Name: Setordji Yao Abotsi

Assignment: Week 1 Homework

Date Due: 12/18/2021

**DATA ANALYSIS REPORT**

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

Answer: The data was analyzed in three different ways.

In the first analysis, I looked at all the different campaigns that were proposed in different categories:

* Successful projects – these are projects that that got successfully executed
* Failed projects – these are projects that got executed but were unsuccessful
* Live projects – these are projects that are currently still alive and being executed
* Canceled projects – these are projects that got canceled for reasons of either no funding or others.

Looking at the data below, there was a lot of projects under the theater category but only 839 were successfully executed. Music category was the second category with most projects and had a total of 540 successful projects. Although the technology category had the third most projects, only 209 projects were successfully executed and had the most canceled projects at 178 total. Film and video category took the third position with the most successfully executed projects. This data visualization however does although shows the categories with the most projects, it also displays those categories with higher percentage of successful project.

In the second analysis, the data visualization below shows the different sub-categories under which projects were represented. That data here is somehow skewed with the higher volume of projects under the plays sub-category. The graph also clearly shows those sub-categories that very successful with none of the projects either canceled or failed or even live.

In the third analysis, the last graph or representation is a timeline trend of the number of projects there were strictly successful, failed or canceled. It is clear from the graph that the month of May saw most successful projects does not state why May was such a successful month. Although July and November had the highest number of projects cancelation, project cancelation remained almost consistent on an average from month to month.

**Q.2.** What are some limitations of this dataset?

Answer:

Some limitations with the data visualization above may be the skewness of the overall data with certain categories being outliers. The dataset does not however show a true comparison of all the different categories performances with the assumption that they all have the same number of projects. A pie chart may probably show a good comparison of the percentages between those projects that were successful, failed, canceled or are still alive.

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

Answer:

We can certainly employ a pie chart for the data visualization as well. This should show all the different projects categories and their individual performances. Another way to look at the data to make a good inference on where to invest will be executing or running some statistical analysis to see what project could have the most return on our investment. For sure not every successful project will have the same return on investment, but the decision can be made by reviewing the statical inferences that should be made from the analysis.

**BONUS**

There are 2 parts to the bonus assignment.

The first part is to successfully create a table that populates the number of successful, failed and canceled campaigns based on the goal ranges provided and then calculate their respective percentages. The below table displays the total number of successful, failed and canceled campaigns as well as their respective percentages for each goal range.

This table is shown in homework spreadsheet under the “Bonus” tab.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Goal** | **Number Successful** | **Number Failed** | **Number Canceled** | **Total Projects** | **Percentage Successful** | **Percentage Failed** | **Percentage Canceled** |
| Less than 1000 | 322 | 113 | 18 | 453 | 71% | 25% | 4% |
| 1000 to 4999 | 932 | 420 | 60 | 1412 | 66% | 30% | 4% |
| 5000 to 9999 | 381 | 283 | 52 | 716 | 53% | 40% | 7% |
| 10000 to 14999 | 168 | 144 | 40 | 352 | 48% | 41% | 11% |
| 15000 to 19999 | 94 | 90 | 17 | 201 | 47% | 45% | 8% |
| 20000 to 24999 | 62 | 72 | 14 | 148 | 42% | 49% | 9% |
| 25000 to 29999 | 55 | 64 | 18 | 137 | 40% | 47% | 13% |
| 30000 to 34999 | 32 | 37 | 13 | 82 | 39% | 45% | 16% |
| 35000 to 39999 | 26 | 22 | 7 | 55 | 47% | 40% | 13% |
| 40000 to 44999 | 21 | 16 | 6 | 43 | 49% | 37% | 14% |
| 45000 to 49999 | 6 | 11 | 4 | 21 | 29% | 52% | 19% |
| Greater than or equal to 50000 | 86 | 258 | 100 | 444 | 19% | 58% | 23% |

Based on the data, in the table above, the percentage outcome for each goal range is represented in the line graph below.

This table is shown in homework spreadsheet under the “Bonus” tab.

STATISTICAL ANALYSIS

The second part of the bonus assignment is the basis statical analysis of the successful and failed campaigns. The tables below show the statistical results of both the successful and failed campaigns.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Successful campaigns** | |  | **Failed campaigns** | |
| Mean | 194.425172 |  | Mean | 17.7098039 |
| Median | 62 |  | Median | 4 |
| Minimum | 1 |  | Minimum | 0 |
| Maximum | 26457 |  | Maximum | 1293 |
| Variance | 712840.987 |  | Variance | 3773.22167 |
| Std Deviation | 844.49238 |  | Std Deviation | 61.446639 |

In the successful campaigns table, we see the mean to be 194.425 and a variance of 712840.987. With that large variance, we can tell the data is skewed making the mean not necessarily be the best way to make a statistical inference. We see the same behavior with the failed campaigns as well where the mean in 17.709 with a variance of 3773.221. In the above table, I will prefer using the median to rather summarize the data.

Comparing both the successful and failed campaigns, we can see that there is more variability in successful campaigns as compared to the failed campaigns. This is because there are more backers of the successful campaigns compare to the unsuccessful.