Kandace Johnson

Module 1 Challenge

June 5, 2023

**Crowdfunding Analysis Report:**

* The crowdfunding data provides several observations, but the three that caught my attention are as follows:
  + Based on initial appearances, the theater category appears to have the highest rate of success and is the most popular.
    - When we examine these categories in more detail, theater accounts for 34% of the total.
    - According to my calculations, theater not only has the lowest overall success, but also has the highest failure rate.
    - Consequently, the popularity of the theater category can be attributed to the fact that there are more projects seeking funding under this category.

|  |  |  |
| --- | --- | --- |
| Categories: | Grand Total | PERCENTAGE |
| Theater: | 319 |  |
| (Film&Video, Food, Games, Journalism, Music, Photography, Publishing, and Technology) summed together: | 610 |  |
| ALL | 929 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Categories | Successful | failed | Grand Total | %Successful | %Fail |
| Theater | **187** | **132** | 319 |  |  |
| Film&Video, Food, Games, Journalism, Music, Photography, Publishing, and Technology | **378** | **232** | 610 |  |  |

* The outcome-based data sets have several limitations, but the following are just a couple:
  + - It is important to note that each outcome-based data set has the limitation that it cannot be interpreted accurately through visual representation.
      * When viewing a set that contains charts and pivot tables, it is very easy to let your eyes mislead you and to assume that you understand what is happening in each set.
    - If you wish to tell the story of data effectively, you must always have an accurate view of the data.
    - A further issue is the lack of category designations for each of the categories.
      * For example, for the category of theater, what type of plays are being funded? Are they musicals, comedies, dramas, horror, or Broadway productions? In the same manner, all other categories in this workbook require further categorization.
    - Furthermore, I find the goal outcome percentage set to be rather broad.
      * This set of data is well understood, but I believe it would be better presented as a pivot table, so the viewer can discern which categories have been successful, failed, or canceled based on the specific goal range in this set.
  + A better interpretation of the data is necessary so as not to mislead the viewer, and additional calculations and modifications should be made, including:
    - Decomposing the data, it may be possible to provide more information to groups who are interested in funding their projects about which projects are likely to receive funding based on their classification.
    - Charts to interpret the following modifications that could be used such as a doughnut chart to show the visualization of the popularity of one’s category:
* It would be more accurate to summarize campaigns based on the mean to provide a more accurate picture of potential success and failure:
  + - By averaging all the values in the data set, the mean represents central tendency.
      * Since medians are not influenced by extremely large values, they are superior to means for data from skewed distributions.
      * Due to these differences, the median would not accurately reflect the actual value we intend to summarize.
    - To conclude, the mean is the most appropriate way to summarize the data.
* A box and whiskers plot is used here to illustrate the data for "successful" (green) and "unsuccessful" (red).
  + - The following visual representation illustrates the variability between the two sets of data. It is evident from the graph that the successful campaign exhibits the largest variation between the two campaigns. On the basis of this graph alone, we can infer that this is accurate as successful campaigns are more likely to be funded.

