Analysis of Kickstarter Campaign Data

Data Science and Analytics Boot Camp – Homework 1

Prepared By: Cam Foster

Date: 15 September 2020

**Analysis Conclusions**

1. Chance of success appears to be dependent on Goal Value (GV). Low GV campaigns were successful 71% of the time, while campaigns with GV greater than $50,000 succeeded only 19% of the time. The range of $5,000-45,000 remained relatively constant with approximately 40-50% succeeding.
2. The Theater, Music, and Film & Video categories had both the highest count of campaigns as well as the highest percentage successful at approximately 64%. Conversely, Technology campaigns had a low percentage successful at approximately 35%, with a significant percentage more than other categories being cancelled.
3. Spring and Summer months historically have had more successful campaigns started than Fall or Winter months.

Based on this data, if the goal was to get a campaign funded you would ideally initiate a Kickstarter campaign in the Spring or Summer months in an artistic category (i.e. Theater, Music, Art, Film/Video) and ask for less than $5000.

Unsurprisingly, most campaigns are from the United States. 3038 out of a total 4114 (73.8%), spanning all major categories, were initiated within the United States. This is not highly informative since the company is based in Brooklyn, New York, but is included for completeness.

**Limitations of this Dataset**

A sample of 4,000 out of 300,000+ (445,000+ as of 2020 according to Wikipedia) Kickstarter campaigns is relatively small. Additionally, without knowing what the sampling criteria or methodology was when obtaining this data, it is impossible to know if any conclusions drawn are indicative of an overall trend or are an artifact of the sampling. For example, when looking at campaign count versus year started, there is a large spike between 2014 and 2016. It is plausible that this is due to a growing popularity around that period. Kickstarter was started in 2009, so it makes logical sense that the first few years would have a lower volume of campaigns. However, there is a sharp decrease in campaign count in 2017 which implies that this dataset was either created sometime during 2017 or did not collect data past that time.

**Additional Tables, Graphs, and Data Visualization Aides**

Histograms and Box-and-Whisker plots would both be valuable for this analysis. The histograms would be a good visualization of backer-count distribution, and the box-and-whisker plots would show the variance of backer count, GV, or average backer donation at the Category or Sub-Category level.

**Statistical Analysis**

Median better summarizes the data in a meaningful way. Looking at the histograms on tab ‘Bonus Statistical Data Charts’, most campaigns had backers near the median for each data subset. The charts were all set to have 10 bins, with the overflow bin starting at approximately half of that subset’s standard deviation.

There is significantly more variance in successful campaigns. This can reasonably be assumed to make sense. There are some very high GV successful campaigns that have a large number of backers, which contrasts with the majority of campaigns that have low backer count and GV. In general, Failed and Cancelled campaigns do not have a lot of these very high backer count campaigns. If they did, they would likely not have been cancelled or have failed.