StarterBook- Data Analysis

Thanu Thavasi Perumal

StarterBook Data Analysis project analyses a snippet of a dataset from an online crowdfunding platform (Kickstarter), Kickstarter helps people to attract investors towards their ideas to get funded and meantime building a community of fans to support their ideas. But still besides all their positive efforts the campaigns still fail or sometimes the projects get cancelled.

The Objective of this project is to carry out an exploratory analysis on the dataset by analyzing across different set of variables and their observations to predict the possibilities of success and failure rates that could help people to achieve their goal.

This Project analyses a dataset that contains data for over 4100 campaigns and their important attributes like Project Name, Funding Goal, Pledged amount, State of the project, Start and end Date, Category and sub-category etc.

Description of the Attributes:

* **State**-Successful, Failed, Cancelled, Live
* **Category**-Various Categories of the project like Television, Theater, Music
* **Year**-Project’s Life Span, this explain the launched date and end date
* **Blurb**-A short description about the project
* **Name** -Name of the project

Methodology:

1.Analysis of State Based on Parent Category.

2.Analysis of State Based on Sub Category.

3. Analysis of State based on the months of a Year.

1.Analysis of State Based on Parent Category:

This analysis explains how the ‘Category’ the projects belong could make a great difference on the state of outcomes (Successful, Failed or Canceled). We also can filter this analysis report for each Countries. It’s quite obvious from this analysis that Theater projects have a highest successful outcome and at the same time the Number of failure are also quite high. Total number of Projects comes under this category can also have a high impact on this result. But the Category Music has a second highest success rate with the much lower possibility to go fail.

2. Analysis of State Based on Sub Category:

By running the same analysis over the State counts and Sub categories we get the following result. This result rhetorically depicts that the sub category “Plays” will the all-time winner and without doubt it steals the show here. And this ‘Play’ category comes under the Parent category of “Theater” as well .But we can’t neglect those sub-categories like Classical Music ,Documentary, Electronic music, hardware, nonfiction, rock , shorts , laptop games and televisions that produces 100% success rate .These two analysis gives us an approximate idea about how the categories can have a influence on the State of outcome but the actual attribute that indirectly drives this result could be the number of projects that falls under each categories .

3. Analysis of State based on the months of a Year.

This Year Vs Count of State analysis helps us understand how starting the projects at various months of an Year plays a crucial role on the success and failure rates. From this analysis its inferred that Projects that starts at the beginning of the second quartile comparatively have a huge Success Rate. And also the Failure Rate closely follows the success rate strategy as the two results have the similar shape on their outcomes most of the times in a year between March and October. At the month of February Success Rate is quite high when compared to the failure Rate that can be because of people’s energy level and motivation that usually comes at the beginning of a new year .Cancelled Projects have a very little dependency with the months of an year because its quite a straight line that shows the Months of an Year have no impact on the “Cancelled” outcome. However during July and November more number of projects get cancelled that could be because of the Patriotic and holiday mood of the people that can have a little bit of impact on the results.

Limitations Of this Dataset:

Though this dataset contains many important variables and their observations to predict the outcome of the campaign there are still some limitations that restricts us from getting a better picture from this prediction analysis. It would be little more beneficial if this data set has the **“Days of the Week”** variable because that could have a high impact on the outcome as well. Also it would be much more informative if we have the variable named **“Cause”** for the failed or cancelled Projects. These are a few of the many limitations but still the **limitation** is a subjective term that depends on what exactly we are looking for from this Dataset.

Other Possible Tables/Graphs:

We also can analyze this dataset with different variables,like blurb(the short description about the project ),the Length of the blurb and the possible outcomes and the Length of the Names of the projects for the successful and Failed campaigns and their distributions can also help us to achieve more accurate predictions . This graph explains that Legth of the Blurbs also can have a positive impact on the outcomes.

Thank You!

Thanu Thavasi Perumal