

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

Gaoming Lin

I. 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.



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

II. ANALYTICAL PROCESS

A. 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.

id	name	category_id	sub_category_id	country_id	duration	goal	pledged	backers	outcome
1	Regalids	2	23	2	30	15000	20	3	failed
2	One To Ten	2	47	1	30	295,2276	946,2262	23	successful
3	Future Gun - A short film about kids and advanced hardware	2	24	2	30	3000	2497	28	failed
4	Fourth Wave Apparel—Vintage Fashion for the Modern Feminist	9	52	2	30	6000	8795	238	successful
5	The Rich Lifestyle	2	9	2	25	2000000	2	2	failed
6	Perks Delivery	11	27	2	37	2500	230	5	failed
7	In a Dark Dark House - Directed by Larry Moss	10	20	2	45	55000	9336	45	failed
8	CHALK:SPONTANEOUS	2	23	2	30	6000	365	4	failed
9	Pat Englehorn - Beer and Sumo	2	54	2	30	50000	0	0	failed
10	Beer elements (all-finished shirts and beard care for men)	4	12	2	30	12000	396	11	failed
11	Digitalstrut Art	11	97	2	20	30000	111	4	failed
12	Experiencing Farm to service the local food bank.	4	107	2	30	2500	405	5	failed
13	Church Hive: Church living made easy!	13	40	2	30	18000	101	2	failed
14	The Visual Media Directory Project	11	26	2	60	10000	2	2	failed
15	Top Street - The Next To Serious Side to Diabetes	1	10	2	30	2500	2325	6	successful
16	"Baby Hungry" Short Film Raising Funds	2	9	2	30	1500	1525	23	successful
17	Turn "Narcosis-Rage" in "Blood" into 12 inches of vinyl!	2	3	2	30	1000	1464	43	successful
18	Fisher! for Chickens	2	47	2	16	2895	3667.13	75	successful
19	ROCKA RECORD FUND	2	38	2	30	3000	5847	99	successful
20	William Charles and On Road Trippers NEW CD	2	35	2	60	800	425	8	failed

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

B. 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.

C. 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.

id	name	sub_category_id	goal	backers	outcome
394	Your Way Game Board	14	250000	0	failed
7703	Ghostbusters™: The Board Game	14	250000	8396	successful
4197	Epic Beer Pong - Portable & Interactive LED Beer Pong	14	202329	9	canceled
8988	BattleAxe	14	110000	130	failed
2129	Xia: Legends of a Drift System	14	100000	3293	successful
4751	Wits & Wagers Epic Geek Edition (Canceled)	14	100000	296	canceled
5769	Seriously?!	14	100000	90	failed
9960	Gloomhaven (Second Printing)	14	100000	40642	successful

Fig. 3. The top8 board game campaigns in the term of raised goal.

D. The Relationship between Secondary-Layer Parameters and Basic-Layer Parameters

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.

III. 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.

A. 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.

	File	Edit	Tables	Rows	Cols	DOE	Fit	Graph	Tools	View	Window	Help
123	Source											
			id	categories	sub_categories	country	duration	goal	pledged	backers	outcome	result
1			38	13	13	92	2	30	25000	0	cancelled	not successful
2			50	5	8	2	30	5000	912	19	cancelled	not successful
3			63	1	10	2	50	50000	4373	30	cancelled	not successful
4			66	13	40	1	30	54900.0276	0	0	cancelled	not successful
5			89	11	31	1	30	5176.314	72.468	1	cancelled	not successful
6			98	3	3	2	30	18000	1979	16	cancelled	not successful
7			99	7	13	2	31	50000	5816	24	cancelled	not successful
8			101	1	10	2	35	1500	9	1	cancelled	not successful
9			133	9	52	2	30	8000	0	0	cancelled	not successful
10			135	11	26	2	30	4303	0	0	cancelled	not successful
11			151	5	8	20	40	9549.8209	77.1017	20	cancelled	not successful
12			169	13	40	2	30	20000	0	0	cancelled	not successful
13			177	12	28	2	30	15000	0	0	cancelled	not successful
14			190	7	66	3	30	29312.2908	17.589	2	cancelled	not successful

Fig. 4. JMP Pro 16 table interface.

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

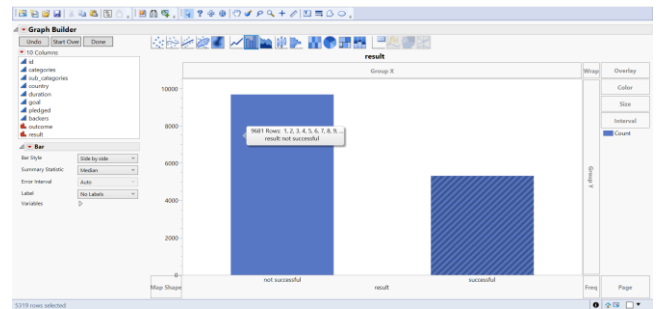


Fig. 5. JMP Pro 16 graph builder interface.

B. Plots

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

1. The relationship between outcome and goal.

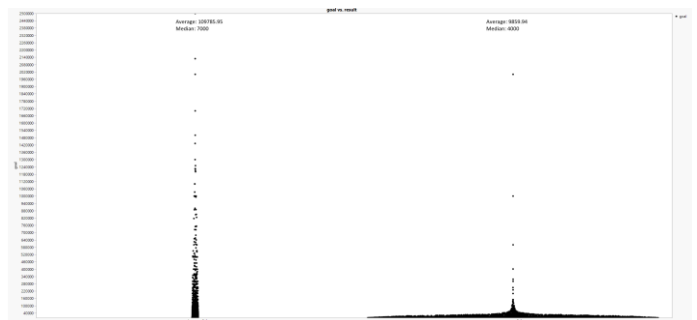


Fig. 6. This plot showed goal had different distribution with different outcome.

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

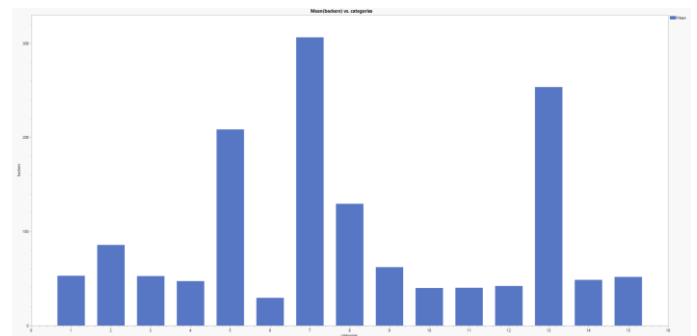


Fig. 7. This plot showed the distribution of backers with different categories

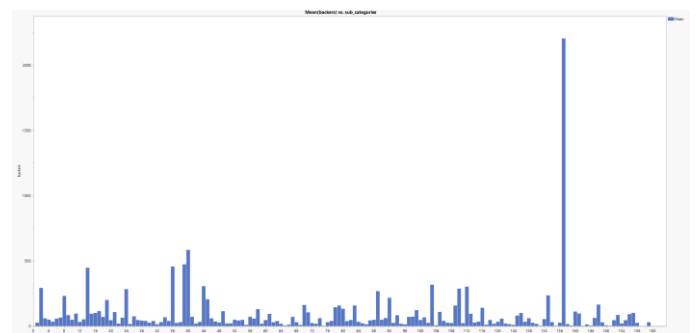


Fig. 8. This plot showed the distribution of backers with different sub-categories.

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

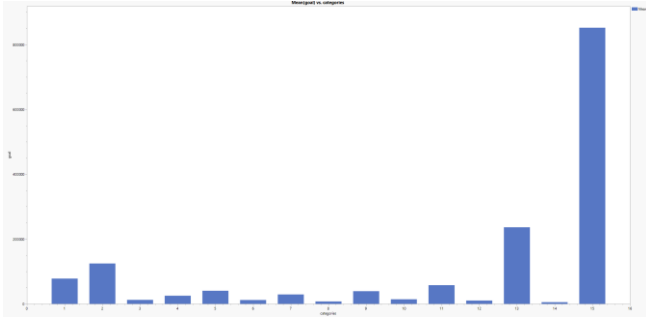


Fig. 9. This plot showed the distribution of goal with different categories.

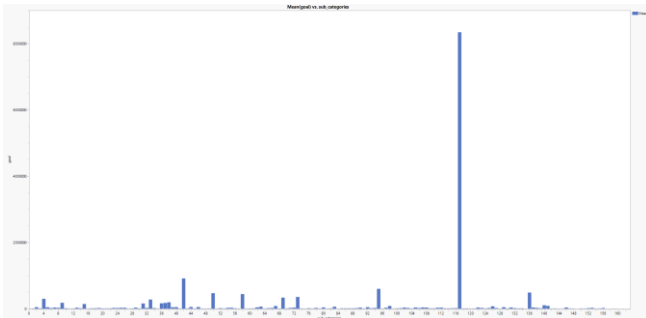


Fig. 10. This plot showed the distribution of goal with different sub-categories.

4. The relationship between goal and backers

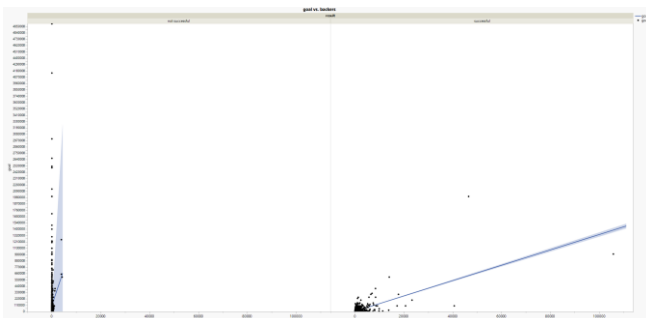


Fig. 11. This plot showed the relationship between goal and backers with successful outcome and not successful outcome, respectively.

5. The relationship between duration and goal

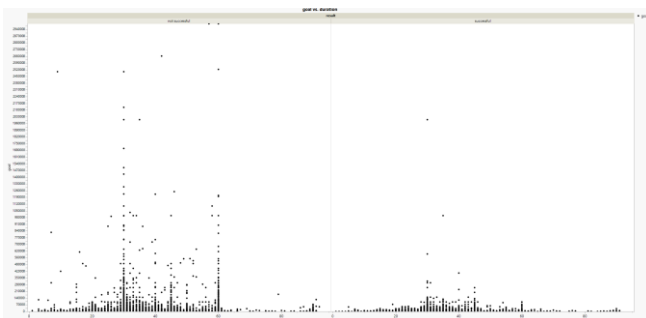


Fig. 12. This plot showed the relationship between goal and duration with successful outcome and not successful outcome, respectively.

IV. 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.

A. 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.

B. 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.