**Excel Homework – Kickstart my Chart**

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

**Conclusion 1:** Based on the dataset provided, more than half (About 53%) of the Kickstarter Campaigns have been successful over the last 9years, 37% of the campaigns have failed, 8% of the campaigns are cancelled and about 1% are currently live.

|  |  |
| --- | --- |
| Total Projects | 4114 |
| Total number of successful projects | 2185 |
| Total number of failed projects | 1530 |
| Total number of canceled projects | 349 |
| Total number of live projects | 50 |
| % of successful projetcs | 53% |
| % of failed projects | 37% |
| % of canceled projects | 8% |
| % of live projects | 1% |

**Conclusion 2**: About 34% of the total campaigns belong to the Category “Theatre” and 26% of the total campaigns belong to the Category “Theatre” and Sub-Category “Plays”. Based on the dataset provided, this concludes that more of the entertainment/Theatre domain projects need Crowdfunding.

|  |  |
| --- | --- |
| Highest Number of Campaings | 1394 |
| % of the Highest number of Campaigns based on Category | 34% |
| % of the Campaigns from Sub-Category Plays | 26% |

**Conclusion 3:** Based on the dataset provided, over a period of 9yrs, There have been more number of Campaigns created in the Month of May.

1. What are some limitations of this dataset?

**Limitation 1:** Kickstarter campaigns are divided into lot of Categories and Sub-Categories and for most of these, the dataset is not large enough to come to conclusive information or results.

**Limitation 2:** This dataset does not include/consider the location of the projects. It would help to understand the statistical conclusions based on the locations.

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

We could create table for the Campaigns based on different countries and pledged amount, Campaigns from different countries Vs statuses. Also can create Bar graphs, Box Plots and so on.

For Example:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Count of state** | **Column Labels** |  |  |  |  |  |
| **Row Labels** | **canceled** | **failed** | **live** | **successful** | **(blank)** | **Grand Total** |
| AT |  | 2 | 1 | 3 |  | 6 |
| AU | 14 | 41 |  | 19 |  | 74 |
| BE |  | 1 |  | 1 |  | 2 |
| CA | 17 | 64 | 1 | 64 |  | 146 |
| CH | 3 | 2 |  | 1 |  | 6 |
| DE | 3 | 27 |  | 23 |  | 53 |
| DK | 4 | 6 |  | 4 |  | 14 |
| ES | 1 | 9 |  | 11 |  | 21 |
| FR | 5 | 10 | 2 | 10 |  | 27 |
| GB | 25 | 205 | 8 | 366 |  | 604 |
| HK | 1 | 2 |  |  |  | 3 |
| IE | 2 | 4 | 1 | 8 |  | 15 |
| IT | 3 | 19 |  | 7 |  | 29 |
| LU |  |  |  | 2 |  | 2 |
| MX | 1 | 8 | 3 |  |  | 12 |
| NL | 4 | 14 | 1 | 2 |  | 21 |
| NO |  | 5 |  | 2 |  | 7 |
| NZ | 4 | 5 |  | 3 |  | 12 |
| SE | 5 | 9 |  | 7 |  | 21 |
| SG |  |  |  | 1 |  | 1 |
| US | 257 | 1097 | 33 | 1651 |  | 3038 |
| (blank) |  |  |  |  |  |  |
| **Grand Total** | **349** | **1530** | **50** | **2185** |  | **4114** |

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

Median summarizes the data more meaningfully mainly because most of the values come under one category with few outliers that are significantly huge and when considered to calculate the mean would not give the best results.

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 variability with successful campaigns because about 53% of the Campaigns are in Successful state, this means that the data set is larger for successful campaigns than that of the other ones.