Berkeley Data Analytics – Fall 2017

Homework #1 – due 11-11-17

Submitted by Angela Detweiler

Report-

1. **What are three conclusions we can make about Kickstarter campaigns given the provided data?**

Based on the data provided we can conclude that between 2009 and the beginning of 2017:

The top 3 categories for number of launched projects were theater, music, and film & video. Music and film & video had over twice as many successful projects funded over failed projects. Amongst the subcategories in theater, plays had the highest number of launched projects (1066) with over 60% being successful. The food category has the lowest success rate (<20%), mostly attributed to the failure of food truck projects being funded (see category pivot and subcategory pivot tabs on spreadsheet “HW1\_Detweiler).

Between 2009 and 2013, the number of successful projects was 3 to 4 times greater than the number of failed projects. Kickstarter project launches increased dramatically after 2014 (977), peaked in 2015 (1225) and decreased in 2016 (949). This trend was observed for successful, failed and canceled projects. The number of failed and successful projects during this period were about the same, whereas canceled projects represented about ¼ of successful projects. The dataset for 2017 is incomplete, therefore we cannot conclude if Kickstarter projects will continue to decrease, or turn around and increase by the end of 2017. There is no correlation between month of year launched and rate of success (See Date Pivot tab).

Projects with funding goals lower than 15,000 had a 50 to 70% success rate. Projects with funding goals between 15,000 and 50,000 hovered at a 50% chance of success, with a sharp decrease in success rate observed for goals set higher than 50,000 (see Bonus Table tab).

1. **What are some of the limitations of this dataset?**

The dataset that we analyzed containing 4114 projects, represents only 1.1% of the total number of projects (379,211) launched to date on Kickstarter1, therefore we should limit the interpretation of our analyses to this dataset, instead of extrapolating it to Kickstarter projects as a whole. I noticed for example that our dataset only includes 9 out of 15 categories, and 41 out of 169 subcategories available on Kickstarter.

Another limitation of this dataset is that the currency varies for different countries, therefore it should be standardized to one currency, so that we can make accurate comparisons of funding contributions across categories/subcategories, etc. Since most projects in this dataset originated in the United States (3038 out of 4114), I think that it is justifiable to convert all non-US currencies to US dollars.

Lastly, several Kickstarter projects offer incentives for contributors. I think that this is a very important factor to take into consideration as to whether a project is successful in being funded or not, therefore it would be useful to have another column called “incentives”, with a “yes” or “no” entry for each project.

1https://www.kickstarter.com/help/stats?ref=about\_subnav

3. **What are some other possible tables/graphs that we could create?**

-We could assess if there is any correlation between success/failure and campaign length, by plotting the project status count versus the number of campaign days. The latter would require subtracting date\_created\_conversion values from the date\_ended\_conversion values.

-It would be interesting to assess which categories/subcategories received percent funding up to 200%, 500%, 1000% or greater. This would require plotting categories/subcategories vs. the different ranges of percent\_funded for the successful projects.

-Plot the average donation per category/subcategory, to evaluate individual’s financial commitment to the different causes.