some limitations of this dataset?**
   1. Misinterpretation: in this context, “success” simply refers to the Kickstarter’s goal being met. It says nothing about the project’s likelihood of success long-term, or what factors are related to that. Users must realize that this dataset speaks more to behavior around donation than project profitability.
   2. The spread of data changes over time, and the ratios exhibited in the aggregated “Outcome x Month” chart are not representative of a single year. Early years in the study show a much higher proportion of successes. In later years, we see the number of successes dip much lower, relative to number of failures or cancellations.
   3. A quick pivot table shows that project countries are spread with varying density. The US made for 73.8% of projects, so we can’t gather too much on the success of startups from country to country. Some nations had less than 5 projects to their name in the dataset.
3. **What are some other possible tables/graphs that we could create?**
   1. There are a few combinations that were not referred to in the assignment. I would be curious as to the relationship between average donation and project success. Another idea would be to chart project outcome versus length of startup, with number of days along the x axis. I would be curious whether the successful projects are meet their goal more quickly than failed projects.