09/23/21

Stephen Lane

UTSA Data Analytics Boot Camp

Excel-Challenge

**Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**

Analysis of the kickstart campaign data revealed the theater represents 34% of all Kickstart campaigns run. The Theater category had an overall success rate of 60%. The second most popular category was Music campaigns which represented 34% of all campaigns run. Music was also the most likely to campaign to succeed with an overall success rate of 77%. The percentage of successful and failed campaigns have an inverse relationship; this is demonstrated as the dollar amount required for the campaign rises the success rates fall while the failure rate rise. The most successful months in which campaigns experience success occurs April through July with the peak of successful project occurring in May.

Food and Journalism were the least successful projects, food campaigns experienced a 70% failure rate and Journalism has not had a campaign that hasn’t required cancelation.

Based on the data analysis Kickstart campaigns are ideal for entertainment project which do not require large sums of capital to start. Projects that require large amounts of funding, experience higher failure rates.

**What are some limitations of this dataset?**

The data is skewed with 74% of all campaigns coming from the United States. It would be challenging if not impossible to draw valid conclusions from underrepresented countries.

Limitations of the data include lack of data definitions, so it’s difficult to draw additional conclusions using some of the columns.

The data is not normalized meaning in the data you have different languages and currencies which have not been converted to single currency. It becomes difficult to compare the amount truly needed without normalizing this data. Additionally, multiple regions are represented in the analysis without any statistical adjustments being made for population and cultural differences.

The data lacks the detail needed to analyze the campaign backers. In a true statistical/business analysis you would want to analyze the demographics of the backers. With this data your can gain information of location, age, income of the backers. Also, without the information on how much each backer contributed, you cannot determine if a large contribution from a single backer contributed to the project success or whether small contributions from many backers contributed to the project’s success.

**What are some other possible tables and/or graphs that we could create?**

* Pivot and graph of Parent Category and Funding Goal to analyze the differences in funding requirements.
* Pivot of graph Country and Category to identify which country each category was popular in. This could reveal some cultural differences/preferences.
* Pivot and graph of Percent Funded and Parent Category with a filter for Country.

**Bonus**

**Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

Mean would best represent the center of the data. While the mean is influenced by outliers there are only a few outliers which with the amount of data present do not significantly impact the mean. If graphed on scatter plot you will observe the mean is closer to the center of the data than the median.

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

There is more variation in the successful projects than the failed projects. I would expect there to be many variations in the number of backers of successful project. Less variation in failed project could be reasonably expected due to the fact that the project failure is linked to the lack of backers.