USC Viterbi Boot Camp – Week 1 Homework (Excel)

This folder contains an Excel workbook of Kickstarter campaigns over the last five years or so, as well as analysis of the success of the campaigns based on category, time of year, goal amount, and number of backers.

**Questions**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
2. What are some limitations of this dataset?
3. What are some other possible tables and/or graphs that we could create?

**Conclusions**

1. Campaigns with smaller goals tend to be successfully more often. (Those with goals less than $10,000 succeeded over 50% of the time, while anything above that number succeeded somewhere in the range of 20-49% of the time.)
2. Some categories are much, much more successful than others. Only theater and music are successful more than half the time, and even these depend on subcategory: for instance, music is uniformly successful in every category except faith and jazz (which have all failed), world music (which have all been canceled), and indie rock (which is only successful over 80% of the time, not 100%). The categories of food and journalism, on the other end of the scale, are almost never successful. It would be very prudent to examine a particular subcategory’s success rate before trying a campaign in it.
3. Location makes some difference as well, though not one as striking as the categories. Campaigns in Australia, for instance, are far less likely to be successful across the board. There have only been approximately 200 campaigns in non-English speaking countries (as opposed to nearly four thousand in countries with English as a primary language), which indicates both that the current data from any particular one of those countries may not be reliable and that Kickstarter is a relative unknown in those places. But even in countries with large, reliable datasets, the pattern of success across categories remains roughly the same.

**Limitations**

The dataset doesn’t break down the pledging numbers in any way, so it’s possible to attain the mean but not the median amount pledged. Median would probably be more useful in some cases, especially for discovering if a few big donors did most of the giving.

The size of campaigns across countries is not easily compared, because each goal is given (from what I can infer; the dataset isn’t explicit here) in the relevant currency. Not only that, but converting every amount of money into a single currency would require using historical data for exchange rates at the time of each campaign, as well as accounting for inflation.

One other piece of data that isn’t present but would be interesting is to see, for successful campaigns, at what point after the start date they reached their initial goal. (Or, for an even more thorough dataset, it could include the amount pledged so far over each day or week that the campaign was live.)

**Other Avenues of Investigation**

There’s a great deal of information left to examine here. What makes a successful campaign fund to 200% instead of barely over a hundred? How does the average donation for a campaign correspond to its goal? The “Staff Pick” and “Spotlight” columns weren’t even touched in this assignment. And textual analysis could be interesting, too: does the length of the name have any interesting correlations, or the length of the description? Are there words that often come up in the blurbs of failed campaigns? And so on.

**Bonus**

*Mean vs median:* The median seems to be a more useful statistic here in understanding the average campaign. Most failed campaigns, for instance, have very few backers: certainly closer to 4 (the median) than 18 (the mean). In fact, only about 300 of the 1500 failed campaigns had more than 18 backers, which doesn’t exactly match the common campaign. There are similar numbers for successful campaigns, too – in both cases, most campaigns are near the median, while the mean is pulled upwards by a few campaigns with an extremely high number of backers.

*Variability*: The variance of a successful campaign is much higher than that of a failed one. This makes sense, because successful campaigns could only need a low number of backers, or could have tens of thousands of backers; but a failed campaign is capped by definition. It could only have had a certain number of backers or else it would have been successful, whereas successful campaigns can have far more than they needed to succeed.