Assessment of Kickstarter Projects

Report, HW1 - Excel

Rob Pascarella

Rice Data Analytics and Visualization

Data Visualization Questions

1. Given the provided data, what are 3 conclusions that can be drawn about Kickstarter campaigns?
   1. US is the major player in the Kickstarter space
   2. Theatre, specifically ‘plays’ standout attempts
   3. Smaller projects typically have greater statistical change of success
   4. Bonus – time of the year shows little correlation to project success
2. What are some limitations of this dataset?
   1. The dataset only gathers from 21 countries. There are 195 countries globally.
   2. Currency vs. Origin of project not tracked; as to measure foreign investment interest.
3. What are some other possible tables and/or graphs that we could create?
   1. We didn’t do anything with average donation size. This could be a valuable study. The larger average donation size would point to larger conviction in the premise of the project, which could very well correlate to success.
   2. It was obviously noted that smaller projects were more successful per the data. But how small? If only $100 was targeted as the goal, and the project was funded at $400 – on a percentage basis that looks like an extremely successful project! But in reality, how hard was it really to draw that level of investment – especially versus a project attempting to attract $100k? Whole different ballgame.

Bonus 2 Verbiage

1. Does the mean or median summarize the data more meaningfully?
   1. The mean is most centrally descriptive of data in which we expect to be normally distributed – as to say we should have some high probabilistic reason to expect a lot of values around a certain area. The median is used more to describe a physical center in which half of the data is to either side. In our case, the fact that the median for unsuccessful projects is so unfathomably low (in regard to seeking investors) – is a most telling data point as to prediction of failure.
2. Using the data, determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

High Variance = data spread wider from the mean

Low Variance = data closer to the mean

High Std Dev = data spread over larger range of values

Low Std Dev = data closer to the mean

* 1. There is more variability in the Successful projects
  2. This makes sense because it is more likely that most of the backer data resides close to 0 with the unsuccessful projects: That is to say, a failed project can only have data within the 0-100% range of funding, while successful projects can be funded from 100%+ to infinitely funded (obviously a larger range of values available).