Challenge 1

Wednesday, June 7, 2023

11:37 PM

* + **Create a report in Microsoft Word, and answer the following questions for Crowdfunding data and pivot tables:**
    - Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
      * The top three crowdfunding categories in descending order were theater (plays), film & video (documentaries, drama, animation, television, shorts, and science fiction), and music (rock, indie rock, electric music, jazz, metal, and world music).
      * An average of 82 crowdfunding projects were launched per month, where there were increases in March and in the spring/summer (May to July/Aug) and less crowdfunding in the fall and winter, with the exception of December and January.
      * An estimated 57% of projects were successful, 37% failed, and 6% of the projects were canceled.
    - What are some limitations of this dataset?
      * The currencies differ between projects; in order to create funding graphs, the funding (goal and pledged) would need to be converted to the same currency to be compared and contrasted. Calculating "Average Donation" also presents a similar problem - the information doesn't shed light on which projects obtained the most pledged money in USD or euros (as examples of often-used currencies). Units are very telling in this data story, and therefore make conclusions opaque.
    - What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
      * As mentioned above, funding in the same currency would add additional value, as well as the amount each backer provided instead of an average. Box-and-whisker plot of mean, median, and mode of each category would provide outliers. Additionally, tree charts of categories and outcomes would provide hierarchical value.

From <[*https://courses.bootcampspot.com/courses/4079/assignments/56631?module\_item\_id=999323*](https://courses.bootcampspot.com/courses/4079/assignments/56631?module_item_id=999323)>

* + **Statistical Analysis - Central Tendency**
    - Use your data to determine whether the mean or the median better summarizes the data.
      * In general, the mean better summarizes the data when the it is normally distributed and the median when the data is not normally distributed. However, there are cases where one may want to use the median even when the data is normally distributed. For example, the median summarizes the data when outliers are a concern or when comparing two data sets that have different distributions. Therefore, it can be assumed, without using a box-and-whisker visual, there are outliers and that the data is not normally distributed. Therefore, the median should be used to summarize both the successful and failed data.
    - Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
      * The variance and standard deviation are both measures of how spread out the values are in a dataset. However, the standard deviation is more commonly used because it is easier to interpret. Variance is how far values are from the mean (average square distance from the mean). Since the median should be used to summarize both the successful and failed data, the variability in both sets is not a reliable statistic to use to summarize the data properly.

***Successful - not normally distributed***

|  |  |
| --- | --- |
| **Statistical Summary** | **Value** |
| **Mean** | 865 |
| **Median** | 207 |
| **Minimum** | 16 |
| **Maximum** | 7295 |
| **Variance** | 1621157 |
| **Standard Dev** | 1273 |

***Failed - not normally distributed***

|  |  |
| --- | --- |
| **Statistical Summary** | **Value** |
| **Mean** | 589 |
| **Median** | 116 |
| **Minimum** | 0 |
| **Maximum** | 6080 |
| **Variance** | 950271 |
| **Standard Dev** | 975 |