**ASSIGNMENT – DATASCIENCE UNIVARIATE ANALYSIS**

**Why 1.5 times IQR? Why not one or two or any other number?**

The IQR method of [outlier detection](https://builtin.com/data-science/how-find-outliers-examples) is a method that dictates that any data point in a [boxplot](https://builtin.com/data-science/boxplot) that’s more than 1.5 IQR points below the first [quartile data](https://builtin.com/data-science/how-to-calculate-quartiles) or more than 1.5 IQR points above the third quartile data is considered an outlier

The choice of using 1.5 times the Interquartile Range (IQR) as a threshold in identifying outliers in a dataset is a common convention in statistics.

The IQR is a measure of statistical dispersion, or spread, and is calculated as the difference between the 75th percentile (Q3) and the 25th percentile (Q1) of the data.

* The interquartile (IQR) method of outlier detection uses 1.5 as its scale to detect outliers because it most closely follows **Gaussian distribution**.
* As a result, the method dictates that any data point that's 1.5 points below the lower bound quartile or above the upper bound quartile is an outlier
* A commonly used rule says that a data point is an outlier if it is more than 1.5 ⋅ IQR ‍ above the third quartile or below the first quartile.
* Said differently, low outliers are below Q 1 − 1.5 ⋅ IQR ‍ and high outliers are above Q 3 + 1.5 ⋅ IQR

When it comes to detecting outliers using the IQR method, multiplying the IQR by 1.5 is a standard practice ***because it strikes a balance between sensitivity to outliers and robustness to noise in the data***.

**This multiplier of 1.5 is not a strict rule but a generally accepted standard** that provides a good compromise between detecting potential outliers and avoiding false alarms.

* Choosing a multiplier value higher than 1.5 would make the method more sensitive to outliers but might also lead to detecting too many false positives, while choosing a lower multiplier would make the method less sensitive to outliers, potentially missing some outliers in the data.
* Ultimately, the choice of 1.5 times the IQR as a threshold in outlier detection is a common convention that has been found to work well in practice for many datasets.
* However, in specific cases or for different types of data, researchers may choose to adjust this multiplier based on the characteristics of the data and the goals of the analysis.