**Excel- Crowdfunding Books Report**

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

Ans: The three conclusions observed in the given data are written as follows:

* The first conclusion we can see is that theatres are the most popular form of entertainment in all the countries and hence the best way to raise the funds, considering 187 campaigns out of 344 were successful worldwide which is significantly higher than the other campaigns. See Table 1:

**A screenshot of a computer

Description automatically generated**

Table

* In contrast, Journalism is the least popular campaign in the world over the years but highly productive nonetheless, because all of the 4 campaigns were successful with 0% rate of failed campaigns or cancelled campaigns. Refer to Table 1.
* If we compare the total campaigns’ success trends among the countries over the given time period, it would be safe to say that US has had the most number of successful campaigns accounting for 436 out of 565 successful campaigns which contributes to 77% of totals, approximately. See Table 2

Table 2

A screenshot of a graph

Description automatically generated

1. **What are some limitations of this dataset?**

Ans: The data provided has several limitations that may affect the interpretation and reliability of the results. Please look at some of these limitations below:

1. High Variability

The file provided has counts that vary significantly between categories. For example, "plays" has 344 outcomes, while "world music" has only 3. This high variability can skew overall analyses and comparisons, making it challenging to generalize findings across categories.

1. Potential for Outliers

The data comprises categories with exceptionally high or low counts. For example, the large number of outcomes for "plays" compared to others might be considered an outlier. Such extreme values can disproportionately influence statistical measures and interpretations.

1. Data Completeness and Accuracy

There’s no indication of how data was collected or if it was cross verified. Missing or incorrect data entries could impact the accuracy of the analysis and conclusions. For example, human errors can always jeopardise the accuracy and reliability of the data.

By addressing these limitations, the accuracy and reliability the statistical analysis and interpretations can be improved.

1. **What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

Ans: There are several datasets that can be used separately to create further potential tables, pivot tables, and graphs. As seen above in Table 3, I have created a Pivot Table which shows the outcome of the campaigns in different countries. It can be filtered by years and shows us which country had how much successful campaigns and how the trends are changing over the years in each country.

The other options that I can think of are using Mean, Median and Mode to create further charts to see the data distribution, or using variance and standard deviation formulas to see the varying trends over the years in different countries and Mode will provide us the with the facts to further add credibility to the story telling of campaigns and how certain recurring methods allow us to increase the potential future campaigns success rate.

**Bonus Statistical Analysis**

1. **Use your data to determine whether the mean or the median better summarises the data.**

Ans: If we create a Pivot table using the number of Backers behind the successful and the unsuccessful campaigns worldwide, we can compare the data distribution trends in both. Look at the following Pivot Table:

A screenshot of a computer

Description automatically generated

The mean is useful for understanding the central tendency. But It is evidently clear from the bar graphs that our data is right-skew. It is due to the extreme values (outliers) existing in both databases. This creates a high variance and standard deviation. Hence, the mean is influenced by extreme data points. See the Box and whisker chart below:

A screenshot of a computer

Description automatically generated

On the contrary, the median provides a measure of central tendency that is less affected by outliers compared to the mean. In this situation, the Australian Bureau of Statistics suggests, “The median is less affected by outliers and skewed data than the mean and is usually the preferred measure of central tendency when the distribution is not symmetrical. “Therefore, median is the best reflector of the given data summary.

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

Ans: The variance measures the average squared deviation from the mean. High variance suggests a large spread of values. The variance for "Successful" outcomes is significantly higher than that for "Unsuccessful" outcomes, indicating greater variability in the "Successful" data. The large differences between the mean and median, coupled with high variance and standard deviation, suggest the presence of significant outliers or skewness in the data, which makes perfect sense as the data suggests the broad range between the minimum and maximum values. The **minimum** number of backers supporting the successful and unsuccessful campaigns are **85 and 1**, whereas **maximum** numbers of backers supporting successful and unsuccessful campaigns are **6,465 and 6,080**, respectively. This supports the observed high variance and standard deviation. Overall, the statistical measures appear reasonable.