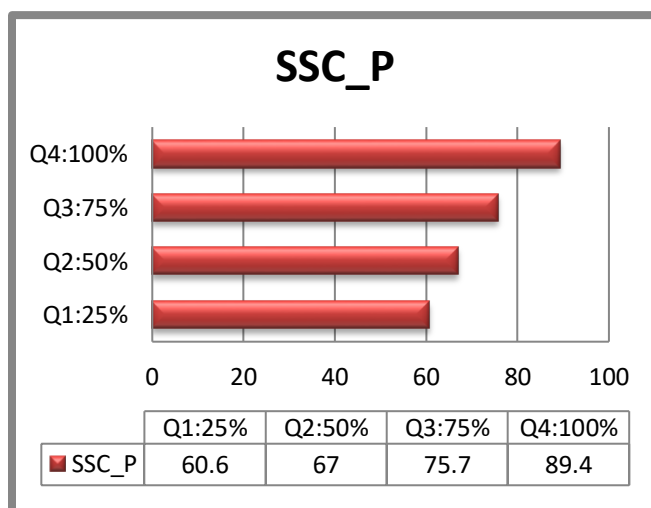


PERCENTILE

- Percentile is a measure that tells the value below which a certain percentage of data falls.
- Here the percentile is represented as quadrants Q1,Q2,Q3 and Q4.
- Table shown below is a placement data percentages of student marks under 25%,50%,75% and 100%.

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108	67	66	66	72	62	288655
Median	108.0	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	1	62.0	63.0	65.0	60.0	56.7	300000.0
Q1:25%	54.5	60.6	60.9	61.0	60.0	57.945	240000.0
Q2:50%	108.0	67.0	65.0	66.0	71.0	62.0	265000.0
Q3:75%	161.5	75.7	73.0	72.0	83.5	66.255	300000.0
80%	172.2	77.0	74.728	72.134	86.0	67.144	NaN
99%	212.86	87.0	91.86	83.86	97.0	76.1142	NaN
Q4:100%	215.0	89.4	97.7	91.0	98.0	77.89	940000.0

- Tables would be plotted as graphs for easy analysing and understanding.



Differences between each quadrants:

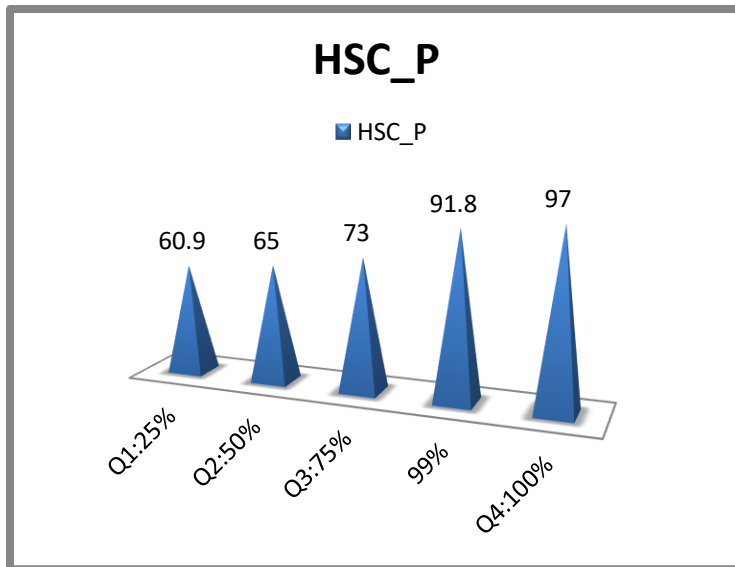
$$Q1 \sim Q2 = 6.4$$

$$Q2 \sim Q3 = 8.7$$

$$Q3 \sim 99\% = 11.3$$

$$Q3 \sim Q4 = 13.7$$

$$99\% \sim 100\% = 2.4$$



Differences between each quadrants:

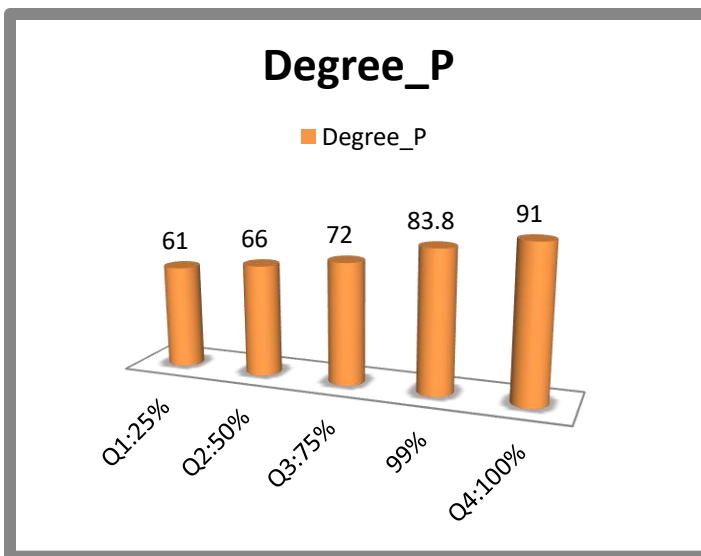
$$Q1 \sim Q2 = 4.1$$

$$Q2 \sim Q3 = 8$$

$$Q3 \sim 99\% = 18.8$$

$$Q3 \sim Q4 = 24.7$$

$$99\% \sim 100\% = 5.8$$



Differences between each quadrants:

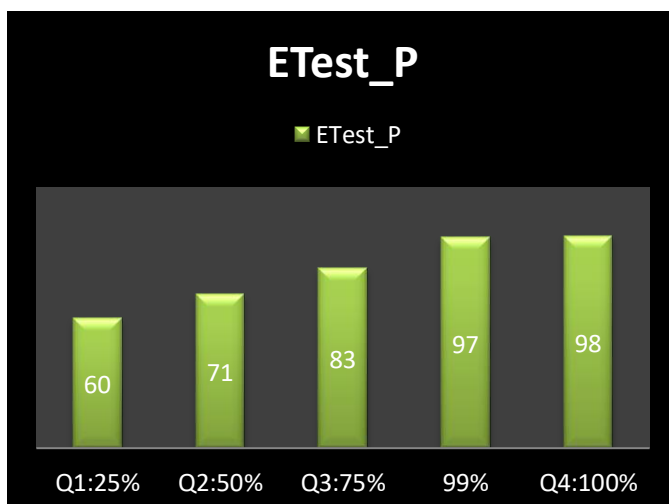
$$Q1 \sim Q2 = 5$$

$$Q2 \sim Q3 = 6$$

$$Q3 \sim 99\% = 11.8$$

$$Q3 \sim Q4 = 19$$

$$99\% \sim 100\% = 7.2$$



Differences between each quadrants:

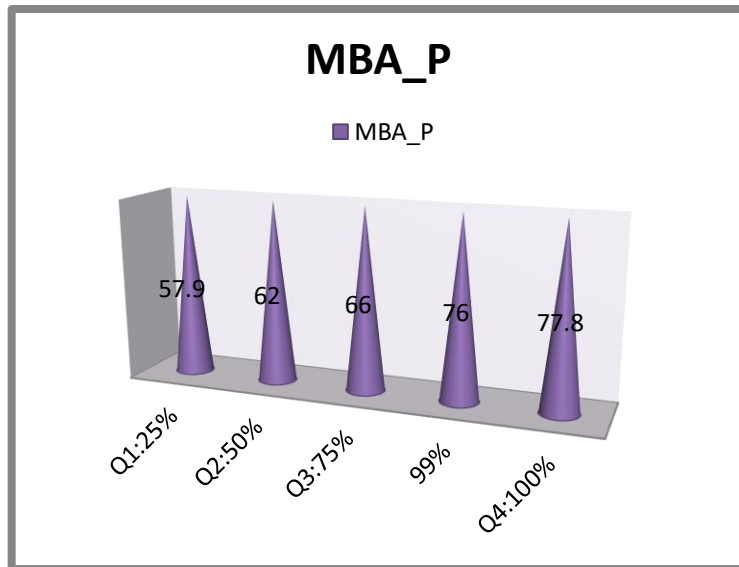
$$Q1 \sim Q2 = 11$$

$$Q2 \sim Q3 = 12$$

$$Q3 \sim 99\% = 14$$

$$Q3 \sim Q4 = 15$$

$$99\% \sim 100\% = 1 \text{ Nearly the same}$$



Differences between each quadrants:

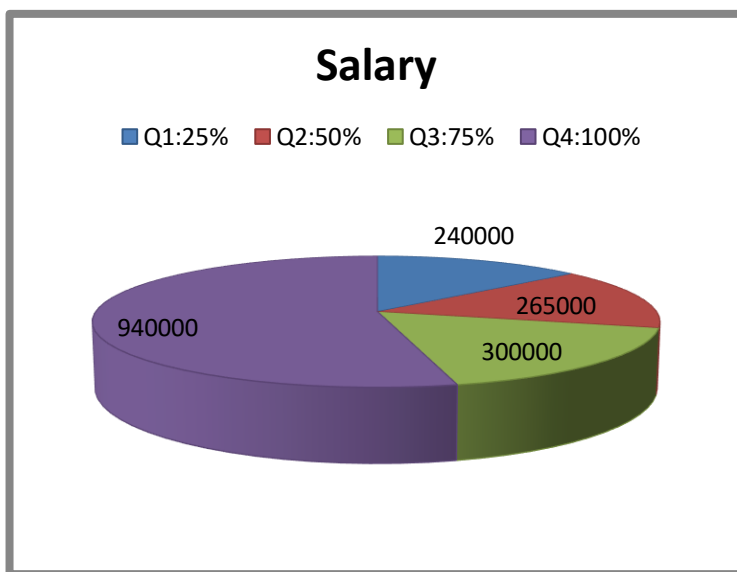
$$Q1 \sim Q2 = 4.1$$

$$Q2 \sim Q3 = 4$$

$$Q3 \sim 99\% = 10$$

$$Q3 \sim Q4 = 11.8$$

$$99\% \sim 100\% = 1.8$$



Differences between each quadrants:

$$Q1 \sim Q2 = 25000$$

$$Q2 \sim Q3 = 35000$$

$$Q3 \sim Q4 = 64000$$

INTER QUARTILE RANGE (IQR)

- IQR is used to replace the acceptable values for outlier values.
- The purpose of IQR is to know the outlier range present in the dataset.

$$\text{IQR} = Q3 - Q1$$

- Lesser Outlier = $Q1 - 1.5(\text{IQR})$
- Greater Outlier = $Q3 + 1.5(\text{IQR})$
- Why 1.5 factor is chosen?
 - * 1.5 IQR is a Rule of Thumb declared by John Tuckey.
 - * IQR is less sensitive to skewed data distributions, as it is based on percentiles.
 - * 1.5 multiplier determines the width of the outlier detection boundaries.
 - * A large multiplier would make the fence wider considers more points as valid points and potentially missing genuine outlier.
 - * A smaller multiplier would make the fence narrower, potentially labeling the valid extreme data points as outlier, which is undesirable.
- IQR Question to find 5 values in day and night.

1) Dataset

	Min	Q1	Median	Q3	Max
Day	32	56	74.5	82.5	99
Night	25.5	78	81	89	98

Solution:

Totally 5 Values.

Day - 32,56,74.5,82.5,99

Night – 25.5,78,81,89,98

