Interquartile Range (IQR)

Why we use 1.5 IQR?

- Using 1 1x IQR would be like setting a super-low bar. For normally distributed data, it would flag about 5% of the data as outliers. That's way too many points to be "special."
- Using 2x IQR would be a very high bar. It would only catch the most extreme values, leaving about 0.06% of the data as outliers. It would miss a lot of interesting anomalies.
- 1.5x IQR hits a sweet spot, flagging about 0.7% of the data. It's a reasonable cutoff that's just tight enough to catch suspicious points without being overly sensitive.

Find any outliers in the below set

	Minimum	Q1	Medium	Q3	Maximum
Day	32	56	74.5	82.5	99
Night	25.5	78	81	89	98

1. Calculation of IQR (Interquartile Range)

The Interquartile Range (IQR) is calculated using the formula:

$$\boldsymbol{IQR = Q3 - Q1}$$

Results:

- For Day Classes,
  $IQR = 82.5 - 56 = 26.5$
- For Night Classes,
  $IQR = 89 - 78 = 11$

Thus, the IQR values are 26.5 for Day Classes and 11 for Night Classes.

---

2. Calculation of Lower Outlier Value

The Lower Outlier boundary is calculated using the formula:

$$\textbf{Lower Outlier} = \boldsymbol{Q1 - (1.5 \times IQR)}$$

Results:

- For Day Classes,
  $56 - (1.5 \times 26.5) = 16.25$
- For Night Classes,
  $78 - (1.5 \times 11) = 61.5$

Therefore, any value below 16.25 for Day Classes or below 61.5 for Night Classes is considered a potential lower outlier.

---

3. Calculation of Upper Outlier Value

The Upper Outlier boundary is determined using the formula:

$$\textbf{Upper Outlier} = \boldsymbol{Q3} + (\mathbf{1.5} \times \boldsymbol{IQR})$$

Results:

- For Day Classes,
  $82.5 + (1.5 \times 26.5) = 122.25$
- For Night Classes,
  $89 + (1.5 \times 11) = 105.5$

Hence, any value above 122.25 for Day Classes or above 105.5 for Night Classes would be considered an upper outlier.

---

4. Findings and Interpretation

- For the Day Classes, all data points fall within the calculated lower and upper boundaries. Therefore, no outliers are identified for this group.
- For the Night Classes, some data value (25.5) falls below the lower boundary (61.5), indicating the presence of lower outliers.
- No upper outliers are found in either category.