

IQR Report – Outlier Detection – Placement Dataset

In this report, I used the **Interquartile Range (IQR) method** to identify outliers in each numeric column of the placement dataset. I first calculated Q1 (25th percentile), Q3 (75th percentile), and then found the IQR by subtracting Q1 from Q3.

To find potential outliers, I applied the $1.5 \times \text{IQR}$ rule, which sets two thresholds:

- **Lesser Bound** = $Q1 - 1.5 \times \text{IQR}$
- **Greater Bound** = $Q3 + 1.5 \times \text{IQR}$

If any value is less than the lesser bound or greater than the greater bound, it is considered an outlier.

I used the exact table below to guide my analysis:

Table (Values Used):

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Q1:25%	54.5	60.6	60.9	61.0	60.0	57.945	240000.0
Q3:75%	161.5	75.7	73.0	72.0	83.5	66.255	300000.0
IQR	107.0	15.1	12.1	11.0	23.5	8.31	60000.0
1.5rule	160.5	22.65	18.15	16.5	35.25	12.465	90000.0
Lesser Bound	-106.0	37.95	42.75	44.5	24.75	45.48	150000.0
Greater Bound	322.0	98.35	91.15	88.5	118.75	78.72	390000.0
Min	1	40.89	37.0	50.0	50.0	51.21	200000.0
Max	215	89.4	97.7	91.0	98.0	77.89	940000.0

Column-by-Column Outlier Check:

1. sl_no (Serial Number)

- Min and Max values are within the limits of -106.0 to 322.0.
No outlier found

2. ssc_p (Secondary School %)

- Values range from 40.89 to 89.4, which fall within 37.95 and 98.35.
No outlier found

3. hsc_p (Higher Secondary %)

- Min = 37.0 is less than 42.75
- Max = 97.7 is greater than 91.15
Two outliers found -> one low and one high

4. degree_p

- Min = 50.0 (greater than lesser bound 44.5)
- Max = 91.0 is slightly higher than 88.5

One possible high outlier

5. etest_p (Employability Test)

- Min = 50.0 and Max = 98.0 fall within the safe range

No outlier found

6. mba_p

- All values (51.21 to 77.89) lie within the range of 45.48 to 78.72

No outlier found

7. salary

- Max = 940000 is much greater than 390000

One high outlier found

Result:

Column	Outlier Found?	Details
sl_no	No	Values within expected range
ssc_p	No	Normal distribution
hsc_p	Yes	Min = 37.0, Max = 97.7 (both outliers)
degree_p	Yes	Max = 91.0 is just above 88.5
etest_p	No	All values within range
mba_p	No	Normal range
salary	Yes	Max = 940000 is an outlier

By applying the IQR method, I was able to identify which columns contain potential outliers in the placement dataset:

- hsc_p had both a low and high outlier
- salary had a very high outlier
- degree_p had a value that was just above the upper limit, so it might be a mild outlier
- All other features were clean and normally distributed

This analysis helped me understand how data varies across features and how outliers can be detected using IQR-based boundaries.