|  |  |
| --- | --- |
| Module 1 Challenge  Analysis on Crowdfunding Campaigns | Abstract  The aim of this report is to uncover some hidden trends that will be helpful to be successful in fund raising via crowdfunding platforms.  Thelge Peiris  Data Analytics Bootcamp – Monash University |

Contents

[Answers to Question 1 – 3 2](#_Toc120207078)

[Q1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns? 2](#_Toc120207079)

[Q2. What are some limitations of Dataset? 3](#_Toc120207080)

[Q3. What are some other possible Tables and / or graphs that we could create, and what additional value would they provide? 3](#_Toc120207081)

[Bonus Statistical Analysis 5](#_Toc120207082)

[Statistical measurements of number of backers 5](#_Toc120207083)

[Successful Campaigns 5](#_Toc120207084)

[Failed Campaigns 6](#_Toc120207085)

[Use your data to deteremine whether the mean or the median better summarises the data. 7](#_Toc120207086)

[Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not? 8](#_Toc120207087)

[References 8](#_Toc120207088)

# Answers to Question 1 – 3

## Q1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

Below conclusions are made based on the 1,000 sample projects launched in 7 countries within 10 years from 2010 to 2019.

* **Conclusion 1:**

**The most common and most successful project category / Sub-category in crowdfunding platform is Theatre / Plays.**

It should be noted that Plays are the only sub-category under Theatre category. There are 344 Theatre / Plays projects out of 1,000 total projects. This represents a 34.4% of total projects. Out of these 344 Theatre / Plays, 187 projects were successful. which shows a 54.36% successful rate for Theatre / Plays projects in crowdfunding platforms. Further, Theatre / Plays projects contributes the largest portion of total successful projects which is 33.1% of 565 successful projects.

* **Conclusion 2:**

**June is the best month to launch a crowdfunding project.**

Projects launched in June has the highest successful rate and the lowest failed rate. July and September also indicate a project successful rate just below that rate of June and an approximately similar failure rate as June.

* **Conclusion 3:**

**There is a positive relationship between the project outcome and number of backers. Campaigns that have greater number of backers are more likely to be successful.**

The average number of backers of successful campaigns are greater than average number of backers of failed campaigns. Median number of backers of successful projects are also higher than that value of failed projects.

## Q2. What are some limitations of Dataset?

* Biasness

Even though there is project data for 7 different countries, a huge majority of campaigns listed in the dataset are launched in US. Therefore, it is hard to mention that it represents 7 countries. The dataset is bias and mostly representing US crowdfunding campaigns.

In an article published by IEEE, mentions that some datasets aimed to represent the visual world have actually become closed worlds unto themselves.

* Missing Data

There are few projects which were earned the 100% of goal, yet the outcome column says “Failed”. If this is not an entry error, that means there are more constrains in addition to Percent funded column that should be used to decide the outcome of the project. The other constrains are missing in the given dataset.

## Q3. What are some other possible Tables and / or graphs that we could create, and what additional value would they provide?

* It is possible to create pie charts to visually express which category is most common among all crowdfunding campaigns. Below chart 1 and chart2 shows that theater is the most common crowdfunding campaign among all categories and also it is the most successful project category.

Chart 1

Chart 2

* The relationship between number of days to complete and the outcome can be found by a scatter plot. As per chart 3 below, there is an exponential relationship between the number of days to complete and the number of successful projects. In other words, there are more successful projects completed in lower number of days and less number of projects were successful when the duration of completion is high. That means if a campaign can meet the target within short time period , the probability of become successful is high.

Chart 3

# Bonus Statistical Analysis

## Statistical measurements of number of backers

### Successful Campaigns

|  |  |
| --- | --- |
| **Successful Campaigns** |  |
| Mean number of backers | 851 |
| Median number of backers | 201 |
| Minimum number of backers | 16 |
| Maximum number of backers | 7295 |
| Variance of the number of backers | 1606217 |
| Standard Deviation of the number of backers | 1267 |

Table 1

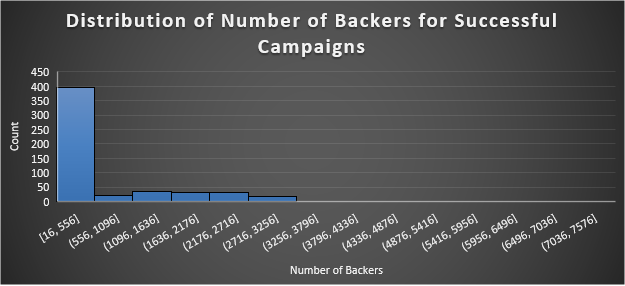


Chart 4

### Failed Campaigns

|  |  |
| --- | --- |
| **Failed Campaigns** |  |
| Mean number of backers | 586 |
| Median number of backers | 114.5 |
| Minimum number of backers | 0 |
| Maximum number of backers | 6080 |
| Variance of the number of backers | 924113 |
| Standard Deviation of the number of backers | 961 |

Table 2

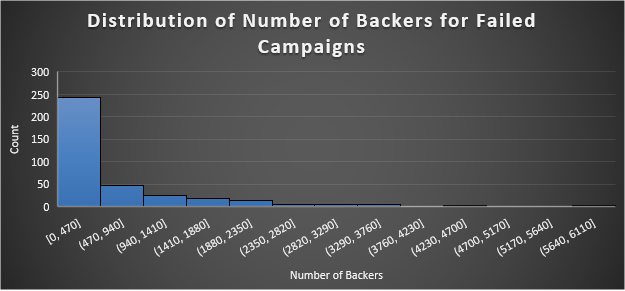


Chart 5

## Use your data to deteremine whether the mean or the median better summarises the data.

As per Table 1, for the successful campaigns, the minimum number of backers is only 16 while the maximum number of backers is 7,295. The middle of the distribution approximately lies around 3,600, but the mean is only 851 which is too far from the middle. The distance between mean and median also considerably high. The large values of variance and standard deviation proves that the mean is too far from a given point of the distribution. Same pattern can be seen in the figures in Table 2 for failed campaigns. These figures indicate that the distribution of number of backers in both types of campaigns are not normal.

Based on above chart 4 and chart 5 which graphically express the distributions, it is clear that the distribution of number of backers is a skewed one for both successful and failed projects. Even though the mean number of backers for successful projects is 851, the histogram in chart 4 shows that the large majority of successful projects have less than 556 backers. Similar to this the histogram in chart 5 indicates that the huge majority of failed projects have less than 470 backers despite the fact that mean figure for that group is 586. Therefore, it can be concluded that mean is not a good indicator to describe the distribution of backers.

In the article named “Reaction times and other skewed distributions” by Rousselet and Wilcox, it is mentioned that mean is a poor measurement of central tendency for skewed distributions and also it says median is a better alternative measurement for skewed distributions. We can prove this by chart 4 and chart 5. For the successful campaigns, the first column of chart 4 indicates that the majority of projects have backers from 16 to 556. The middle of this bracket is around 270, and the median of number of backers for successful campaigns is 201 as per Table 1. This implies that the median is a best indicator than the mean for the backers’ dataset. Similar to successful campaigns, the failed campaigns also showing a skewed distribution where the median is a better measurement than mean to describe the distribution.

## Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?



Table 3



Table 4

Based on Table 3 and Table 4 data, there is arount 20% of successful projects are out of mean + 2 stdv location and only 17% of failed projects are out of mean + 2 stdv location. This indicates that there is more variability with successful projects than failed projects. That makes sense because the standard deviation of successful projects is greater than that of failed projects.

# References

A. Torralba and A. A. Efros, "Unbiased look at dataset bias," CVPR 2011, 2011, pp. 1521-1528, doi: 10.1109/CVPR.2011.5995347.

Rousselet, G. A. & Wilcox, R. R. (2019). Reaction times and other skewed distributions: problems with the mean and the median. doi: https://doi.org/10.1101/383935