

Three Conclusions about Kickstarter campaigns

1. Among the 4114 projects and data collected, about half of the projects' crowdfunding are successful, and rest of the projects are either failed, canceled, or still live. This conclusion summarizes how successful the Kickstarter's crowdfunding campaigns are.
2. Among the 4114 crowdfunding projects, about one-third of the projects belong to the parent category of theater and about one-third of the projects belongs to the sub-category of plays inside of theater. This indicates that the Kickstarter's crowdfunding campaigns are quite limited to theaters and plays in this category, they are not very diverse.
3. The crowdfunding projects created in the summer, especially in May and June, are easier to be successful compared to the projects created in November and December. For the projects that failed, they are the least likely to happen in February, March and April. This indicated that there tends to be more projects in the summers compared to in the winters. However, since the number of projects is not constant among the years, we cannot simply draw this conclusion based on this one line graph. We can use percentage to verify the validity of this conclusion, which I will expand in the future work.

Limitations and Assumptions of Dataset

1. This dataset contains only 4114 projects among over 300,000 crowdfunding projects that are launched only at Kickstarter. However, due to the limitations of excel on handling bigger data, we assume that this is the population instead of the sample.
2. This dataset contains only projects that are launched or created between May 2009 and March 2017, only launched at Kickstarter and only represent projects from 22 selected countries.
3. We assume that there are no human errors consisting in this dataset such as missing data or typing error and no outliers that can affect our analysis and conclusions.
4. This dataset only evaluates the trend for success based on the projects' crowdfunding amounts, backer's counts, creation and ending dates, country, and categories. Other possible factors that might affect the success of projects such as the government policy, culture and so on are not included in this data analytics.

Future Work

1. We can calculate P-value, t-test, and hypothesis test to verify the validity of this dataset.
2. Create a pivot chart line graph to see how the numbers of successful projects are changed from 2009 to 2017 to test which year have the maximum number of successful projects.
3. The theaters parent category and the plays sub-category have the greatest number of successful projects and the greatest number of failed projects. Therefore, we cannot conclude that theaters and plays are two of the categories and sub-categories that have the greatest numbers of successful projects. We need to create another pivot table and pivot chart to evaluate the relationships between the percentages of successful projects over all projects and the categories to further test this.
4. This dataset also includes information regarding the backer's number and the average donations in this dataset. We can create a pivot table and chart to test the relationships between the average donations and successful projects.
5. We can also calculate Z-score for backer's count to see if we can find any trends.

6. We can also use IQR to verify if this dataset contains any outliers
7. We can create a pivot table and chart to test the relationship between the country that project launched and number of successful projects.
8. As mentioned in my conclusion, we can do another graph to verify the relationship between month and number of successful projects by using the percentage of successful projects instead of number of successful projects.

Bonus Statistical Analysis

After the calculation, the mean or average of backer's counts for all the successful projects is 194, and the median is 62. For the unsuccessful projects, the mean is 18 and the median is 4. Median summarizes the data more meaningfully compared to mean in this scenario. After I sorted the backer's count for both successful and unsuccessful projects, I found out that roughly 80% of the data falls in the category of 0 to the mean. This shows that if we use mean instead of median to summarize the data, it will not best represent the data's central tendency. Furthermore, this data type shows that it is ordinary, and it does not fall into the normal distribution. For this type of situation, we use median to represent the central tendency of the data.

Based on the calculation, the variance and standard deviation of successful campaigns are much higher than that of the unsuccessful campaigns. For the successful campaigns, majority of the backer's count falls in the range of 0 to 8000, and there are 2 backer's count that fall into the range of 20000 and 30000. For the unsuccessful campaign, all the data falls into the range of 0 to 1300. Higher variance and higher standard deviation imply bigger variability. Therefore, this result aligns with the data. Further down, the backer's count and the number of successful projects contains a correlation. The successful campaigns tend to have much larger backer's count compared to unsuccessful count, and the unsuccessful campaigns tends have a smaller range of backer's count. Since the variability shows how spread out a set of data is, it implies that successful campaigns have more variability than unsuccessful campaigns.