Unit 1 Homework: Kickstart My Chart

Haley Huhtala

3/11/2020

Background:

Projects can receive crowdfunding through Kickstarter if they meet or exceed the initial funding goal they set for themselves. The 4,000 projects started between 5/17/2009 and 3/15/2017 were analyzed to determine if project category, sub-category, start date, or goal (in USD) could help predict which projects would be successfully funded.

Results and Discussion:

There were a range of number of project submissions per category and sub-category and the status of the projects also varied across the categorization. The most prevalent statuses were successfully funded and failed to receive funding with live, as in still ongoing at the time of data collection and cancelled making up a minority of projects. Of the nine categories Theater, Music and Technology have the most projects submitted (Figure 1). Journalism has only 24 projects compared to the others that have more than 200.

**Figure 1:**  The number of projects per category is represented by the bar height and the data label above the bars. Additionally, each bar is divided by the status of the projects and color coded according to the legend.

Drilling down in the 41 sub-categories, plays are the most prevalent projects followed by rock music and wearable technology which are each less than a quarter of the play submissions (Figure 2).

**Figure 2:** The sub-categories of the top three categories are shown to preserve readability. The number of projects per category is represented by the bar height and the status is again color coded based on the legend.

The overall success rate of the data set was 53%. Theater, music, and film and video were the only categories that had a better success rate. (Figure 3).

**Figure 3:** The bars are color coded based on the percent of each project in the different statuses as depicted by the colors in the legend.

Now looking at the data in terms of the US dollar value of their goal, the general trend is as the project goal gets higher the project is less likely to be successful. Projects with goals less than $5,00 were more likely than the overall average to be successful. It is also worth noting the projects with a goal between $5,00 and $10,000 had an average success rate of 52%, right below the overall average of 53%. Projects with budgets between $40,000 and $45,000 interestingly had an average success rate of 50%, this is higher than a linear relationship between goal and percent status would predict and visually appears to be an inflection point because the average success rate of projects with a goal of $45,000 to $50,000 was 28%.

**Figure 4:** The bars are color coded based on the percent of each project in the different statuses as depicted by the colors in the legend.

Limitations:

The sample selection of this data set limits this analysis. Many choices had to be made to collect this data including date range, number of projects, selection method (random or other), categories to include, and others. One could reasonably question is this data relevant to projects in 2020 and is this set representative of the all the projects? There are a few methods used in this analysis that present limitations as well. The currency was converted using 3/9/2020 conversion rates, not the rate at the time the project was raising money which could lead to project goals being grouped differently. Additionally, statistical significance testing was not performed on the conclusions so the results are possibly due to statistical variation not underlying trends.

Future Analysis:

Future analysis could include further investigating the successful projects in the $45,000 to $50,000 goal range to try to determine why they are on average more successful than projects in the adjacent ranges. Looking at the category and sub-category breakdown of that range could offer additional insights. This total analysis could also be run again on more recent data to see if the trends observed here still hold.

Question for the TA: I’m a Chemist so I was taught to write reports in a passive voice. Is the same convention used in Data Analytics? Do you think it leads to my writing sounding less compelling? Thanks.