* Create a report in Microsoft Word, and answer the following questions:
  + Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

[Akira Hayashi]

1. From provided data, we couldn’t come up with “the magic formula” which guarantee successful outcome of crowdfunding projects. In other words, I cannot definitively say “if you do this, your project will succeed” from below perspective (for example, “if you create $X size of projects and start in winter in Y industry, you have Z% chance of success):

a. From seasonal perspective – Shown in “OutcomeByMonth” tab, there seems to be no good or bad month to start projects. It seems project started in summertime (i.e., June-July) seems to have slightly higher success but not high enough to guarantee the success of projects.

b. From project size perspective - There seems to be no correlation between the size of project and success (or failure), some big projects succeeded, and some others didn’t. The same is true for smaller projects in funding size. One more note to make in making this statement “b.” is some of the funding amounts are expressed in different currencies and I don’t have any currency conversion rate at hand at the time of this analysis. If we plug in right currency conversion table to express all projects in USD, the analysis might show more accurate trend. It should be noted, however, many of the projects are funded in US dollars (USD).

c. From category/sub-category perspective - We (namely I), also observed no correlation in between success/fail and industries. As shown in both “CampaignStatus per Category” and “CampaignStatus per Sub-category” tabs, proportion of fails/successes (despite different categories/sub-categories have different case numbers) are the same all across (more or less).

d. From yearly perspective – out of curiosity, though it wasn’t required for this module 1 challenge, I also created a pivot table from yearly perspective, hoping either success or failure trend would change as the years go by, but the result didn’t reveal any yearly trend (shown in “OutcomeByYear”) tab.

2. Shown in “StatisticalAnalysis” tab, likelihood of success or failure is not captured in given datasets. Failure or success factors are outside of our obtained data - whatever they were. It could be that some projects had better solicitation model than others, or it could be by just a sheer luck. The “mean” values are close each other between successful and failed projects and other values we have analyzed such as “median”, “minimum” as well as “maximum” number of backers are close enough thus unable to tell us any good stories. The only oddity was the “variance of the number of backers” where failed is almost 6 times higher. Further study is needed which is outside of our required time window for this analysis.

3. The conclusion is inconclusive – as I mentioned before, we were unable to find “the magic formula” to say definitively what attributed to the success or failure of each project but it doesn’t necessarily mean everything is happening by sheer luck. It is therefore fair to say our conclusions, with given data and depth of analysis with given time window is inconclusive.

* + What are some limitations of this dataset?

The size of record is only 1000 and I am not sure if this is effective and meaningful sample size. Also, the data is US (English) centric. Global market perspective might not be shown in this data. The data also does not clearly show uniqueness of each dataset. For example, a project could have failed first time, re-branded under different name then succeeded.

* + What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Perhaps some other economic data such as stock market price or unemployment rate by months which might show whether we have more funding available in our society (and overlay analysis with them) might reveal more interesting stories. If projects were starting in economic downtime, for example, they might have had tougher time in succeeding. I’d also be interested in looking at regional or geographical variables.

* Use your data to determine whether the mean or the median better summarizes the data.

Neither mean nor median has shown strong trends.

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

As shown in “the variance of the number of backers” in “StatisticalAnalysis” tab failed projects is 6 times higher than successful projects. During our study group preparation meeting for in writing of this report, this was unexpected and didn’t make sense to us. We were expecting failed projects, in general, would have lower variance. “The minimum number of backers” = 0 for failed projects, on the other hand, made sense to us since no one was backing the projects (thus it failed).