

Success rate analysis of projects on Kickstarter in Excel

Over \$2 billion has been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

In this project, the Kickstarter database (4,000 past projects) was analyzed. Obtained results, allowed to uncover some hidden market trends, and help to determine the main features required to run the successful projects.

Input Data

Calculations and analysis were performed using Microsoft Excel (Office 365) program (Microsoft Excel, <https://office.microsoft.com/excel>).

All data can be found in the Excel file: [01-Excel_Challenge.xlsx](#)

This workbook has 6 sheets:

- Table
- PivotTable_1
- PivotTable_2
- PivotTable_3
- Bonus
- Bonus_Stat

The Kickstarter database can be found in the Table sheet. It contains the following data::

name – the name of the project

blurb – brief description of the project

goal – the required amount of money

pledged – obtained amount of money

state – the associated campaign was successful, failed, canceled, or is live

country – country (US, GB, AU etc.)

currency – the world currencies by countries

deadline – date when project was ended (in Unix timestamps)

launched_at – expected date when the project would be ended (in Unix timestamps)

backers_count – the total amount of backers donated for the project

Category and Sub-Category – type of industry project belongs (for example, food, music, games etc.)

Calculations and Formulas:

Essential calculations and formatting were performed in the Table spreadsheet:

- 1) Conditional formatting was used to fill each cell in the **state** column with a different color, depending on whether the associated campaign: successful (**green**), failed (**red**), canceled (**yellow**), or live (**blue**).
- 2) A new column called **Percent Funded** (column O) was created. These data are used to uncover how much money a campaign made to reach its initial goal. Formula is:

$$\text{Percent Funded} = \frac{\text{pledged} \cdot 100}{\text{goal}}$$

Then, conditional formatting was used to fill each cell in the Percent Funded column using a three-color scale. The scale starts at 0 and became a dark shade of red, transitioning to green at 100, and blue at 200.

- 3) A new column called **Average Donation** (Column P) was created. These data are used to show how much each backer for the project paid on average. Formula is:

$$\text{Average Donation} = \frac{\text{pledged}}{\text{backers_count}}$$

- 4) The **Category and Sub-Category** column N was split into two parts: (i) **Category** (column Q), and (ii) **Sub-Category** (column R). Excel's functions LEFT and RIGTH were used to do it. Example of used formulas for row #2 is presented below:

$$\text{Category} = \text{LEFT}(N2, \text{FIND}("/", N2) - 1)$$

$$\text{Sub - Category} = \text{RIGHT}(N2, \text{LEN}(N2) - \text{FIND}("/", N2))$$

- 5) To determine the dates when project was started and ended, the **deadline** and **launched_at** columns (with dates in Unix timestamps format) were converted it into an into Excel's date format using formulas:

$$\text{Date Created Conversion} = (((\text{launched_at}/60)/60)/24) + \text{DATE}(1970,1,1)$$

$$\text{Date Ended Conversion} = (((\text{deadline}/60)/60)/24) + \text{DATE}(1970,1,1)$$

Two new columns **Date Created Conversion** (column S) and **Date Ended Conversion** (column T) were created.

Data analysis:

PivotTable_1. A new sheet with a pivot table (PivotTable_1) was created to analyze initial worksheet to count how many campaigns were successful, failed, canceled, or live per **Category** (**Fig.1, PivotTable_1**).

Additionally, a stacked column pivot chart that can be filtered by country based on the table have been produced (**Fig. 2, PivotTable_1**).

	A	B	C	D	E	F
1	country	(All)				
2						
3	Count of pledged Column Labels					
4	Row Labels	successful	failed	canceled	live	Grand Total
5	film & video	300	180	40		520
6	food	34	140	20	6	200
7	games	80	140			220
8	journalism			24		24
9	music	540	120	20	20	700
10	photography	103	117			220
11	publishing	80	127	30		237
12	technology	209	213	178		600
13	theater	839	493	37	24	1393
14	Grand Total	2185	1530	349	50	4114

Fig. 1. Pivot table: Filter = country; Rows = Category; Columns = state; Values = Count of pledged.

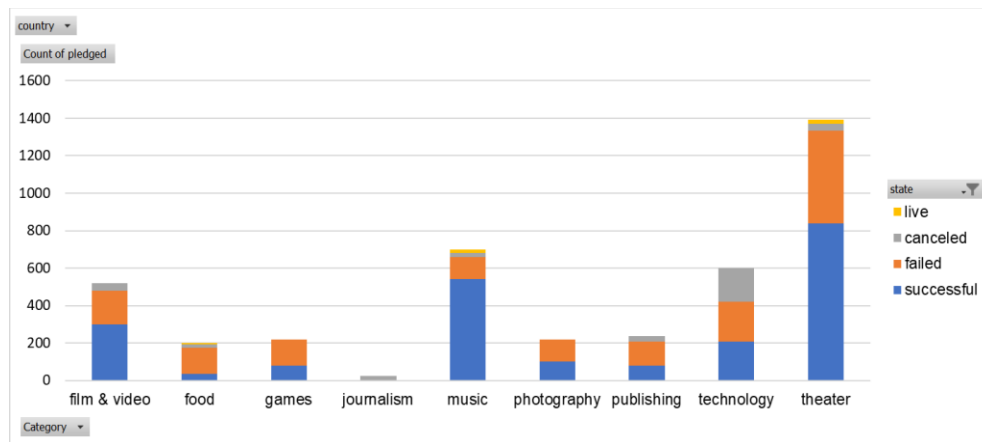


Fig. 2. A stacked column pivot chart that can be filtered by country based on the table in PivotTable_1.

PivotTable_2. A new sheet with a pivot table (PivotTable_2) that will analyze the initial sheet to count how many campaigns were successful, failed, canceled, or live per Sub-Category (**Fig.3**). Additionally, a stacked column pivot chart that can be filtered by country based on the table have been produced (**Fig. 4, PivotTable_2**).

	A	B	C	D	E	F
1	country	(All)				
2	Category	(All)				
3						
4	Count of pledged	Column Labels				
5	Row Labels	successful	failed	canceled	live	Grand Total
6	animation		100			100
7	art books			20		20
8	audio			24		24
9	children's books		40			40
10	classical music	40				40
11	documentary	180				180
12	drama		80			80
13	electronic music	40				40
14	faith		40		20	60
15	fiction		40			40
16	food trucks		120	20		140
17	gadgets		20			20
18	hardware	140				140
19	indie rock	140	20			160
20	jazz		60			60
21	makerspaces	9	11			20
22	metal	20				20
23	mobile games		40			40
24	musical	60	60	20		140
25	nature		20			20
26	nonfiction	60				60
27	people		20			20
28	photobooks	103	57			160
29	places		20			20
30	plays	694	353		19	1066
31	pop	40				40
32	radio & podcast	20				20
33	restaurants		20			20
34	rock	260				260
35	science fiction			40		40
36	shorts	60				60
37	small batch	34			6	40
38	space exploratio	40	2	18		60
39	spaces	85	80	17	5	187
40	tabletop games	80				80
41	television	60				60
42	translations		47	10		57
43	video games		100			100
44	wearables	20	120	60		200
45	web		60	100		160
46	world music			20		20
47	Grand Total	2185	1530	349	50	4114
48						

Fig. 3. Pivot table: Filter = country and Category; Rows = Sub-Category; Columns = state; Values = Count of pledged.

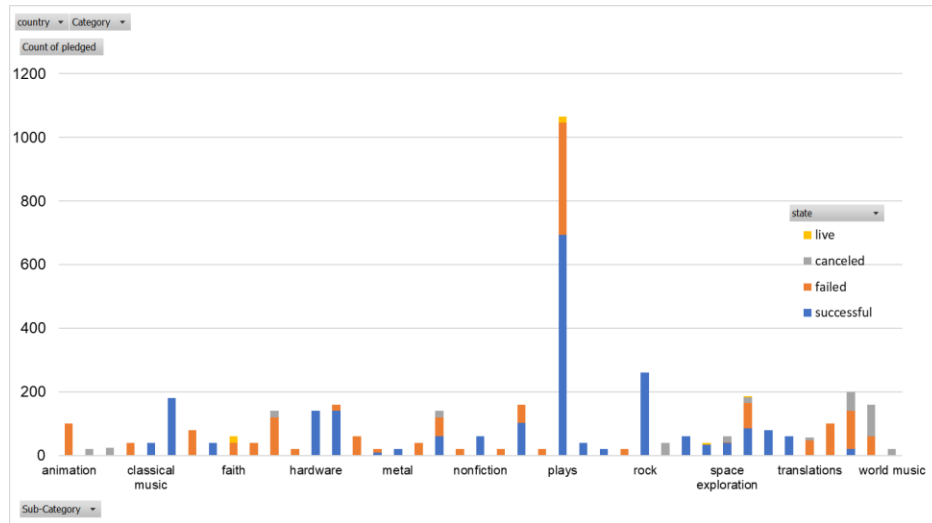


Fig.4. A stacked column pivot chart that can be filtered by country and Category based on the table in PivotTable_2.

PivotTable_3. A new sheet with a pivot table (PivotTable_3) was created with a column of state, rows of Date Created Conversion, values based on the count of state, and filters based on Category and Years (Fig.5). Also, a pivot chart line graph that visualizes this new table is presented on Fig.6 (PivotTable_3).

1	Category	(All)			
2	Years	(All)			
3					
4	Count of state	Column Labels			
5	Row Labels	successful	failed	canceled	Grand Total
6	Jan	182	149	34	365
7	Feb	202	106	27	335
8	Mar	180	108	28	316
9	Apr	192	102	27	321
10	May	234	126	26	386
11	Jun	211	147	27	385
12	Jul	194	150	43	387
13	Aug	166	134	33	333
14	Sep	147	127	24	298
15	Oct	183	149	20	352
16	Nov	183	114	37	334
17	Dec	111	118	23	252
18	Grand Total	2185	1530	349	4064

Fig. 5. Pivot table: Filter = Category and Years; Rows = Date Created Conversion; Columns = state; Values = Count of state.

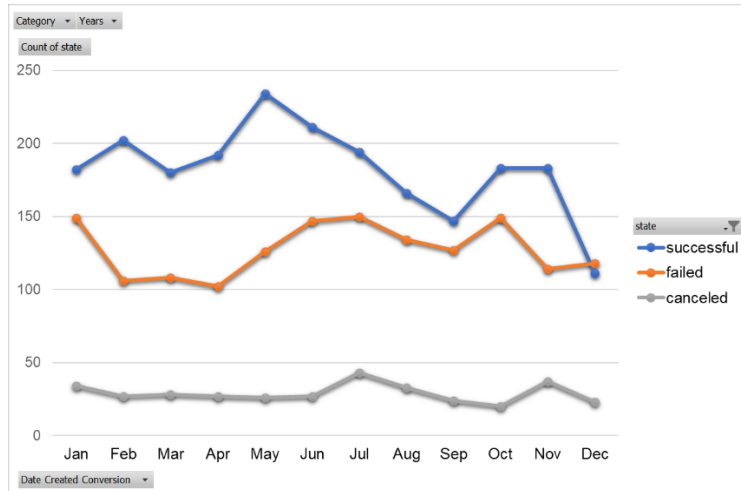


Fig.6. Visualization of PivotTable_3.

Conclusions:

- Journalism is a less successful and less profitable type of campaign.
- Obtained results showed that two of the most successful and profitable categories of the projects are Theater and Music: Theater > Music
- In the Theater-Category, the most successful division is Plays. In the case of the Music-Category, Rock music is the most fortunate.
- The number of Cancelled projects does not have seasonal behavior: it is almost constant during the year. However, the number of successful projects is very sensitive to the season: it is increasing in the 1st half of the year (from January till June), but after June it is significantly decreasing till December.

Limitations of this dataset:

- The considered data set includes 4,000 past projects (from 300,000) only, which does not fully represent the population on Kickstarter campaigns. With a larger set, the ratio between successful and unsuccessful projects can be different.
- To get the final conclusion about the most successful Category of projects, we need to compare Kickstarter results with data obtained from other crowdfunding platforms.
- Additionally, it would be important to have data about the real profit of every project during the lifetime. Unfortunately, this dataset does not have this type of information.

Other possible tables/graphs that could be created:

- It will be very interesting to see the geographical distribution of successful and unsuccessful projects to see the preferences of different countries. To do this the pivot table: Filter = state, Rows = Category or Sub-Category, Columns = country, Values = Count of state.

Bonus:

A new sheet (Bonus) was created. In this part I have calculated: (i) the amount of successful, failed, canceled, or live projects with respect to the goal within the ranges (**Fig.7, Bonus**); (ii) calculated the total

number of projects per goal category; and (iii) determined the percentage of successful, failed, canceled, or live projects with respect to the goal (**Fig.7**).

To count how many successful, failed, and canceled projects were created with goals within the ranges the following function Using the COUNTIFS() were applied (example of formulas for successful projects):

$$\text{Number Successful} = \text{COUNTIFS}(\text{Table!}\$D:\$D, "<1000", \text{Table!}\$F:\$F, "successful")$$

$$\text{Total Projects} = \text{Number Successful} + \text{Number Failed} + \text{Number Canceled}$$

$$\text{Percentage Successful} = \frac{\text{Number Successful}}{\text{Total Projects}} \%$$

	A	B	C	D	E	F	G	H
1	Goal	Number Successful	Number Failed	Number Canceled	Total Projects	Percentage Successful	Percentage Failed	Percentage Canceled
2	Less than 1000	322	113	18	453	71%	25%	4%
3	1000 to 4999	931	420	60	1411	66%	30%	4%
4	5000 to 9999	380	282	51	713	53%	40%	7%
5	10000 to 14999	168	144	40	352	48%	41%	11%
6	15000 to 19999	94	90	17	201	47%	45%	8%
7	20000 to 24999	62	72	14	148	42%	49%	9%
8	25000 to 29999	55	64	18	137	40%	47%	13%
9	30000 to 34999	32	37	13	82	39%	45%	16%
10	35000 to 39999	26	22	7	55	47%	40%	13%
11	40000 to 44999	21	16	6	43	49%	37%	14%
12	45000 to 49999	6	11	4	21	29%	52%	19%
13	Greater than or equal to 50000	86	258	100	444	19%	58%	23%

Fig. 7. Number and percentage of successful, failed, canceled, or live projects with respect to the type of goal.

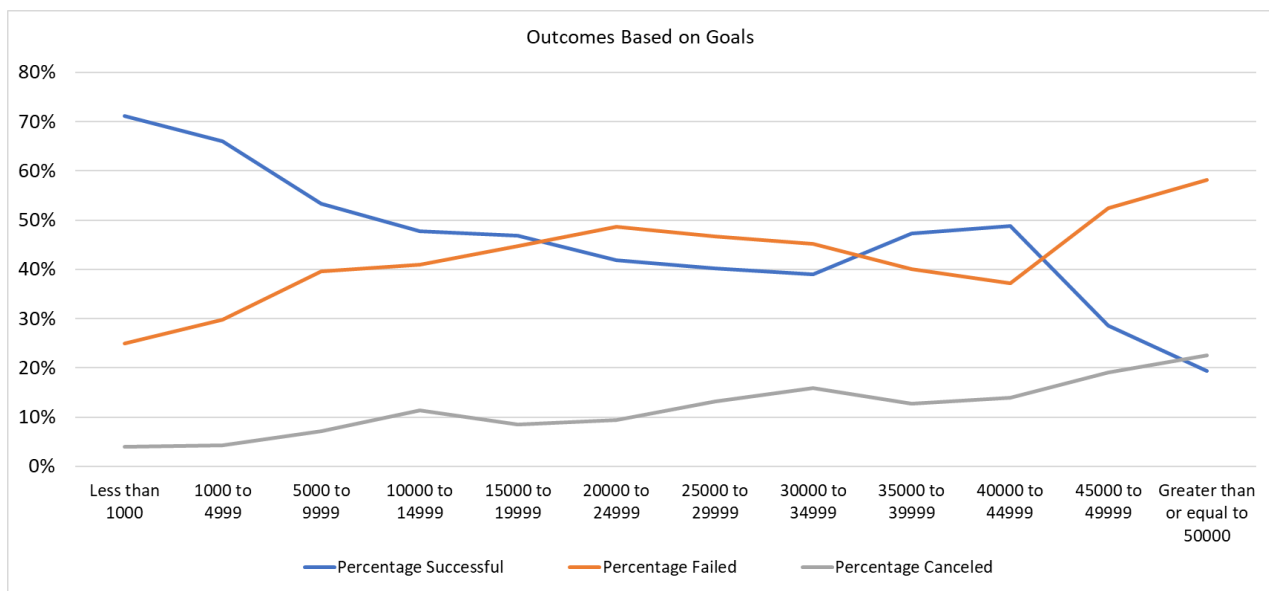


Fig. 8. The relationship between a goal's amount and its chances at success, failure, or cancellation.

Conclusion:

These results showed, that curves of "Percentage Failed " projects and "Percentage successful" projects are an approximate mirror image of each other: both, the higher percentage of successful projects and the

smallest percentage of the failed projects related to the smallest amount of requested money (goal), with increasing the goal the trend becomes opposite.

Bonus Statistical Analysis

In this part, the number of backers of successful and unsuccessful campaigns was evaluated by creating the summary statistics table (**Fig. 9** and **Fig.10**). A new worksheet (Bonus_Stat) in the workbook was created. Then a column each for the number of backers of successful campaigns and unsuccessful campaigns were generated (**Fig.9**).

	B	C	D	E	F
1	state	backers_count		state	backers_count
2	successful	182		failed	0
3	successful	79		failed	1
4	successful	35		failed	10
5	successful	150		failed	0
6	successful	284		failed	7
7	successful	47		failed	0
8	successful	58		failed	1
9	successful	57		failed	2
10	successful	12		failed	3
11	successful	20		failed	10

Fig.9. The number of backers of successful campaigns and unsuccessful campaigns.

These data were used to evaluate the following statistical data for successful campaigns, and then for unsuccessful campaigns (see **Fig.10**):

- The mean number of backers
- The median number of backers
- The minimum number of backers.
- The maximum number of backers.
- The variance of the number of backers.
- The standard deviation of the number of backers.

H	I	J
	Statistics for successful	Statistics for faild
The mean number of backers.	194	18
The median number of backers.	62	4
The minimum number of backers.	1	0
The maximum number of backers.	26457	1293
The variance of the number of backers.	712841	3773
The standard deviation of the number of backers	844	61

Fig. 10. The summary statistics table.

Conclusions:

- According to obtained statistical data, both data sets are very dispersed: a big difference between the minimum and maximum values. This affects the calculated data:
- For both sets the median summarizes data more meaningfully to the relevant value. Additionally, The high values of variance allowed to conclude that the mean is not that useful.
- There is more variability with successful campaigns: the variance of the number of backers is 712841, which is in 189 times higher than for unsuccessful campaign.
- Additional tests should be applied to the data sets to check which features don't add much value to the model and which are of importance.