Asmision Challenge Report ---- Data Analysis for The Success of Kickstarter Campaigns

Gaoming Lin

# Introduction

The goal of this report is to implement data analysis for the success of Kickstarter campaigns based on given materials. Then data-driven recommendations should be given.

Based on tables in the SQL file, there are 13 parameters. Campaign IDs and campaign names are unrelated parameters. Currency IDs can be used to transform the amount of goal and pledged, which can be ignored after transformation. Same as launched time and deadline time, they can be neglected after the calculation of campaign duration. As a result, there are 8 related parameters, which can be classified into 3 layers.

The top layer is the ultimate parameter “outcome”, and it shows the result of the campaign. The secondary layer has 3 parameters: “goal”, “pledged” and “backers”. They can influence the top parameter directly. There might be some internal relationships between secondary parameters.

Other parameters, such as duration, country, categories and sub-categories, belong to the basic layer. They are key variables that influence secondary parameters.

A screenshot of a computer

Description automatically generated with low confidence

Fig. 1. The hierarchy of three kinds of parameters.

# Analytical process

## Pre-Operation

A SQL file, containing several groups of business data, is given in this admission challenge. However, related information is separated into several tables, which is not beneficial for data analysis. As a result, it is necessary to summarize all related information into one table “campaign”.

Next, transform all data of goals and pledged into USD format. Finally, calculate campaign durations. The table, named as “final” after pre-operation, is shown in Fig.2.

Graphical user interface

Description automatically generated

Fig. 2. The complete table has 15000 rows of data. Only 20 rows are shown here as an example.

## The Relationship between “Outcome” and Secondary-Layer Parameters

There are three secondary-layer parameters. So, the relationship between “outcome” and “goal”, the relationship between “outcome” and “pledged” and the relationship between “outcome” and “backers” should be analyzed. Among them, the relationship between “outcome” and “goal” can answer the first question in part 1 (Are the goals for dollars raised significantly different between campaigns that are successful and unsuccessful?).

**Tab. 1. The table of relationship between “Outcome” and Secondary-Layer Parameters. Results were calculated by MySQL.**

|  |  |  |  |
| --- | --- | --- | --- |
| **OUTCOME** | **NUMBER** | **AVERAGE GOAL (USD)** | **MEDIAN GOAL (USD)** |
| successful | 5319 | 9859.94 | 4000 |
| not successful | 9681 | 109785.95 | 7000 |
| **AVERAGE PLEDGED (USD)** | **MEDIAN PLEDGED (USD)** | **AVERAGE**  **BACKERS** | **MEDIAN BACKERS** |
| 22591.16 | 5196 | 282.15 | 70 |
| 1592.7 | 100 | 18.17 | 3 |

Median was more suitable to represent the characteristics of their relationships, compared with average values. It is obvious that successful campaigns usually have lower goals, higher pledged and more backers.

## The Internal Relationship between Secondary-Layer Parameters

The analysis of possible internal relationship among “goal”, “backers” and “pledged” should be implemented. The relationship between “goal” and “backers” can answer the fourth question in part 1 (What was the amount the most successful board game company raised? How many backers did they have?) and the second question in part 3 (How many backers will be needed to meet their goal?).

Board game belongs to the category of "game". There are 8 subcategories in "game" category: 13. Games; 14. Tabletop Games; 44. Video Games; 66. Mobile Games; 70. Playing Cards; 113. Puzzles; 122. Live Games; 134. Gaming Hardware. Among them, 14. Tabletop Games, 70. Playing Cards and 113. Puzzles are board game items. The result (Fig. 3.) showed that the most successful board game campaign is Ghostbusters™: The Board Game with 250000 USD goal and 8396 backers.

Graphical user interface

Description automatically generatedFig. 3. The top8 board game campaigns in the term of raised goal.

## The Relationship between Secondary-Layer Parameters and Basic-Layer Parmeters

There are 12 relationships between the secondary layer and the basic layer. Among them, the relationship between “categories” and “backers” and the relationship between “sub-categories” and “backers” can answer the second question in part 1 (What is the top/bottom 3 categories with the most backers? What is the top/bottom 3 subcategories by backers?) and the third question in part 3 (How many backers can the company realistically expect, based on trends in their category?). The relationship between “categories” and “goal” and the relationship between “sub-categories” and “goal” can answer the third question in part 1 (What is the top/bottom 3 categories that have raised the most money? What is the top/bottom 3 subcategories that have raised the most money?). The relationship between “country” and “pledged” and the relationship between “country” and “backers” can answer the fifth question in part 1 (Rank the top three countries with the most successful campaigns in terms of dollars (total amount pledged), and in terms of the number of campaigns backed.). The relationship between “duration” and “goal” can answer the sixth question in part 1 (Do longer, or shorter campaigns tend to raise more money? Why?).

In term of backers, the top 3 categories are 7, 13, 5, while the bottom 3 categories are 14, 15, 6. In term of backers, the top 3 sub-categories are 14, 8, 44, while the bottom 3 categories are 149, 131, 65. In term of goal, the top 3 categories are 2, 13, 1, while the bottom 3 categories are 14, 8, 6. In term of backers, the top 3 categories are 4, 58, 9, while the bottom 3 categories are 149, 159, 144. In term of pledged, the top 3 countries are 2, 1, 3, while in term of backers the top 3 countries are 2, 1, 4.

More details and data can be found in submitted SQL file. More reliable and straightforward conclusions will be drawn in visualization analysis.

# Visualization

Visualization is an important method to presenting data and conclusion straightforward with graphs. In this report, visualization was utilized to show data’s distribution and ranks of different parameters.

## Tools and Operation

JMP Pro 16, a very famous statistics analysis software, was used as visualization tools. The SQL table, including 15000 rows, was copied to JMP Pro, shown in Fig. 3.

Table

Description automatically generated

Fig. 4. JMP Pro 16 table interface.

Then JMP graph builder can visualize the data to satisfy the challenge requirements, shown in Fig. 4.

Graphical user interface, application

Description automatically generated

Fig. 5. JMP Pro 16 graph builder interface.

## Plots

Five plots were given to solve five problems in this business analysis.

1. The relationship between outcome and goal.

Chart

Description automatically generated with low confidence

Fig. 6. This plot showed goal had different distribution with different outcome.

2. The distribution of backers with different categories/sub-categories.

Chart, bar chart

Description automatically generated

Fig. 7. This plot showed the distribution of backers with different categories

Chart, histogram

Description automatically generated

Fig. 8. This plot showed the distribution of backers with different sub-categories.

1. The distribution of goal with different categories/sub-categories.

Chart, histogram

Description automatically generated

Fig. 9. This plot showed the distribution of goal with different categories.

Chart, histogram

Description automatically generated

Fig. 10. This plot showed the distribution of goal with different sub-categories.

1. The relationship between goal and backers

Chart

Description automatically generated

Fig. 11. This plot showed the relationship between goal and backers with successful outcome and not successful outcome, respectively.

1. The relationship between duration and goal

Chart

Description automatically generated

Fig. 12. This plot showed the relationship between goal and duration with successful outcome and not successful outcome, respectively.

# BUSINESS RECOMMENDATION

There are three questions: 1. What is a realistic Kickstarter campaign goal (in dollars) should the company aim to raise; 2. How many backers will be needed to meet their goal; 3. How many backers can the company realistically expect, based on trends in their category. Visualization analysis will be given based on these three questions.

## Recommendations

Based on Fig. 6, the successful campaigns’ goals usually concentrated in the range of [0, 8000] USD. Their median is 4000 and their mean is 9859.94. However, the unsuccessful campaigns’ goals had a much wider distribution. And they did not have obvious concentration in a specific range. So, a realistic campaign goal should be about 4000.

Based on Fig. 11, according to the result of successful campaigns, the goal and backers have positive correlation. As for a campaign with a goal of 4000, about 100 backers were needed. However, the unsuccessful campaigns tended to have much fewer backers and much higher goals.

Based on Fig. 7, this plot showed the distribution of backers with different categories. In term of backers, the top 3 categories are 7, 13, 5, while the bottom 3 categories are 14, 15, 6. To order to get more backers, the companies can increase the ratio of the categories, whose IDs are 7, 13 or 5, in their campaigns.

## Conclusion

The outcome of campaigns is influenced by many parameters. The goal, pledged and backers are secondary parameters that can lead to outcomes directly. A too high goal will lead to an unsuccessful outcome. Successful campaigns usually have goals distributing in a specific range. Among these three secondary parameters, there are also internal relationships. A too high goal will lead to a few backers and pledged.

The basic parameters are categories, sub-categories, country and duration. They will influence goals, backers and pledged, then they will influence outcomes indirectly. Only specific categories, sub-categories, countries and the range of duration can be beneficial to improve the secondary parameters in order to get more successful outcomes.