Excel-challenge – Rupali Surve

Assumption:

The starterbook data has many currencies listed in column H, currency column. I have assumed that the currencies have been converted to USD in order to calculate Goal and Pledged columns already.

Please let me know if these were expected to be converted per currency and I will resubmit with new calculations.

P.S.: I have also included statistical analysis report below.

Kickstarter campaigns report

After analyzing available data, we can conclude below 3 points pertaining to Kickstarter.

1. Top 3 categories with most successful campaigns are, Theater, Music and Film & Video. Theater has the most successful campaigns of all.
2. Plays subcategory (parent category - Theater) has the most successful campaigns.
3. Month of May has most successful campaigns.

Bonus – Statistical analysis

* Determine whether the mean or the median summarizes the data more meaningfully

**Conclusion:** Analyzing the data, it is apparent that the median summarizes the backers data more meaningfully.

To conclude this, I have used several measures of variability including variance and standard deviation.

For ‘successful campaigns’ and ‘unsuccessful campaigns’ analysis, below are the measure of central tendency and variability that have been calculated,

|  |  |
| --- | --- |
| Successful Campaigns |  |
| The mean number of backers. | 194 |
| The median number of backers. | 62 |
| The minimum number of backers. | 1 |
| The maximum number of backers. | 26457 |
| The variance of the number of backers. | 712841 |
| The standard deviation of the number of backers. | 844 |

|  |  |
| --- | --- |
| Unsuccessful Campaigns |  |
| The mean number of backers. | 18 |
| The median number of backers. | 4 |
| The minimum number of backers. | 0 |
| The maximum number of backers. | 1293 |
| The variance of the number of backers. | 3773 |
| The standard deviation of the number of backers. | 61 |

The mean is very appropriate when the distribution is symmetrical.

However, comparing minimum and maximum number of backers for both successful and unsuccessful campaigns, we can presumably say that the distribution is skewed as the mean 194 is not in the middle of the range.

A better measure of the center for skewed distribution would be the median.

In both cases, the mean is greater than the median. This indicates distribution is skewed to the right. I have used scatter plot to show the same.

* Determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

**Conclusion:** There is more variability with the successful campaigns when compared to unsuccessful campaigns.

To conclude this, I have used several measures of variability for both successful and unsuccessful campaigns, as shown and described below, including Interquartile range and coefficient of variation.

IQR is an easy way to measure how spread out numbers are, without worrying about outliers.

The coefficient of variation (CV) is a measure of relative variability.

|  |  |
| --- | --- |
| Successful Campaigns |  |
| The mean number of backers. | 194 |
| The median number of backers. | 62 |
| The minimum number of backers. | 1 |
| The maximum number of backers. | 26457 |
| The variance of the number of backers. | 712841 |
| The standard deviation of the number of backers. | 844 |

|  |  |
| --- | --- |
| Quartile1 | 29 |
| Quartile3 | 141 |
| InterQuartileRange IQR | 112 |
| Coefficient of Variation CV | 4.35 |

|  |  |
| --- | --- |
| Unsuccessful Campaigns |  |
| The mean number of backers. | 18 |
| The median number of backers. | 4 |
| The minimum number of backers. | 0 |
| The maximum number of backers. | 1293 |
| The variance of the number of backers. | 3773 |
| The standard deviation of the number of backers. | 61 |

|  |  |
| --- | --- |
| Quartile1 | 1 |
| Quartile3 | 12 |
| InterQuartileRange IQR | 11 |
| Coefficient of Variation CV | 3.39 |

Based on above statistics, we have a larger variance for successful campaigns suggesting the backers count is more widely spread out from the mean when compared to unsuccessful campaigns.

Also, if we compare, coefficient of variation, successful campaigns show higher percentage of variation when compared to unsuccessful campaigns.

Both the successful and unsuccessful campaigns have CV >=1 indicates high variation.