MODULE 1 CHALLENGE

Crowdfunding Analysis with Excel

Abstract

The scope of this module challenge is to use Excel to uncover market trends by organizing and analyzing a database of 1000 sample projects from crowdfunding platforms like Kickstarter and Indiegogo. The data for this datasheet was generated by edX Boot Camps LLC for the purpose of this module challenge.

Daiana Spataru

June 14, 2023

Conclusion on the crowdfunding campaigns

Given the provided data, there are many conclusions that can be drawn about crowdfunding campaigns. Below are three conclusions.

Conclusion #1: Focusing on the data visualization from Category Stats, theater has the greatest number of Crowdfunding. It is even more evident when looking at the Sub-Category Stats graph. The months of June and July have the highest number of successful Crowdfunding. The US has the highest total number of Crowdfunding campaigns by far and heavily influences the analysis.

Conclusion #2: Excluding the Journalism category which comparatively has the lowest number of Crowdfunding, photography and technology have the lowest percentage of failed Crowdfunding.

Category	Failure Rate
film & video	34%
food	43%
games	48%
journalism	0%
music	38%
photography	26%
publishing	36%
technology	29%
theater	38%

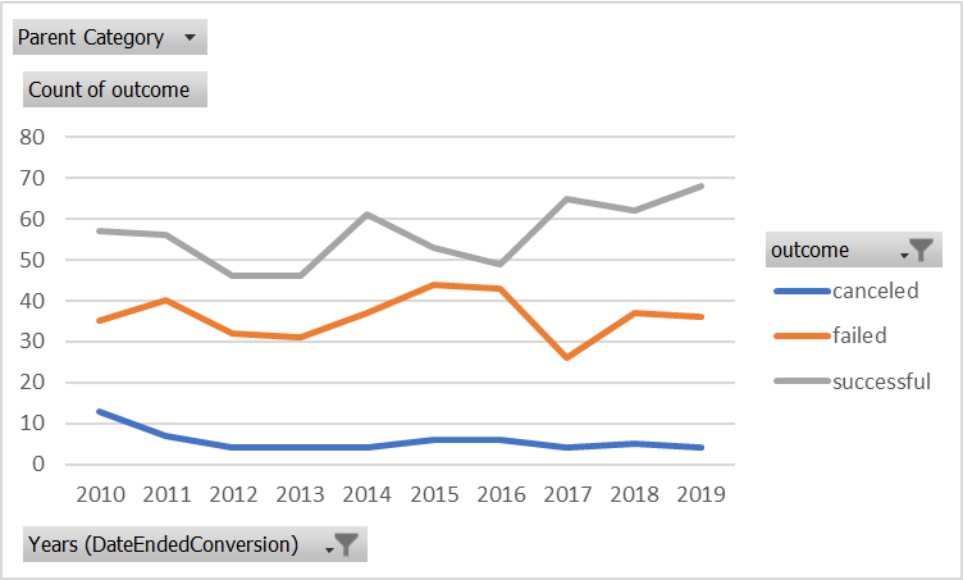
Again, excluding journalism, photography also had the lowest funding goal amongst all the parent categories which may be a factor for its low failure rate.

	Failed Crowdfunding		Successful Crowdfunding	
Parent Category	**Sum of goal**	**Count of outcome**	**Sum of goal**	**Count of outcome**
film & video	4547900	60	2612400	102
food	1155200	20	399600	22
games	1876300	23	373600	21
journalism			25700	4
music	3670200	66	2798000	99
photography	268900	11	374800	26
publishing	2108400	24	1074000	40
technology	1272000	28	1875300	64
theater	8908000	132	5385600	187
Grand Total	**23806900**	**364**	**14919000**	**565**

Comparing the difference between the Sum of goal column for successful and unsuccessful Crowdfunding across each category, the general trend shows that setting a

Conclusion #3: The graph below visualizes the data for the number of Crowdfunding that have succeeded, failed, and been canceled since 2010. The greatest number of canceled Crowdfunding occurred in 2010. The blue line in the graph below represents the number of canceled Crowdfunding.

Ignoring the data from 2020 as it only contains partial information from the early months of 2020, the number of successful Crowdfunding shows a general upward trend. In 2016, the number of successful and unsuccessful Crowdfunding were very similar. Up until 2016, the trendline of unsuccessful Crowdfunding generally matched the trendline for successful Crowdfunding; when the number of successful Crowdfunding increased, the number of unsuccessful Crowdfunding increased as well, and vice-versa. However, after 2016 the data shows the inverse relationship between the trendlines of successful and unsuccessful Crowdfunding. This could be related to the economic growth that was seen in 2017.



Limitations of the dataset

Granularity of data is lost. The success and failure of each Crowdfunding campaign is binary and does not provide details about the performance over the lifespan of the campaign. Additionally, a lot of this data contains outliers especially when it comes to the funding goals which makes it difficult to compare "apples to apples" in terms of success rate. Certain categories of campaigns may have more success because of their smaller funding goals. Furthermore, the US has the highest total number of Crowdfunding campaigns by far and heavily influences the analysis. Lastly, the blurb is quite generic and does not really provide any meaningful data about the purpose/goals of the fundraiser itself.

Other possible tables/graphs

There are various tables and graphs that could be constructed to help further analyze the dataset.

- As mentioned above, it would be interesting to exclude any significant outliers from the data in terms of funding goals or create a new category to sort the data into similar funding goals and compare the successful and unsuccessful campaigns in this manner (This is done in the next section of the activity).
- Additionally, we can calculate the duration of a campaign and see if it plays any role in its success.

- Creating box and whisker plots for each category with respect to the average donation, number of backers, length of a campaign, and percent funded, of both successful and unsuccessful campaigns could help reveal interesting patterns about the dataset.
- Excluding US from the analysis may provide different results.
- Lastly, it might be interesting to see if the Country of origin influences the success of campaigns. Some countries may be less likely to participate in Crowdfunding and could skew the data.

Crowdfunding Goal Analysis

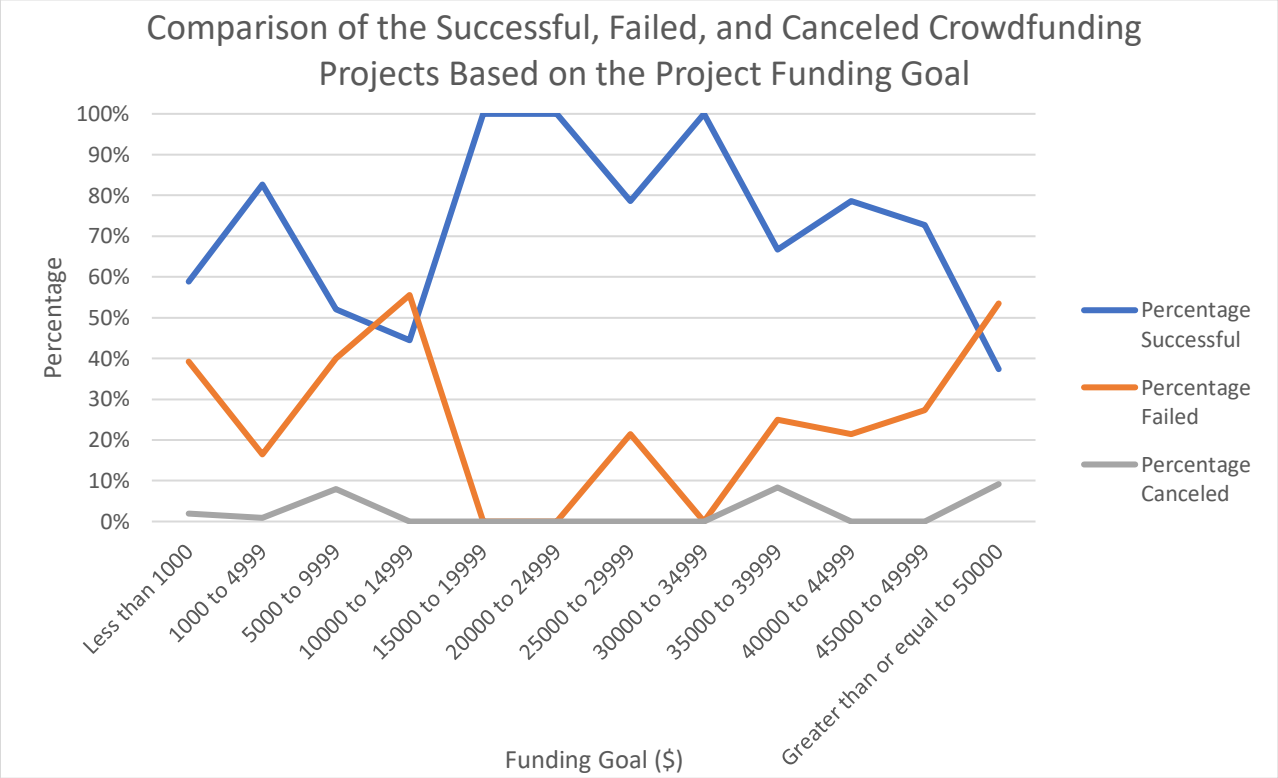
In this section, a new sheet with eight columns was created to group the outcomes of a campaign based on the goal. The data was split into 12 ranges and the following metrics were calculated using formulas in Excel.

Goal	Number Successful	Number Failed	Number Canceled	Total Projects	Percentage Successful	Percentage Failed	Percentage Canceled
Less than 1000	30	20	1	51	59%	39%	2%
1000 to 4999	191	38	2	231	83%	16%	1%
5000 to 9999	164	126	25	315	52%	40%	8%
10000 to 14999	4	5	0	9	44%	56%	0%
15000 to 19999	10	0	0	10	100%	0%	0%
20000 to 24999	7	0	0	7	100%	0%	0%
25000 to 29999	11	3	0	14	79%	21%	0%
30000 to 34999	7	0	0	7	100%	0%	0%
35000 to 39999	8	3	1	12	67%	25%	8%
40000 to 44999	11	3	0	14	79%	21%	0%
45000 to 49999	8	3	0	11	73%	27%	0%
Greater than or equal to 50000	114	163	28	305	37%	53%	9%

The COUNTIFS function in excel was used to calculate the number of successful, failed, and canceled Crowdfunding based on the 12 ranges. Below is an example calculation for a goal range of 5000-9999:

$$= COUNTIFS(Crowdfunding!\$G\$2:\$G\$1001, "successful", Crowdfunding!\$D\$2:\$D\$1001, ">4999", Crowdfunding!\$D\$2:\$D\$1001, "<10000")$$

Where the G column from the Crowdfunding sheet contains the data for the outcome of a campaign, and the D column from the Crowdfunding sheet contains the goal numerical value for each campaign. A line chart was created to visualize the relationship between a goal amount and its chances of success, failure, or cancellation.



Statistical Analysis

According to the information provided in the assignment, the number of campaign backers are often used to determine the success of a Crowdfunding campaign. A table of the backers_count for successful and unsuccessful campaigns was created to perform a summary statistics table including the mean, median, minimum, maximum, variance, and standard deviation of the number of backers.

The summaries are found in the following tables below:

Successful Campaigns	Result	Formula Used in Excel
Mean number of backers:	851.1469027	=AVERAGE(backers_count)
The median number of backers:	201	=MEDIAN(backers_count)
The minimum number of backers:	16	=MIN(backers_count)
The maximum number of backers:	7295	=MAX(backers_count)
The variance of the number of backers:	1603373.732	=VAR.P(backers_count)
The standard deviation of the number of backers:	1266.243947	=STDEV.P(backers_count)

Unsuccessful Campaigns	Result	Formula Used in Excel
Mean number of backers:	585.6153846	=AVERAGE(backers_count)
The median number of backers:	114.5	=MEDIAN(backers_count)
The minimum number of backers:	0	=MIN(backers_count)
The maximum number of backers:	6080	=MAX(backers_count)
The variance of the number of backers:	921574.6817	=VAR.P(backers_count)
The standard deviation of the number of backers:	959.9868133	=STDEV.P(backers_count)

The mean and median are measures of central tendency which are summary statistics that represent the center point of typical value of a dataset. The mean is typically better when the data follows a symmetric distribution. When the data is skewed, the median is more useful because the mean will be distorted by outliers.

For the successful campaigns, the median is 201 which is a big difference from the mean. The median being 201 means that 50% of number of backers for successful campaigns is <=200. The mean and median being drastically different from each other suggests that the data is skewed. In addition, because the max number of backers is 7295, this value is very far from the mean which suggests that above the median, 50% of the data falls under a larger range than the data below the median. As a result, the median should be used to represent the measure of central tendency for the successful campaigns.

For the unsuccessful campaigns, a similar conclusion can be drawn. The median is 114.5 which is a big difference from the mean. The median being 114.5 means that 50% of number of backers for successful campaigns is <=114.5. The mean and median being drastically different from each other suggests that the data is skewed. In addition, because the max number of backers is 6080, this value is very far from the mean which suggests that above the median, 50% of the data falls under a larger range than below the median. As a result, the median should be used to represent the measure of central tendency for the successful campaigns.

The standard deviation is the average amount of variability in your dataset. It tells you, on average, how far each score lies from the mean. The larger the standard deviation, the more variable the data set is. The calculation of the standard deviation for the successful and unsuccessful campaigns are 1266.24 and 959.99 respectively. Just by looking at the numbers, there is more variability in the successful campaigns than the unsuccessful campaigns. This could be explained by the fact that successful campaigns normally succeed by exceeding their goal which many not necessarily be directly related to the number of backers but rather, the amount each backer contributed which is not captured in these numbers. For example, a campaign with a target goal for 100,000 could have had only 5 backers that contributed 20,000 each. In this extreme case, the number of backs is quite low, but the campaign still succeeded. This could be one reason why the variability of the successful campaigns is higher. In contrast, it is more likely that a campaign which failed had a lower number of backers on average which is seen in the statistics summary table as well. However, it is important to remember that the data is skewed, and variability is harder to define.