**Report**

**Assignment #1: Excel Kickstarter Analysis**

**Christian Ott**

The provided Kickstarter data set (about 4000 rows) reveals a number of interesting insights about Kickstarter campaigns.

The project assignment has the following questions:

1. What are three conclusions we can make about Kickstarter campaigns given the provided data?
2. What are some of the limitations of this dataset?
3. What are some other possible tables/graphs that we could create?

**Answer to 1: Conclusions**

There is a lot that can be said about the data. Based on the three pivot tables and plots the project assignment suggest to create, we directly observe the following:

(1) Pivot 1 "Outcome vs. Category"

Theater (performing arts) projects make up the largest fraction of projects across all categories. However, they also have a high failure rate of above 30%.

Second are music projects in terms of total number, but they are more successful on average than the Theater projects. Food projects have the lowest success rate (about 15%-20% by eye).

The smallest number of projects (24) were submitted in the Journalism category and all were cancelled. Since all other categories have at least about ten times more submissions, the number of Journalism projects is too small to draw conclusions based on their outcomes.

(2) Pivot 2 "Outcome vs. Sub-Category"

Based on this analysis, we draw the obvious conclusion that the vast majority of projects submitted to Kickstarter (or at least contained in the preset dataset; we don't know how representative it is!) are in the Theater--Plays subcategory. There are more than three times as many projects in this subcategory than in any other. More than 50% of these projects turn out successful.

The second-largest subcategory is the Music--Rock subcategory. Interestingly, all projects in this subcategory are successful.

Subcategories with particularly high failure (and cancellation) rates include animation, drama, video games, web, science fiction, food trucks, jazz, and others. Highly successful categories in addition to rock include indie rock, pop, documentary, nonfiction (writing?), tabletop games, classical music, short movies, television, small batch, and hardware.

(3) Pivot 3 "Outcomes vs. Starting Month"

Here, the obvious conclusion is that it is not a good idea to start a Kickstarter project near the end of the year. While the overall number of projects started then is the lowest, the success rate is also the lowest in December (44%). This is probably due to two reasons: (a) People have less freely spendable money at the end of the year because of their holiday expenses (Christmas presents! Travel). (b) And this is an assertion that could be tested with more data: Most of the pledges come in soon after the project is launched and the pledge rate fades over time. If (b) wasn't the case and the pledge rate would be (nearly) constant of the duration of most projects, then even those projects started in December should be able to make up for their poor start.

The highest success rate is for projects started in May (61%). So apparently it is a good idea to start your project in late Spring (May)!

**Answer to 2: Limitations of the Dataset**

The data set has a number of limitations. These include but are not necessarily limited to:

* Size -- There have been almost 400,000 Kickstarter projects thus far. Our Dataset with 4000 is about 1% of these projects. This may not be a sufficiently large sample to be statistically significant. This could be checked.
* We do not know how the projects in our sample were chosen. A random choice would give the sample the best chance of being statistical significant.
* There is no information on the pledge rate as a function of time since project start.
* There is no detailed information on how the projects were promoted by their owners. It's clear that more heavily promoted projects will tend to do better. The data show ("Spotlight"/"Staff Pick" columns) that those projects that were in the "Spotlight" (i.e. probably promoted by Kickstarter on its webpage[?]) did better. It'd also be interesting to know how the pledge rate varies with promotion events.
* Analysis by country is not really possible since many countries have very few projects in the dataset.

**Answer to 3: Further Analysis**

There is much room left for further analysis. Here are some examples of what we could look at:

* One important thing to consider are fractional success rates as a function of category, subcategory, project goal, and start date.
* Understanding average pledge size as a function of category, subcategory, project goal can be very useful for setting up new projects.
* Getting an understanding how project duration correlates with project success will be another useful analysis (Pivot table, plot).
* Similarly, understanding outcome as a function of project goal and category will be interesting. There may be some systematics there that can provide useful guidance for new project submissions.
* Looking at how trends in all of the above change of the years. It's likely that over time, people learned from their past mistakes (or from others mistakes!) and changed their approaches.
* One could also consider trends by country, but some countries have a very small number of projects in the dataset.
* And more!