01

KAGGLE'S NEW YORK STOCK EXCHANGE S&P 500 DATASET

Descriptive statistics

02

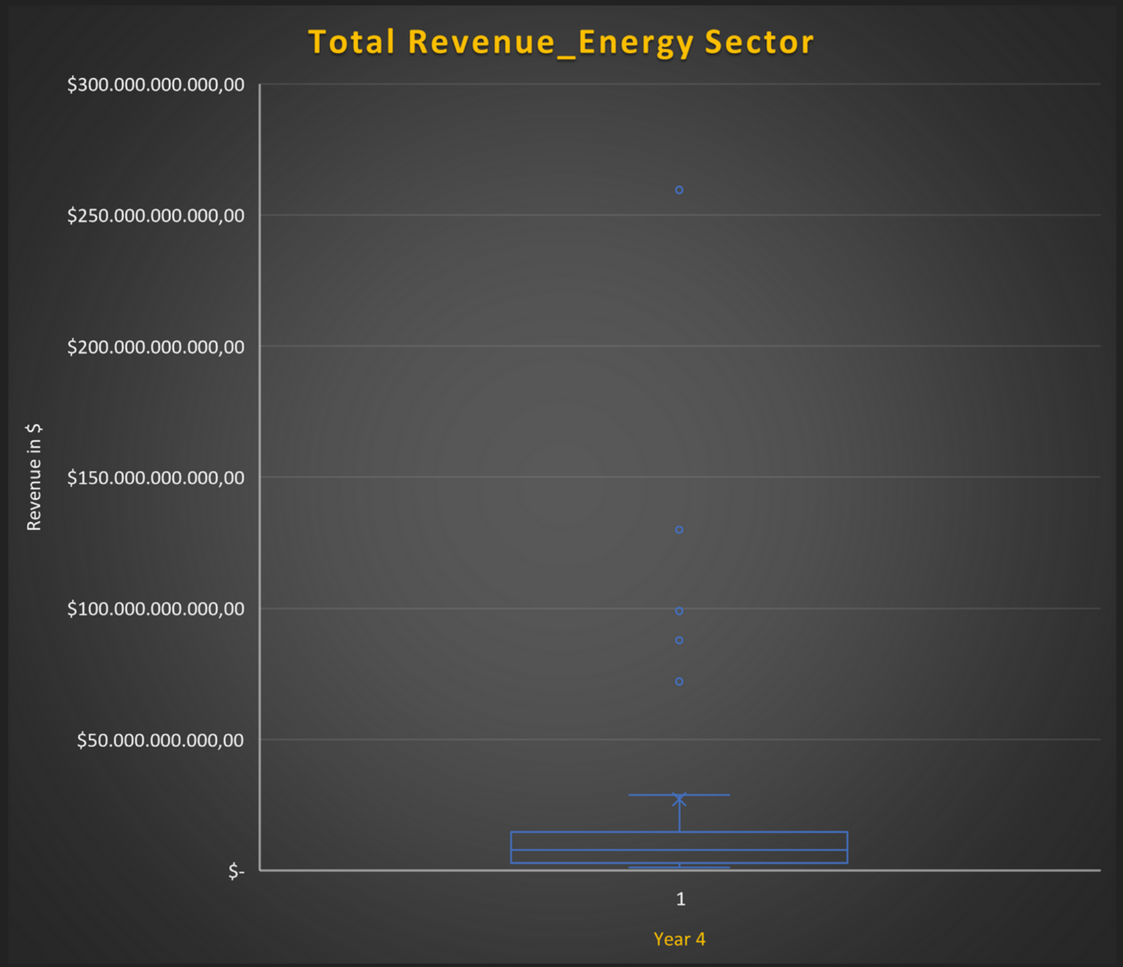
Table of Contents

Presentation Outline

Question:

How do the Energy and Information Technology sectors compare in terms of total revenue in Year 4?

Energy Sector



Mean (x̄)
$ 27.216.366.677,42
The mean is represented by the X in the whisker plot.

Median
$7.763.206.000,00
The median is represented by the vertical line inside the box.

Mode
#N/A since the dataset does not contain any duplicate.

Standard Deviation
$52.799.285.326,36

Skewness
3,24
Skewness is higher than +1, therefore the distribution is highly (right) skewed since the mean > median.

Minimum
$1.181.704.000,00

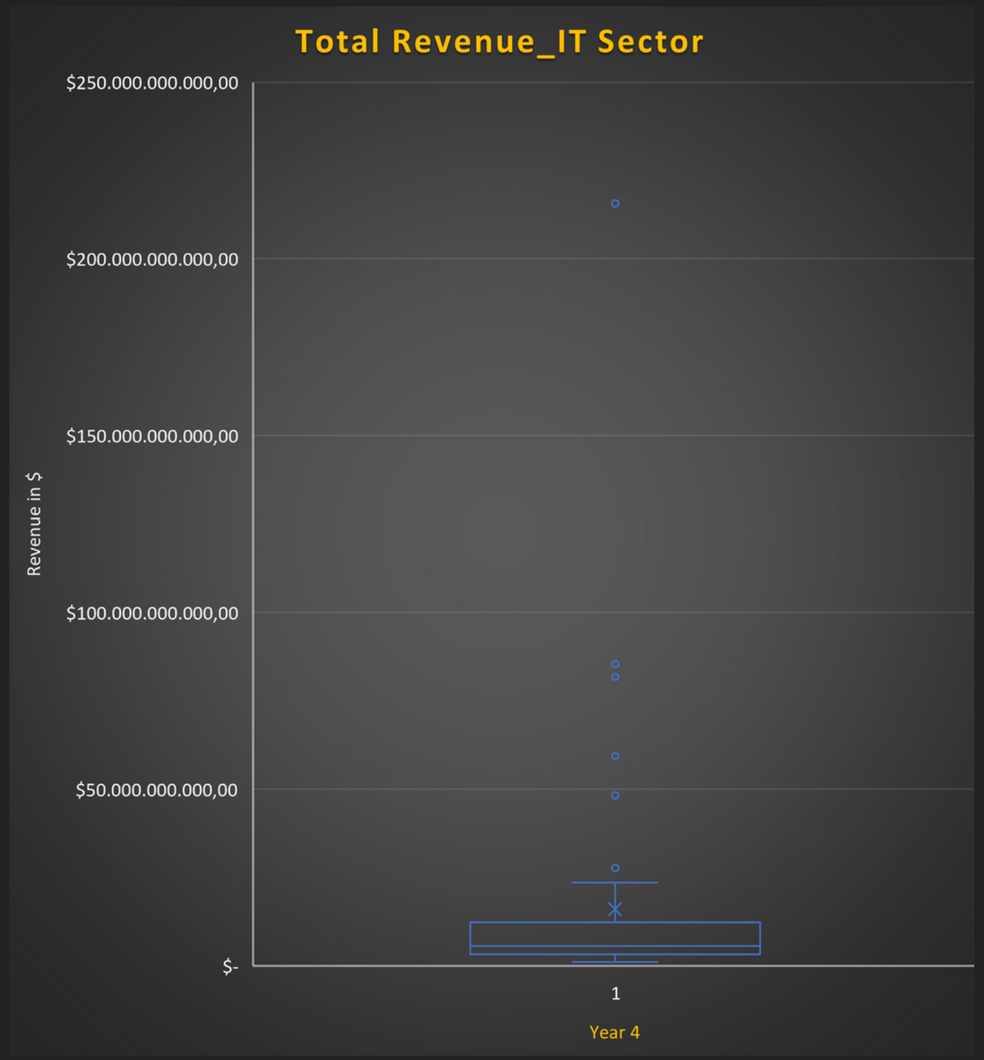
Maximum
$259.488.000.000,00

Range
$258.306.296.000,00

Outliers are represented by the dots in this whisker plot.

Summary Statistics

IT Sector



Mean (x̄)
$16.035.034.618,18
The mean is represented by the X in the whisker plot.

Median
$5.568.700.000,00
The median is represented by the vertical line inside the box.

Mode
#N/A since the dataset does not contain any duplicate.

Standard Deviation
$32.615.202.621,01

Skewness
4,67
Skewness is higher than +1, therefore the distribution is highly (right) skewed since the mean > median.

Minimum
$1.059.366.000,00

Maximum
$215.639.000.000,00

Range
$214.579.634.000,00

Outliers are represented by the dots in this whisker plot.

Summary Statistics

Observations

06

+50%

- of companies in the Energy sector generated a total revenue less than $10.000.000.000,00 in year 4.

- of companies in the IT sector generated a total revenue less than $7.000.000.000,00 in year 4.

When looking at the ticker symbol list, the company in the energy sector which qualifies for the highest revenue share in year 4 is XOM (31% revenue share) while in the same year AAPL dominates the IT sector with a revenue share of about 24%.

In both cases the distribution is right or positively skewed since the mean value is higher than the median value. However, the skewness degree in the IT sector is higher than the one displayed by the Energy sector which means that the distribution is more asymmetrical in the case of the IT sector.

Since data are not normally distributed and outliers are present, the mean and the standard deviation are not recommended to describe the dataset and the 5 Number Summary (i.e. median, max, min and range) represent a more appropriate way to summarize our findings.

Sources

07

Canva

Projectdata NYSE

https://365datascience.com/measuring-asymmetry-skewness/
https://help.gooddata.com/doc/en/reporting-and-dashboards/maql-analytical-query-language/maql-expression-reference/aggregation-functions/statistical-functions/predictive-statistical-use-cases/normality-testing-skewness-and-kurtosis