Interquartile Range & Replacing Outliers

Dataset: Placement.csv

3.Topic : IQR

IQR formula = Q3=Q1

Lesser Outlier = Q1-1.5*IQR

Greater Outlier=Q3-1.5*IQR

Why 1.5?

1.5 is taken because it is a balanced cutoff. It will take care not going too high or too low.

Example : IQR=20,

Lets say formula is 1*IQR= 1*20 = 20

1.5*IQR=1.5*20=30

3*1QR=3*20=60

Here we see that 20 and 60 are low and high variations. But 30 is a middle value which works well for data.

```
[49]: descriptive_analysis_table
```

Out[49]:

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Q1:25th	54.5	60.6	60.9	61	60	57.945	240000
Q2:50th	108	67	65	66	71	62	265000
Q3:75th	161.5	75.7	73	72	83.5	66.255	300000
Q4:100th	215	89.4	97.7	91	98	77.89	940000
IQR	107	15.1	12.1	11	23.5	8.31	60000
1.5rule	160.5	22.65	18.15	16.5	35.25	12.465	90000
lesser_outlier	-106	37.95	42.75	44.5	24.75	45.48	150000
greater_outlier	322	98.35	91.15	88.5	118.75	78.72	390000
min	1	40.89	42.75	50	50	51.21	200000
max	215	89.4	91.15	88.5	98	77.89	390000

The highlighted values are replaced with the outlier values.