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

|  |  |
| --- | --- |
| **Failed campaign** | |
| The mean number of backers | 585.6154 |
| The median number of backers | 114.5 |
| The minimum number of backers | 0 |
| The maximum number of backers | 6080 |
| The variance of the number of backers | 924113.5 |
| The standard deviation of the number of backers | 961.3082 |

In order to determine which central tendency best represents the given data set, let us consider the below data from successful and failed campaigns:

|  |  |
| --- | --- |
| **Successful campaign** | |
| The mean number of backers | 851.1469 |
| The median number of backers | 201 |
| The minimum number of backers | 16 |
| The maximum number of backers | 7295 |
| The variance of the number of backers | 1606216.6 |
| The standard deviation of the number of backers | 1267.366 |

There are 565 successful campaigns, with 480898 backers in total.   
There are 364 failed campaigns, with 213164 backers in total.

For successful campaigns, Mean is 851 and median is 201. Mean vs median differs by over 650. Here, mean is significantly higher than the median.

For failed campaigns, mean is 586 and median is 114.5. Mean vs median differs by over 470. Here also, mean is very high than the median.

Given dataset represents skewed distribution since the mean and median are not equal. When the mean is higher than the median, it means the distribution is skewed to the right.

Since the mean is very high than the median, it also suggests the presence of outliers with large values in the data. Statistically, median is usually not affected by the outliers while mean is sensitive to outliers.

Since the given data represents a skewed distribution, median better summarises the data.

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

Variability in a dataset refers to the dispersion of the data or how spread out the data is from the mean/average.

Given data set is skewed distribution. Variability in a skewed data is usually determined from the interquartile range and standard deviation. Larger the standard deviation and IQR is, the more the variability.

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | **Standard deviation** | **IQR** | **Variance** |
| **Successful campaign** | 1267.366 | 1152 | 1606217 |
| **Failed campaign** | 961.3 | 746.5 | 924113.5 |

Here the values of IQR and Standard deviation of successful campaign is larger than the IQR and standard deviation of failed campaign.

To conclude, Successful campaigns has more variability in it than the failed campaigns.