Gigi Jones

G2gomez@hotmail.com

Berkeley Continuing Studies

Data Analysis Bootcamp

Homework #1

Follow-Up Questions (Post-Data Analysis)

1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**
   1. Theater (839 projects), Music (540 projects), and Film & Video (300 projects) are the top three “successful” categories.
   2. However, Theater (493 projects), Technology (213 projects), and Film & Video (180 projects) are the top three “failed” categories.
   3. Plays (694 projects) is the most “successful” sub-category.
2. **What are some limitations of this dataset?**
   1. There are some missing data. All three pivot tables showed a default “blank” category.
   2. The early years of the data (i.e., 2009-2011) seem questionable. In 2009, all of the projects were successful. In 2010 and 2011, the line chart shows gaps or breaks over time.
   3. The last year of data 2017 only goes Mar 2017 suggesting that at the time of downloading the data, the 2017 had not completed yet. So we only had partial year data for 2017.
3. **What are some other possible tables and/or graphs that we could create?**
   1. I would try to answer the research question: Does the goal funded size of a project make a difference? That is, do smaller goal projects see more successful pledges more than larger goal projects? Thus, I would create a new column that breaks up the goal amounts to five groups: small, medium-small, medium, medium-large, and large currency amounts. I would create a graph against the category versus goal funding size.
   2. Do “staff pick” or “spotlight” projects get more pledges than those not handpicked by staff or spotlighted? I would create a pivot table of staff pick vs. campaign “status” and another one with “spotlight” vs. campaign “status.” I might want to do filter by country in case one country gets more attention than others.
   3. Finally, I wonder if some ‘frequently used’ text could be analyzed from the blurbs. While qualitative, I have a hunch that common ‘catchy’ words might be used in successful project versions ‘steer-clear-of-these-words-or-risk-failure/cancel’ outcome. I doubt Excel is the best application to do such analysis, but I am sure this idea could be done with lots of time and research. I would think about “Word Clouds” would be fun visualizations to get information by category (type of projects) as well as campaign status (successful, fail, cancel, or live).

Bonus Statistical Analysis Questions

**1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

Both mean and median provide insightful descriptive statistics. We see that on average that successful campaigns have more backers (182 funders) than failed campaigns (about 20 funders).

While the average might have some outliers (more or fewer funders), the median shows about 57 funders for successful projects compared to the median of 4 for failed projects.

**2. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

Yes, there is a lot of variability both in variance and standard deviation from the mean when compared to successful and failed campaigns.

In other words, we can see the spread of the data from the average. The variance is greater for successful campaigns who generate/attract more backers (a standard deviation of 816 backers) than failed campaigns with a standard deviation of 67 backers.