UNIVERSITY OF TORONTO

MODULE – 1
CHALLENGE REPORT

DATA BOOTCAMP

Submitted by:

PARAMDEEP SINGH BIRDI

Date:

February 26, 2023

Instructor:

PIRO DHIMITRI

TABLE OF CONTENTS

EXECUTIVE SUMMARY

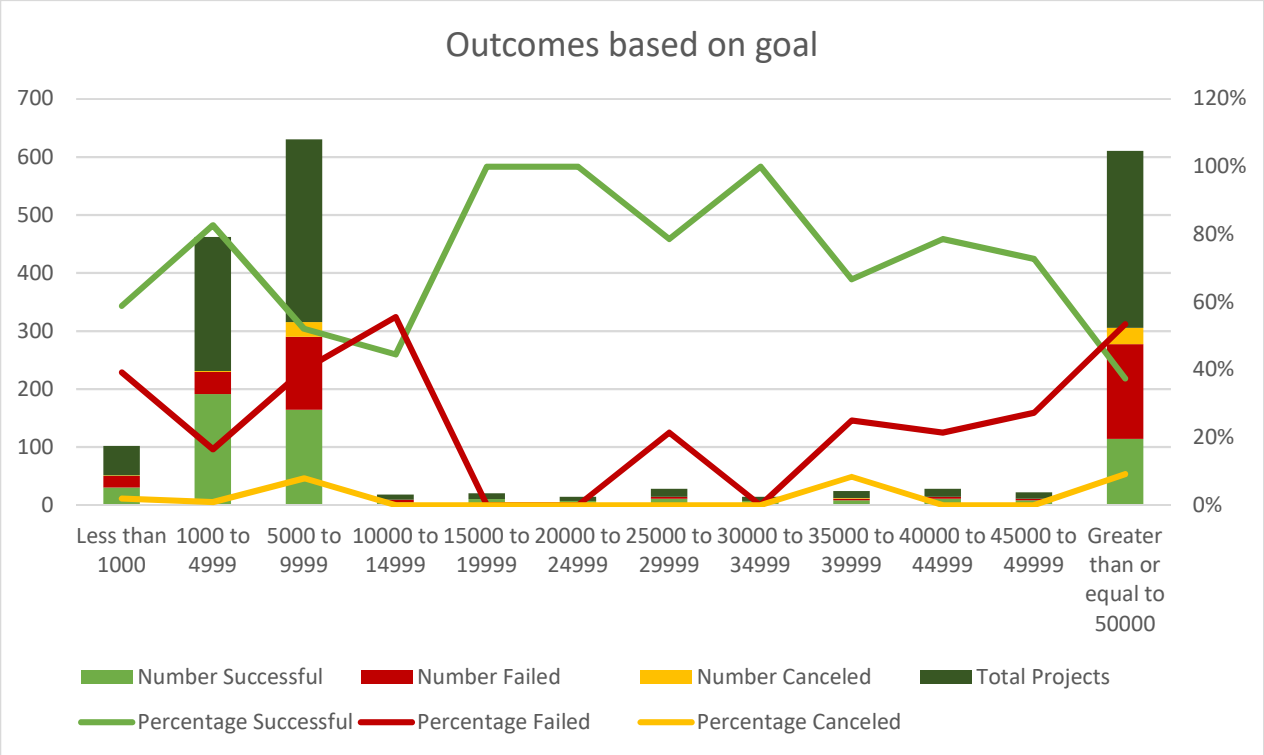
The executive summary describes the current popularity of crowdfunding platforms, such as Kickstarter and Indiegogo, and the importance of achieving the initial funding goal. Despite the increasing trend of using crowdfunding, not every project has found success. Therefore, to find the key to success, organizations analyze previous projects to uncover any hidden trends. In Module - 1 challenge, a database of 1,000 sample projects will be organized and analyzed to discover any patterns that could lead to success. Additionally, crowdfunding campaigns are often evaluated based on the number of backers they receive. In order to gain a deeper understanding of this metric, this report further analyzes the number of backers for successful and unsuccessful campaigns using a summary statistics table.

Based on the give dataset (not attached to this report), this report identifies trends and draw conclusions about the factors that contribute to a successful campaign. It aims to answer the following five questions:

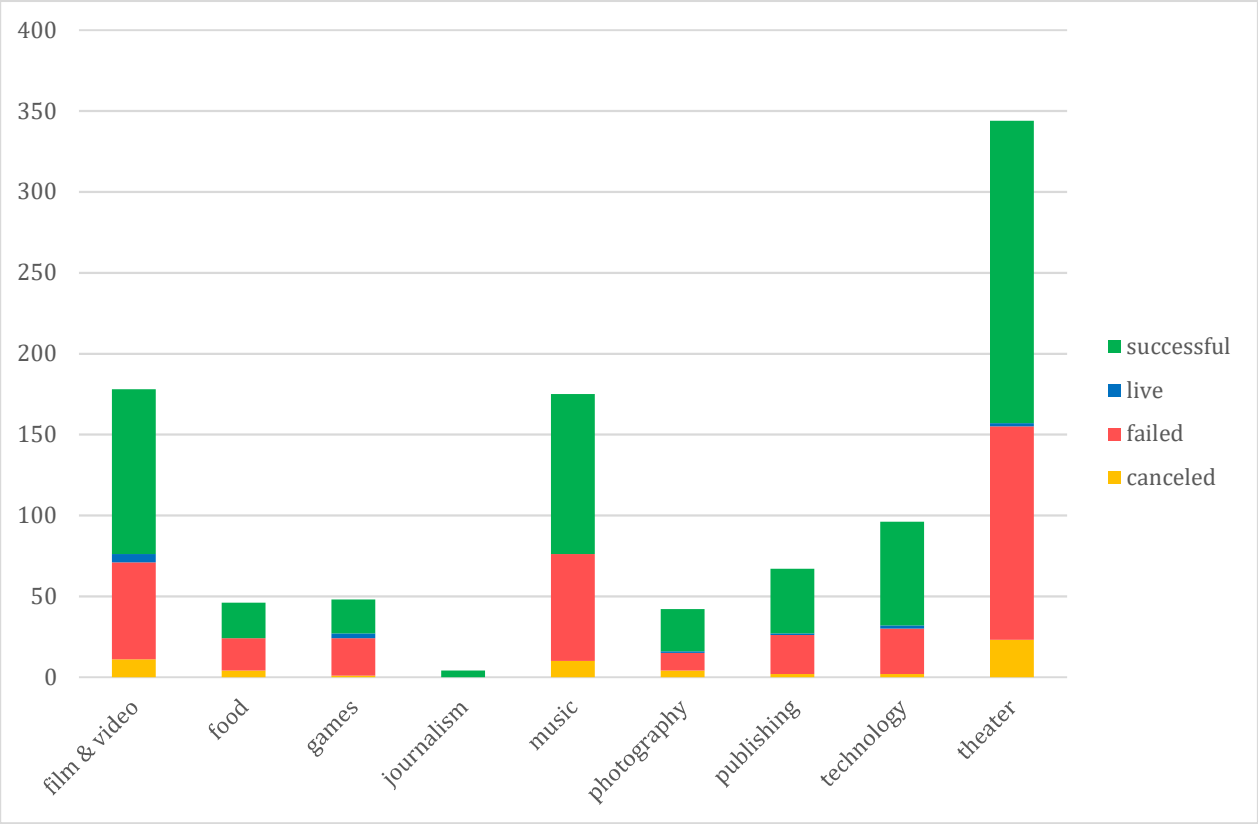
1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
2. What are some limitations of this dataset?
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
4. Using the given data to determine whether the mean or the median better summarizes the data.
5. Using the given data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

1. CONCLUSIONS

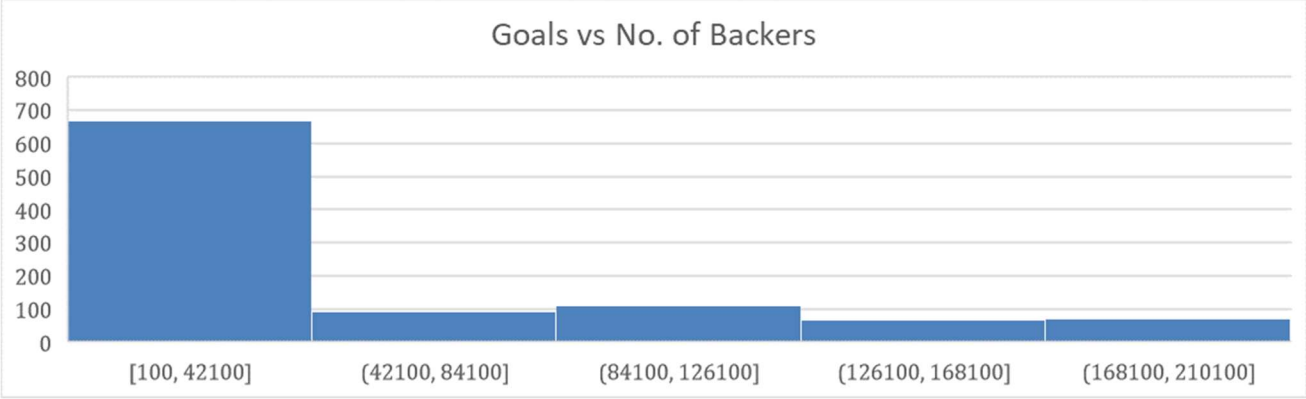
1. A higher funding goal does not always lead to more backers or greater success. As evident from the following chart, projects with lower funding goals often have higher success rates, indicating that setting realistic goals may be more effective.
   Projects with higher goal amounts (greater than or equal to $50,000) had the lowest success rates, with only 37% of these projects being successful. Projects with goal amounts between $10,000 and $49,999 had moderate success rates ranging from 52% to 79%. In contrast, projects with goal amounts less than $10,000 had the highest success rates, ranging from 59% to 100%.



2. The category of a project can greatly impact its success rate. It can be inferred from the following chart that the highest success rates were seen in the categories of journalism, technology and photography, with success rates of 100%, 67% and 62%, respectively. The highest failure rates were seen in the categories of games and food, with failure rates of 48% and 43%, respectively. This may indicate that these categories are more challenging or risky to fund. The largest number of projects overall were in the "theater" category, with 344 projects. The "film & video" and "music" categories were also relatively popular with 178 and 175 projects, respectively.



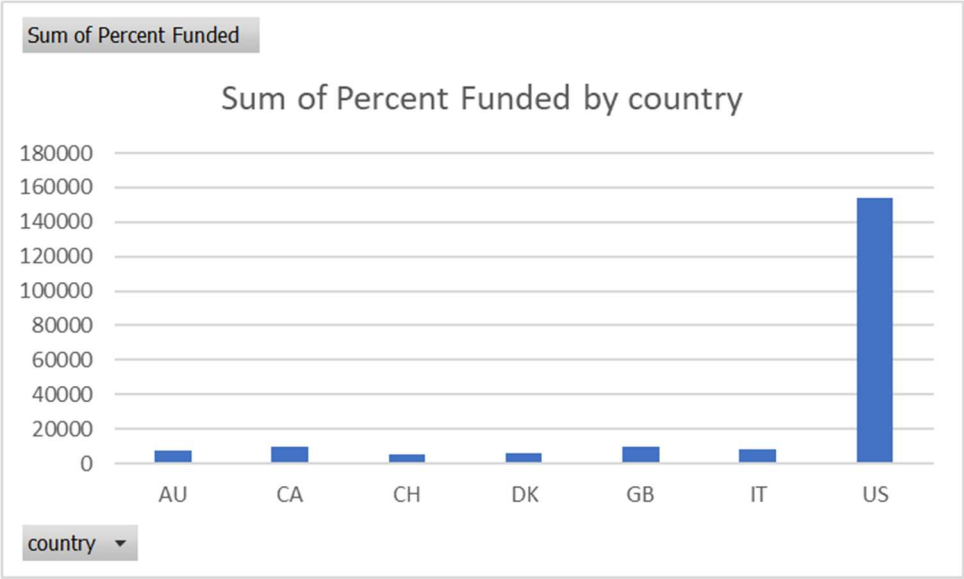
3. The number of backers for successful campaigns is, on average, higher than for unsuccessful campaigns, indicating that the support of a large number of people is crucial for success. As evident from the goal to backer's ratio and the following histogram, projects with higher funding goals tend to have fewer backers, suggesting that it may be more challenging to generate support for projects that require a significant amount of funding (Larger bin size as been used to accommodate the graph in this report).



4. Successful and failed campaigns have similar durations of their lives when compared to each other, indicating that the campaigns generated similar momentum before their verdict.

Descriptive Statistics	Failed	Successful
The mean difference of start and end dates of the project.	16	15
The median difference of start and end dates of the project.	12	12
The minimum difference of start and end dates of the project.	0	0
The maximum difference of start and end dates of the project.	59	58
The variance of the difference of start and end dates of the project.	193.18	168.79
The standard deviation of the difference of start and end dates of the project.	13.90	13.17

5. Most of the funding currency is attributed to US$ which indicates that the crowdfunding projects are most popular in the countries that accept US$ as their mainstream currency and receive major support of the backers in these countries. It would be beneficial to showcase one's project in one of these countries to achieve their goals.



2. LIMITATIONS

Limitations of this dataset include:

1. The data only includes a sample of 1,000 projects and may not be representative of all crowdfunding campaigns.

2. The data does not provide information on the quality of the projects or the marketing strategies used by the campaigns such as the mode and number of the campaigns, etc., which could impact their success.

3. The data does not include information on the demographics of the backers other than the currency of their donation. This could provide valuable insights such as the financial credibility of the backers.

3. ADDITIONAL TABLES/ CHARTS/ GRAPHS

The following visualizations could provide further insights into the factors that contribute to a successful crowdfunding campaign:

1. A histogram can be used to show the distribution of funding goals and the number of backers or funding pledged within each goal range, which could provide insights into the most common funding goals and how many campaigns are successful or unsuccessful within each range.

2. A scatter plot can be used to plot the relationship between the number of backers and the funding goal, which could help to identify any patterns or trends in the data.

3. If the reasons for failed projects are provided, a pareto chart can be used to identify the most frequent reasons as it provides a clear and visual representation of the data that can help to identify the key areas that require attention (80/20 rule may also be applicable).

4. MEAN OR MEDIAN

To determine whether the mean or median better summarizes the data, we need to look at the skewness of the data. Skewness is a measure of the asymmetry of a distribution. If the data is skewed to the left, the mean will be less than the median, and if it is skewed to the right, the mean will be greater than the median. If the data is symmetric, the mean and median will be the same.
Looking at the given data, the mean number of backers for successful campaigns is 851.15, while the median is 201. For unsuccessful campaigns, the mean is 585.62 and the median is 114.5. Based on this information, we can see that the median better summarizes the data for both successful and unsuccessful campaigns because the median is closer to the second quartile and represents the middle of the data.

5. VARIABILITY

The variance and standard deviation of the number of backers are both higher for successful campaigns than for unsuccessful ones. This suggests that there is more variability in the number of backers for successful campaigns, with some campaigns having very high numbers of backers and others having fewer backers. In contrast, unsuccessful campaigns tend to have a more consistent number of backers.

Descriptive Statistics	Successful	Unsuccessful
The mean number of backers	851.1469	585.6153846
The median number of backers	201	114.5
The minimum number of backers	16	0
The maximum number of backers	7295	6080
The variance of the number of backers	1606217	924113.455
The standard deviation of the number of backers	1267.366	961.3081998
First Quartile	127.5	38
Second Quartile	201	114.5
Third Quartile	1288.5	789.5
Inter-Quartile Range	1161	751.5
Outliers	1741.5	1127.25