Austin Wen

1. Excel Homework

03/15/2020

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

**Conclusion 1**: Theater was the most popular category from this sample dataset, and journalism was the least popular category.

**Conclusion 2**: On average, the projects had 58.8% success rate\*. But the success rate varied from year to year. Year 2011 had the highest success rate at 83%, and 2015 had the lowest success rate at 52%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year\State** | **Successful** | **Failed** | **Total** | **Success Rate** |
| **2009** | 9 | 4 | 13 | 69% |
| **2010** | 49 | 15 | 64 | 77% |
| **2011** | 136 | 28 | 164 | 83% |
| **2012** | 216 | 60 | 276 | 78% |
| **2013** | 200 | 67 | 267 | 75% |
| **2014** | 474 | 422 | 896 | 53% |
| **2015** | 567 | 527 | 1094 | 52% |
| **2016** | 475 | 376 | 851 | 56% |
| **2017** | 59 | 31 | 90 | 66% |
| **Total Project** | 2185 | 1530 | 3715 | 59% |

\*We defined the success rate as total number of successful projects divided by the sum of total number of successful and failed projects.

**Conclusion 3**: Kickstarter projects with the Staff Pick designation had higher success rate than the projects without the Staff Pick designation.

In Table 1 below, it shows that projects with the Staff Pick designation (True) had 88% success rate, comparing to 54% success rate for those projects without the Staff Pick designation (False).

**Table 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Staff Pick/State | Successful | Failed | Total | Success Rate |
| TRUE | 486 | 66 | 552 | 88% |
| FALSE | 1699 | 1464 | 3163 | 54% |
| Total | 2185\* | 1530 | 3715 |  |
| Overall Average | 58.8% | 41.2% |  |  |

We then performed the Chi-square test of homogeneity to check whether the difference in success rate is statistically significant at 5%. Our hypotheses are as follow:

*Ho: Equal success rate with or without Staff Pick designation.*

*Ha: Not equal success rate between the Staff Pick projects and those projects without.*

If the two groups have the same expected success rate as overall average. The expected numbers of success and fail are shown in Table 2. And the Chi-Square calculation is shown in Table 3.

**Table 2**

|  |  |  |  |
| --- | --- | --- | --- |
| Expected Value | Staff Pick/State | Successful | Failed |
| TRUE | 324.66 | 227.34 |
| FALSE | 1860.34 | 1302.66 |

**Table 3**

|  |  |  |  |
| --- | --- | --- | --- |
| Chi-Square | Staff Pick/State | Successful | Failed |
| TRUE | 80.18 | 114.50 |
| FALSE | 13.99 | 19.98 |

The sum of Chi-Square is 228.65 with a degree of freedom of 1. This yields a p-value less than 0.1%; therefore, we can reject the Ho hypothesis. It is statistically significant in success rate difference between projects with staff pick designation and projects without

1. What are some limitations of this dataset?

First, I noticed this sample dataset includes goal and pledged amounts in 13 different currencies, and we do not have any foreign exchange rate data available. Without properly convert the goal and pledge amounts to a base currency, comparison between projects or categories would be misleading.

Second, we were not provided with any detailed definition what each column stood for. For example, I noticed all successful projects’ “spotlight” column were TRUE, but I do not know the causation between column “state” and column “spotlight”. Similarly, having the “staff pick” designation did correlate with higher success rate, but it is impossible to conclude the “staff pick” designation caused the project to succeed.

1. What are some other possible tables and/or graphs that we could create?
   1. Success Rate by Category Bar Chart
   2. Count of State by Years with Category filter Pivot Table and Chart

A screenshot of a cell phone

Description automatically generated

A screenshot of a map

Description automatically generated

* 1. Number of projects per country and per year

A screenshot of a social media post

Description automatically generated