**HW01 - EXCEL – STARTERBOOK**

The summary of Kickstarter Campaigns:

1. Successful projects: 53%
2. Failed Projects: 37%
3. Canceled projects: 8%
4. Live projects: 1%

**In first look, this Campaign is assessed to be effectively in term of the total numbers of project status (successful, failed, canceled or live).**

Now, on deeper analysis, parent categories have highest numbers of successful starter projects:

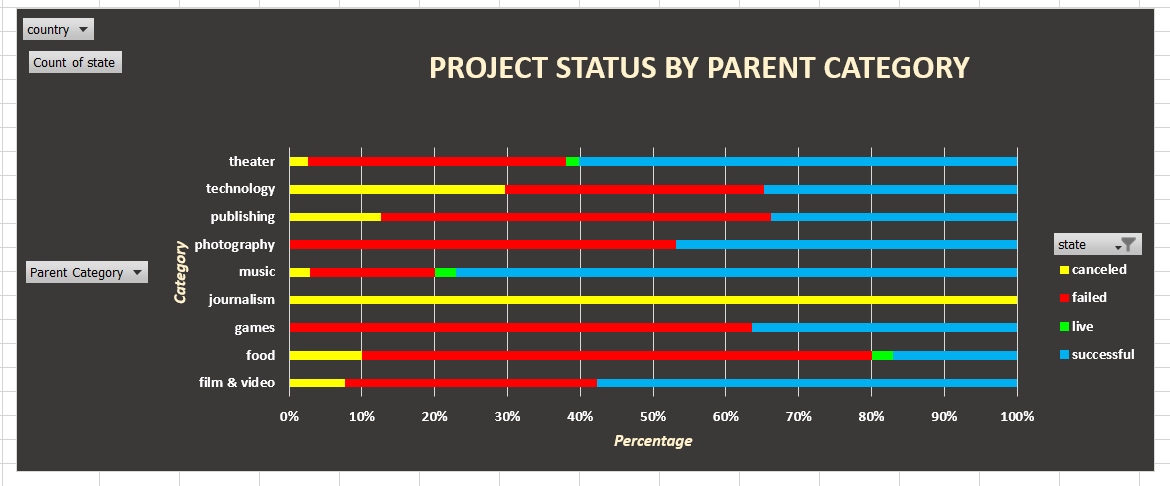
1. Theater (839)
2. Music (540)
3. Film & Video (300)

While as, parent categories have highest numbers of failed starter projects:

1. Theater (493)
2. Technology (213)
3. Film & Video (180)

**In term of ratio of successful over failed projects for each parent category, music gained the highest rate of appx. 78%. Then, there are only three parent categories having over 50% of success rate: music, film & video and theater respectively. Meanwhile, the highest rate of failure is for food (70%), then the second tier is for games (65%), and the third tier for Technology and publishing (appx. 60%). (As illustrated in Figure 1.1)**

Figure1.1:



**BRIEF ON ANALYTICAL FINDINGS:**

**In term of successful state of sub-category, the ranking is as follows:**

1. **Plays (theater): 694**
2. **Rock (music): 260**
3. **Documentary (Film & video): 180**

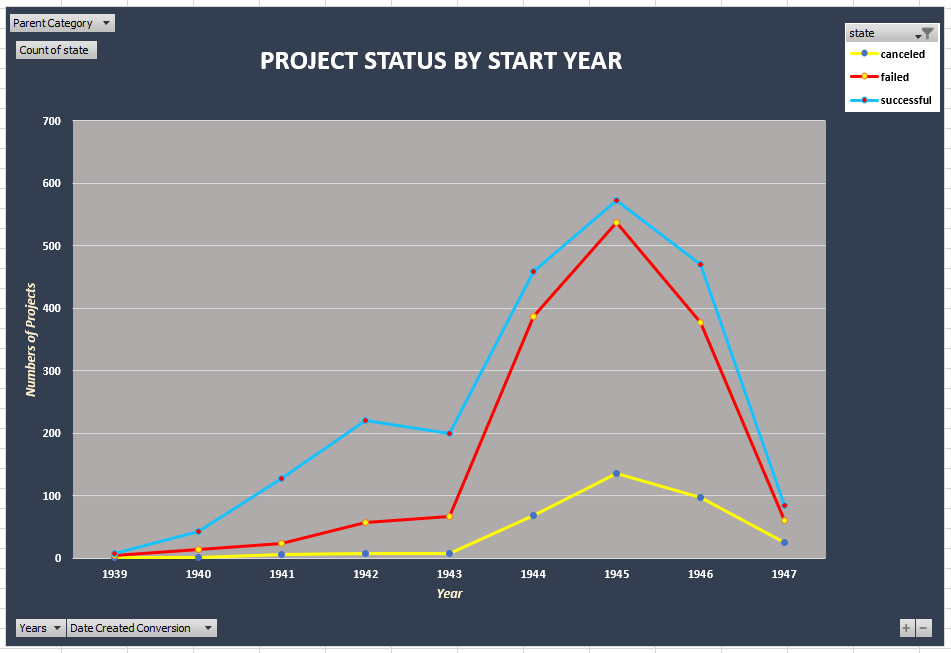
**However, plays also recorded the highest number of failed projects (353). The second role belongs to food trucks (120) and wearables – technology (120) and the third position belongs to video games (100) and animation (100).**

In term of conversion time, there is probably the trend of project status cycle:

1. From January to April: the number of successful projects decreased gradually in small range of 200 and jump up to reach the peak at appx. 230 in May, then keeps the continuous and sharp down movement until December, even though there is the improvement of upside reversal in October and November. At December, the numbers of successful and failed projects have closed to the same at appx. 105.
2. The failed project has the similar trend to successful projects, but at the smaller range of 100-150 per month.
3. The monthly number of canceled projects is stable at 25 in average.

Upon the graph for years of created conversion time, there is very clearly form of investment flow booming in year 1945. The flow of investment as well as the successful rate increased sharply in the period of 1941 – 1945. The downtrend is sharply clear from 1945 to 1947. (As illustrated Figure 1.2)

Figure 1.2



**Hence, we can conclude that the campaign does not show its effectiveness and performance as great as it was before 1976.**

**LIMITATIONS ON DATASET:**

The analysis is based on numbers of three groups: successful, canceled and failed of two main categories: parent and sub-category. And we can see only few categories turned to be successful among the majority of categories in failure status. However, the numbers of projects launched in this minority is dominant, like plays.

That is why in the first review on the total numbers of projects ignoring category consideration, it seems this campaign sounds be successfully. But in my opinion, it sounds this analysis needed more deeper statistical graphs or tables to test its reliability of dataset, the quantitative analysis on project efficiency has not been concerned yet. For instance, how effective ranking project state is based on fund size criteria.

Furthermore, the limitation of the dataset could be the effects of extreme values or other saying is the outliers can distort the accuracy of the dataset’s statistical analysis.

**SUGGESTIONS ON FURTHER STATISTICAL ANALYSYS:**

1. Hypothetical testing on correlation between successful and failed groups in each category or sub-category.

Firstly, I used the formula of correlation coefficient for two datasets: failed and successful projects in context of sub-category. This **Correl. is +0.92**. This figure sounds significantly close to +1. That is not the good sign. We need this figure is as small as possible so these two sub datasets can be recognized as independent sets.

1. The quantitative analysis on project efficiency has not been concerned yet. For instance, how effective ranking project state is based on fund size. I think that is the reason why there are two more bonus graphs to be required: outcomes based on goal and the table of crucial statistical ratios of successful and failed projects in term of backers count criteria.
2. The further graphs should be used:

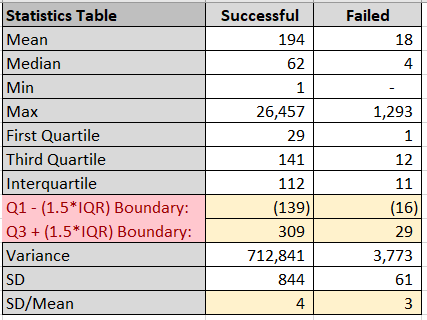
* Quantiles and Boxplot.
* Statistical Summary Table

1. **Analysis on outcomes based on goal:**

There is the obvious trend of percentage of successful and failed projects in terms of fund goal:

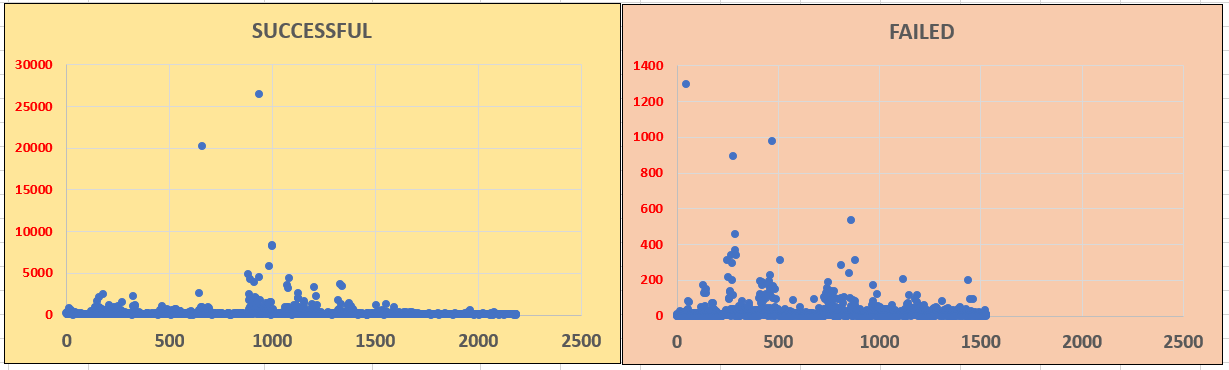
* Successful percentage declines from 70% to 20% when goal size increases from 1000 to greater than 50000.
* Oppositely, the failure percentage increases from 25% to 60% and the cancellation percentage also increases from appx. 5% to over 20%.
* The percentages of successful and failed projects are stable in the goal range of 10,000 to 45,000 at the range of 40% - 50%. Whereas, the percentage of canceled projects is stable at 10% in goal range of 10,000 to 25,000. With the goal range of 20,000 – 35,000, the percentage of successful projects keeps stably below 10 % of the percentage of failed projects.
* The greatly clear visualized conclusion that projects with the goal under 1,000 has the highest successful percentage (70%) and lowest failure percentage (20%). The project having goal over 50,000 has the highest percentage of failure (60%) and lowest percentage of successful (under 20%).

1. **Analysis on outcomes based on backer counts:**



The number of backers in successful projects fluctuates in a greatly wider range than the number of backers in failed projects. It could be the max number of successful projects is extreme high in comparing to the min and the max of the failed projects as well.

Hence, in my opinion, for successful projects, the mean should not be used but the median can be used instead. But for failed projects, as its closer scatter slotting range of backers counts, the mean is more reliable with its max and min values less extreme far from its mean value (as illustrated in Figure 1.3).



Furthermore, the ratio of SD over mean of successful is higher than that of failed projects about 37%. Since this percentage is greatly high, we can conclude that successful projects have clearly higher variability than failed projects do.

However, if the two datasets of successful and failed projects are analyzed by using Boxplot graphing, there are the big numbers of outliers appeared in the graphs for both datasets (as illustrated in Figure 1.4). **does it make sense?**

Figure 1.4

