Placement Dataset IQR Analysis

Dataset:

```
dataset
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	NaN
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
210	211	M	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.0	Mkt&Fin	74.49	Placed	400000.0
211	212	M	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.0	Mkt&Fin	53.62	Placed	275000.0
212	213	M	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.0	Mkt&Fin	69.72	Placed	295000.0
213	214	F	74.00	Others	66.00	Others	Commerce	58.00	Comm&Mgmt	No	70.0	Mkt&HR	60.23	Placed	204000.0
214	215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.0	Mkt&HR	60.22	Not Placed	NaN

Quantitative Analysis:

```
dataset[quan]
```

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
0	67.00	91.00	58.00	55.0	58.80	270000.0
1	79.33	78.33	77.48	86.5	66.28	200000.0
2	65.00	68.00	64.00	75.0	57.80	250000.0
3	56.00	52.00	52.00	66.0	59.43	NaN
4	85.80	73.60	73.30	96.8	55.50	425000.0
...	...	...	...	...	...	...
210	80.60	82.00	77.60	91.0	74.49	400000.0
211	58.00	60.00	72.00	74.0	53.62	275000.0
212	67.00	67.00	73.00	59.0	69.72	295000.0
213	74.00	66.00	58.00	70.0	60.23	204000.0
214	62.00	58.00	53.00	89.0	60.22	NaN

215 rows × 6 columns

Output:

IQR	15.1	12.1	11.0	23.5	8.31	60000.0
1.5rule	22.65	18.15	16.5	35.25	12.465	90000.0
lesser_outlier	37.95	42.75	44.5	24.75	45.48	150000.0
greater_outlier	98.35	91.15	88.5	118.75	78.72	390000.0
min	40.89	37.0	50.0	50.0	51.21	200000.0
max	89.4	97.7	91.0	98.0	77.89	940000.0

IQR Comparison:

If the IQR value is low , then the 50% of middle data points are closer to each other points and it is closer to the median.

Please find the IQR from low variability to high variability
Low variability - High variability
mba_p < degree_p < hsc_p < ssc_p < etest_p < salary

Let's check if we have outliers in any of the columns or not

ssc_p:
Min value (40.89) > Less_outlier (37.95) so no lesser outliers
Max value (89.4) < greater_outlier (98.35) so no greater outliers

hsc_p:
Min value (37.0) < Less_outlier (42.75) so it has lesser outliers
Max value (97.7) > greater_outlier (91.15) so it has greater outliers

degree_p:
Min value (50.0) > Less_outlier (44.5) so it has lesser outliers
Max value (91.0) > greater_outlier (88.5) so it has greater outliers

etest_p:
Min value (50.0) > Less_outlier (24.75) so it has lesser outliers
Max value (98.0) < greater_outlier (118.75) so it has greater outliers

mba_p:
Min value (51.21) > Less_outlier (45.48) so it has lesser outliers
Max value (77.89) < greater_outlier (78.72) so it has greater outliers

salary:
Min value (200000) > Less_outlier (150000) so it has lesser outliers
Max value (940000) > greater_outlier (390000) so it has greater outliers

Below is final observation:

Outliers	**ssc_p**	**hsc_p**	**degree_p**	**etest_p**	**mba_p**	**salary**
Are Lesser Outliers available?	No	**Yes**	No	No	No	No
Are Greater Outliers available?	No	**Yes**	**Yes**	No	No	**Yes**

As per the above results, hsc_p has both lower and upper bound outliers and salary has only upper bound outliers