Jason Ree

Cohort 6 – Georgia Tech Data Science Boot Camp

6/15/2019

**Excel HW: Questions**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

* Conclusion 1:

From the perspective of parent categories, the nine categories can broadly be divided into categories related to technology and arts (it can arguably be assumed that food can be an art). Three of the five largest categories, theater, music, and film/video pertain to the arts and account for 63.5% of all campaigns with 60.2%, 77.1%, and 57.7% successful campaigns, respectively. This indirectly displays the shortage of funding for the arts, and presents the utilization of crowd funding platforms for the purposes of a) demonstrating creative and prospective ideas to the public, b) obtaining an early fan-base and crowd support, and c) obtaining the funding required for such art campaigns to begin production.

The third and sixth largest categories are technology and games, where typically large amounts of resources are required for production. They also benefit from crowd funding platforms to see the potential market demands of their future products. An intriguing statistic of these two categories is that their success rates are much lower than the top art-related categories at 34.8% (technology) and 36.4% (games). Such outcomes may be correlated to technology and games requiring larger initial and sunk costs to begin development and/or production, which in turn may be harder to reach their funding goals.

Further analysis and visualizations into categories’ size of funding goals in respect to their funding success rates may be required to provide a more in depth visualization of the provided data. However, within the scope of the exercise, a new campaign related to the arts can likely expect a successful outcome, while the chances may be significantly lower for technology or game related campaigns. Also, the smaller categories (food and journalism) have low success rates, possibly due to being unpopular with Kickstarter backers and enthusiasts; thus, suggesting that starting a campaign in such categories may not be advised.

* Conclusion 2:

After dividing the data into sub-categories, the first notable observation was that twelve sub-categories (classical music, documentary, electronic music, hardware, metal, nonfiction, pop, radio/podcast, rock, shorts, tabletop games, and television) had 100% success rates of reaching their goals; followed by five sub-categories (indie rock, small batch, space exploration, plays, and photobooks) that had percentages above the total success rate percentage of 53.1%. Intriguingly, the sub-category with the largest number of campaigns is plays accounting for 25.1% of all campaigns, also with an astonishing 65.1% success rate. Such sub-categories may have low funding goals which may be easier to achieve.

Similarly, it was very interesting to find that twenty sub-categories (animation, art books, audio, children’s books, drama, faith, fiction, food trucks, gadgets, jazz, mobile games, nature, people, places, restaurants, sci-fi, translations, video games, web, world music) had 0% success rates. Again further analysis into goal size may provide more insight as to why such sub-categories have such a low success rate.

Through filtering the sub-categories by their parent categories, it could be observed that, in general, certain sub-categories succeed dominantly within their parent category. In other words, it is interesting to observe that the sub-categories either 100% succeed or 100% fail. For example, for the food category, the small batch sub-category accounts for all successes at a 100% rate within the parent category (food trucks and restaurants have 100% failure rates). Likewise, among the six sub-categories within the film/video parent category, documentary, shorts, and television have 100% success, while the remaining three (animation, drama, and sci-fi) have 100% failure or cancel rates. It could be very beneficial to refer to how past campaigns of the same sub-category and parent category performed when launching a new crowd funding campaign.

* Conclusion 3:

When observing the campaign outcomes based on months, the total campaigns launched per month are pretty consistent at around 300 campaigns. Campaigns launched in May (233 successful campaigns) have the highest success rate and those launched in December (111 successful campaigns) have the lowest. Overall observing the time series trend it seems that campaigns may have a higher chance of success if launched in January to June than the latter half of the year. However, based on the linear regression line and its R-squared value (0.32) it is difficult to state that the date in which a campaign is launched is correlated to its success.

1. What are some limitations of this dataset?

Some limitations of the dataset are the following:

1. State (successful, failed, canceled) of the campaigns: the goals may be astronomically large or small to the point where the success rates may suffer or benefit due to either being too difficult or easy to achieve, respectively. In other words, it is a bit difficult to determine whether setting an unachievable funding goal was a major reason for the campaign to fail or whether the idea was unable to draw the support of the Kickstarter backers.
2. Limited to quantitative analysis (for this exercise): perhaps the data can be evaluated further using NLP or text mining techniques to assess the quality or content of the campaigns for qualitative analysis. This may allow discovering unknown characteristics present in successful campaigns when compared to failed or canceled campaigns.
3. What are some other possible tables and/or graphs that we could create?

One interesting relationship I would like to see is whether the size of the campaign goal has an impact on the success rate or state of the campaign (i.e. whether a campaign that sets its goals too high or too low, and how effective this goal is in affecting the success or failure of the campaign). Adding tables that further divide the funding goals filtered by the sub-categories would be the next step to take for further observations. Other types of analyses I would like to perform is qualitative analysis utilizing NLP and text mining techniques to see if there are common contextual factors among successful campaigns. In addition, I would do extensive analyses on the play sub-category to see whether this popular (most campaigns and high success rates) has notable traits that could explain for its high success rates.