5 Number summary:

* Minimum
* First quartile (Q1, 25 Percentile)
* Median
* Third Quartile – 75 percentile (Q3)
* Maximum

All the above techniques are used to find the Outliers. This is very important for feature engineering.

Let us discuss all the above techniques with an example.

**Dataset : (1,2,2,2,3,3,3,4,5,5,5,6,6,6,6,7,8,8,9,27}**

**First of all, let us discuss about,**

* **What is outlier?**
* **Process to find the outliers?**

**What is outlier?**

An outlier is a data point that is significantly different from the majority of other data points in a dataset. Outliers can be much higher or lower in value than the other points in the dataset.

**Example1:**

Let’s imagine in a group of 5 students, the test grades were 9, 8, 9, 7, and 2. The last value seems to be an outlier because it falls below the main pattern of the other grades.

**Example2:**

Let us take n a real-world example; the average height of a giraffe is **about 16 feet tall**. However, there have been recent discoveries of two giraffes that stand at **9 feet and 8.5 feet**, respectively. These two giraffes would be considered outliers in comparison to the general giraffe population.

Now, let’s come to our data science scenario.

**Process to find the outliers?**

First of all, we need to find out the **higher fence and lower fence**

**Lower fence = Q1 - 1.5 (IQR)**

**Higher Fence = Q3 + 1.5 (IQR)**

Here **IQR** mean **Inter Quartile Range.**

**IQR = Q3 – Q1 (difference between Q3 and Q1)**

**Q1 = (25/100) \* (n+1)**

**= 5.25th index**

In 5.25th index we have 2 numbers 3, 3. So need to find the mean (3+3)/2=**3.**

**Q3 = (75/100) \* (n+1)**

**= 15.75th index**

15.75th index, we have 8 and 7, so need to find mean (8+7)/2 = **7.5**

**So IQR = 7.5 – 3 = 4.5**

**Lower fence = Q1 - 1.5 (IQR) = 3 – 1.5 (4.5) = - 3.65**

**Higher Fence = Q3 + 1.5 (IQR) = 7.5 + 1.5 (4.5) = 14.25**

**Ranging from [- 3.65 🡨-🡪14.25]**

**So below are the findings**

1. **Minimum = 1**
2. **Q1 = 3**
3. **Median = 5 [(5+5)/2]**
4. **Q3 = 7.5**
5. **Maximum = 9**