IQR Assignments

1. **Why are we multiplying 1.5 to IQR value while calculating lesser and greater outliers?**

Lesser outliers: Q1- 1.5(IQR)
Greater outliers: Q3+ 1.5(IQR)

Here , 1.5 is the scale which controls sensitivity of the range of data and hence the decision rule.

Why 1.5?

1. Scale is 1
    a. If the scale is 1 then , we may consider few data points as outliers
2. Scale is 2
    a. If the scale is 2 then, we may consider outliers as data points
3. Scale is 1.7
    a. This is the right value to get only the valid data points for models

Why 1.5 , not 1.7?

To maintain symmetrical we are considering 1.5 as the final scale.

Note: This scale (1.5) may change based on the distribution of data.

2. Solve below assignment

Interquartile Range(IQR)

a. The interquartile range. Compare the two interquartile ranges.

b. Any outliers in either set.

The five number summary for the day and night classes is

	Minimum	Q_1	Median	Q_3	Maximum
Day	32	56	74.5	82.5	99
Night	25.5	78	81	89	98

1. **Lets calculate all the values for Day data:**
    a. IQR = Q3-Q1 => 82.5-56 => 26.5
    b. Calculate Lesser outliers
        i. Q1- 1.5 (IQR)
            1. 56 - 1.5 (26.5) => 16.25
    c. Calculate greater outliers
        i. Q3 + 1.5 (IQR)

1. 82.5 + 1.5(26.5) => 122.25

2. **Lets calculate all the values for Night data**
   - **a.** IQR = Q3-Q1 => 89-78 => 11
   - b. Calculate Lesser outliers
     - i. Q1- 1.5 (IQR)
       1. 78 - 1.5 (11) => 61.5
   - c. Calculate greater outliers
     - i. Q3 + 1.5 (IQR)
       1. 89 + 1.5(11) => 105.5

Answers:

- **a. The interquartile range. Compare the interquartile range**
  - i. The IQR value for Day data is 26.5 and The IQR value for Night data is 11.
  - ii. Note: If IQR value is small then we can say that all the 50% of middle dataset are closer together and it is consistent.
  - iii. **Night has lower IQR than Day so Night has low variability which means 50% of middle datasets are closer to each other and from the median and it is consistent.**
    1. here variability describes how far apart data points lie from each other and from the center of a distribution
- **b. Any outliers in either set:**
  - **i. Day:**
    1. Lesser outliers of Day is 16.25 which is greater than the minimum value i.e 32 so there is no lesser outliers in Day
    2. Greater outlier of Day is 122.25 which is greater than the maximum value 99 so there is no greater outlier in Day
  - ii. Night:
    1. Lesser outliers of Day is 61.5 which is greater than the minimum value i.e 25.5 so there is no lesser outliers in Day
    2. Greater outlier of Day is 105.25 which is greater than the maximum value 98 so there is no greater outlier in Day

Overall, there is no outliers in both Day and Night