# **Kickstart Dataset Assignment**

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

* After analysing the dataset, it can be observed that about 53% of the total campaigns are successful in comparison to 37% failed. Successful campaigns have shorter durations (as shown in Table 1), a lower funding goals (Table 1) and a higher number of backers than failed projects.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | **successful** | **failed** | **canceled** | **live** | **Grand Total** |
| Campaign | 2185 | 1530 | 349 | 50 | 4114 |
| Proportion (%) | 53% | 37% | 9% | 1% | 100.00% |
| Average of Duration (Day) | 32 | 35 | 36 | 35 | 34 |
| Average of goal ($) | 9867 | 60556 | 517985 | 19381 | 71939 |
| Average pledge ($) | 18579 | 1856 | 7286 | 3922 | 11224 |
| Average of backers\_count | 194 | 18 | 27 | 36 | 113 |
| backers\_count | 424819 | 27096 | 9530 | 1801 | 463246 |
| % backers\_count | 92.0% | 6% | 2% | 0.4% | 100.00% |
|  |  |  |  |  |  |

Table : Summary of campaign

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| **Row Labels** | **successful** | **failed** | **canceled** | **live** | **Grand Total** |  |
| film & video | 300 | 180 | 40 |  | 520 |  |
| food | 34 | 140 | 20 | 6 | 200 |  |
| games | 80 | 140 |  |  | 220 |  |
| journalism |  |  | 24 |  | 24 |  |
| music | 540 | 120 | 20 | 20 | 700 |  |
| photography | 103 | 117 |  |  | 220 |  |
| publishing | 80 | 127 | 30 |  | 237 |  |
| technology | 209 | 213 | 178 |  | 600 |  |
| theater | 839 | 493 | 37 | 24 | 1393 |  |
| **Grand Total** | **2185** | **1530** | **349** | **50** | **4114** |  |
|  |  |  |  |  |  |  |

Table : Summary of campaign per category

A screenshot of a cell phone

Description automatically generated

Figure : Percentage of campaign per category

A screenshot of a video game

Description automatically generated

Figure : State by category

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **Successful** | **Failed** | **Canceled** | **Live** |
| film & video | 57.69% | 34.62% | 7.69% | 0.00% |
| food | 17.00% | 70.00% | 10.00% | 3.00% |
| games | 36.36% | 63.64% | 0.00% | 0.00% |
| journalism | 0.00% | 0.00% | 100.00% | 0.00% |
| music | 77.14% | 17.14% | 2.86% | 2.86% |
| photography | 46.82% | 53.18% | 0.00% | 0.00% |
| publishing | 33.76% | 53.59% | 12.66% | 0.00% |
| technology | 34.83% | 35.50% | 29.67% | 0.00% |
| theater | 60.23% | 35.39% | 2.66% | 1.72% |

Table : Summary category VS State

The average successful campaign has a goal 6 times less than failed campaign as well as average pledged amount of about 10 times more than failed projects. Successful campaigns have on average 10 times more backers than failed campaign. It means that the more people back a campaign, the more likely it’s going to succeed.

Further analysis was carried out on category to uncover any trends. I analyzed project success rate according to each category. Table 3 shows that Music, Theater and film and video seems to achieve 77%, 60%, and 58% respectively which is greater than the average success rate of all campaigns (53%). All Journalism campaign were canceled, and food were least successful. Game Technology and Publishing have a relatively similar success rate.

* In the sub-categories, Rock has more successful campaigns in terms of total counts. From figure 3 below, it is apparent that most campaign are in the Play sub-category and it has most of the successful and failed projects (based on total counts per each sub-category);

A picture containing sitting, person, parked

Description automatically generated

Figure : State per sub-category

* Of all the campaigns so far, there are more successful campaigns in May (254) followed by June (211). Canceled campaigns are fairly consistent over the months less than 50.

A close up of a map

Description automatically generated

Figure : State per month

# What are some limitations of this dataset?

* The Dataset contains limited number of sub-categories. If the category is broken down further, if may be more helpful to spot the state of a specific sub-categories. For example, the plays sub-category has the most successful and failed rate. It’s difficult to tell what kind of plays yield the most successful counts and which doesn’t.
* The dataset has regional limitation: The United states accounts for most of the campaign as shown in figure 5. It might not be a good representation of the success rate in general.

# A screenshot of a computer Description automatically generated

Figure : State Vs Country

# 3. What are some other possible tables and/or graphs that we could create?

* Goal Amounts per state
* State per Country
* Backers per State
* Pledged amounts by Category and Sub-Category
* A Scatter Plot between Goal Amount and Pledge Amount

# Bonus Questions

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

The mean summarizes the data more meaningfully because it takes into account all the values whereas the median lacks the representatives of data as it only selects the positional middle value.

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

There’s more variability with successful campaigns which make sense because investors want to have a stake in interesting project even with a small percentage. In this case retail investors will invest a relatively small amount as well. Some other projects are completed bought out by fewer investors.