Module 1 Report

Kirill Zavalin

1. All categories have a comparable distribution of failed and successfully funded projects, though proper statistical analysis would have to be done to say this with certainty. Certain categories have a much greater volume of projects, but also have a greater number of successfully funded projects, so one should not be deterred by running in a category with many other participants. If one looks at funding over time, June and July seem to have the greatest number of projects funded within a year. June in particular is the month with the highest percentage of funded projects. However, if one looks year by year, this trend is not always cleanly observed. Finally, projects that have particularly high goal amounts are less likely to be funded.

2. This data set has limited entries in certain categories, particularly journalism, having very few entries, making it difficult to gauge success in crowdfunding projects in those specific categories. Subcategories suffer from this even more, or analysis of different categories by select countries. Another limitation is that the listed currency is different for different countries, though this could be fixed by converting all to USD. However, without doing this, the separation of funding campaigns by goal amounts is not entirely appropriate, since these amounts are not normalized to a standard. Another limitation is not knowing the specific amounts pledged by backers. Though some information can be deduced from calculating an average contribution, this can not truly be determined without having the data. This could be important in calculating the importance of strategies that attract few high-paying backers vs many moderately-paying backers. Another limitation is not knowing the timeline of project funding for each specific project. For instance, accumulating a critical mass of backing can be important in attracting further backers as they perceive that the project has potential for success. Furthermore, many limitations depend on specific research questions in mind.

3. Standardizing currency would be a useful addition to the existing table. A line graph showing total and successfully funded project over years would be useful to see long-term developments in crowdfunding. This could be filtered by category and country for additional detail. Another useful table and graph could be used to look at greatly successful projects, and to see correlations between number of overpledged projects (over 200% funded) and categories, countries, and goal amount. This could suggest the criteria for having high funding success. This would be a pivot table where categories, countries, or goal amount (for each respective comparison) would be placed as rows. Another useful graph would be a line graph that graphs all failed projects with percent funded on x-axis, and frequency on y-axis. This would demonstrate the distribution of funding within failed projects, and one could analyze whether it is normal. This could hint at whether projects that come close to being funded end up gaining momentum to be funded, and whether many projects that receive little funding stall out in this spot. This data could be augmented if one knew the funding these projects received during the initial period of fundraising, and if this initial momentum of success or failure affected whether the projects had eventual success or failure. This information, unfortunately, is not available, and is of the limitations of this dataset.

Bonus Statistical Analysis Answers

Median summarized the data much better than the mean. There are clearly outlier projects with an exceptionally high number of backers that skew the mean towards large values. This is represented by the high variance/standard deviation, and by the large difference between mean and median values.

More variability is observed in successful campaigns, which makes sense, because this category selects for projects that gained full funding or funding that greatly exceeds the goal amount. However, the failed projects category locks in projects that had a limited number of backers (which was the cause for having failed), and thus selects out a more restricted range of backers within the projects.