1. Analyse the mean , median, mode for below

```
descriptive
```

[42]:

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108.0	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
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

Summary of Mean Median & Mode:

Mean:

The average salary is ₹288,655, with an average SSC score of 67.30%, HSC score of 66.33%, degree score of 66.37%, entrance test score of 72.10%, and MBA score of 62.28%.

Median:

The median salary is ₹265,000, indicating that half the data falls below this amount. The median academic scores are generally close to their means, except for the entrance test score, which has a median of 71%.

Mode:

The mode (most frequent value) for salary is ₹300,000, suggesting a concentration of salaries around this value. Similarly, academic scores cluster around 62%-65%.

2.Analyse the Percentile for below data

```
descriptive
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108.0	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
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
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

Summary of Percentile:

1. SSC_P (Secondary School Certificate Percentage)
   - Q1:25%: 60.6%
   - Q2:50% (Median): 67.0%
     - Difference (Q2 - Q1): +6.4%
   - Q3:75%: 75.7%
     - Difference (Q3 - Q2): +8.7%
   - 99%: 87.0%
     - Difference (99% - Q3): +11.3%
   - Q4:100%: 89.4%
     - Difference (Q4 - 99%): +2.4%

Summary: SSC_P increases by 6.4% from Q1 to Q2, by 8.7% from Q2 to Q3, and by 11.3% from Q3 to the 99th percentile. There is a smaller rise of 2.4% from the 99th percentile to the maximum (Q4:100%).

2. HSC_P (Higher Secondary Certificate Percentage)

- Q1:25%: 60.9%
- Q2:50% (Median): 65.0%
  - Difference (Q2 - Q1): +4.1%
- Q3:75%: 73.0%
  - Difference (Q3 - Q2): +8.0%
- 99%: 91.86%
  - Difference (99% - Q3): +18.86%
- Q4:100%: 97.7%
  - Difference (Q4 - 99%): +5.84%

Summary: HSC_P increases by 5.0% from Q1 to Q2, by 8.0% from Q2 to Q3, and by a significant 18.86% from Q3 to the 99th percentile. The rise between the 99th percentile and Q4 is 5.84%.

3. Degree_P (Degree Percentage)

- Q1:25%: 61.0%
- Q2:50% (Median): 66.0%
  - Difference (Q2 - Q1): +5.0%
- Q3:75%: 72.0%
  - Difference (Q3 - Q2): +6.0%
- 99%: 83.86%
  - Difference (99% - Q3): +11.86%
- Q4:100%: 91.0%
  - Difference (Q4 - 99%): +7.14%

Summary: Degree_P shows consistent increases of 6.0% between Q1 to Q2 and Q2 to Q3, with a larger 11.86% jump between Q3 and the 99th percentile. The final rise between the 99th percentile and Q4 is 7.14%.

4. ETest_P (Entrance Test Percentage)

- Q1:25%: 60.0%
- Q2:50% (Median): 71.0%
  - Difference (Q2 - Q1): +11.0%
- Q3:75%: 72.0%
  - Difference (Q3 - Q2): +1.0%
- 99%: 97.0%
  - Difference (99% - Q3): +25.0%
- Q4:100%: 98.0%
  - Difference (Q4 - 99%): +1.0%

Summary: ETest_P shows an 11.0% increase from Q1 to Q2, a small 1.0% rise from Q2 to Q3, and a large jump of 25.0% from Q3 to the 99th percentile. The increase between the 99th percentile and Q4 is minimal at 1.0%.

5. MBA_P (MBA Percentage)

- Q1:25%: 57.945%
- Q2:50% (Median): 62.0%
  - Difference (Q2 - Q1): +4.055%
- Q3:75%: 66.255%
  - Difference (Q3 - Q2): +4.255%
- 99%: 76.1142%
  - Difference (99% - Q3): +9.8592%
- Q4:100%: 77.89%
  - Difference (Q4 - 99%): +1.7758%

Summary: MBA_P shows moderate increases of around 4.0% between Q1 to Q2 and Q2 to Q3, with a larger 9.86% rise from Q3 to the 99th percentile. The final increase from the 99th percentile to Q4 is 1.78%.

6. Salary
   - Q1:25%: ₹240,000
   - Q2:50% (Median): ₹265,000
      - Difference (Q2 - Q1): +₹25,000
   - Q3:75%: ₹300,000
      - Difference (Q3 - Q2): +₹35,000
   - 99%: After pre-processing data , will check the percentile of 99 for the salary column

   - Q4:100%: ₹940,000
      - Difference (Q4 - 99%): ₹0 (Same as 99%)

Summary: Salary shows moderate increases of ₹25,000 from Q1 to Q2, and ₹35,000 from Q2 to Q3, followed by a dramatic jump of ₹640,000 from Q3 to 100 percentile.

3.IQR lesser /greater outlier

35]:

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108.0	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
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
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
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	-106.0	37.95	42.75	44.5	24.75	45.48	150000.0
Greater	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

Let's analyze the columns for outliers using the 1.5 IQR rule.

1. ssc_p
   - Lesser: 37.95
   - Greater: 98.35
   - **Outliers**: None (Min = 40.89, Max = 89.4 within range)
2. hsc_p
   - Lesser: 42.75

    - Greater: 91.15
    - **Outliers**: Both lesser & greater outliers present slightly (Min = 37.0 ,Max= 97.7
3. degree_p
    - Lesser: 44.5
    - Greater: 88.5
    - **Outliers**: there is greater outlier (Min = 50.0, Max = 91.0 close to upper bound)
4. etest_p
    - Lesser: 24.75
    - Greater: 118.75
    - **Outliers**: None (Min = 50.0, Max = 98.0 within range)
5. mba_p
    - Lesser: 45.48
    - Greater: 78.72
    - **Outliers**: None (Min = 51.21, Max = 77.89 within range)
6. salary
    - Lesser: 150000
    - Greater: 390000
    - **Outliers**: Yes , there is greater outlier
        - Min = 200000 (Not an outlier, as it is above 150000)
        - Max = 940000 (Outlier, as it is greater than 390000!)

Conclusion:

- The only clear outlier is in the salary column: 940000.
- Other columns do not have extreme outliers based on the 1.5 IQR rule.

IQR:Ref

Let's say we have the dataset:

5, 7, 8, 9, 10, 12, 15, 100

- The **minimum** value is **5**.
- The **maximum** value is **100**.

Detecting Outliers Using IQR:

1. **Find Q1 (First Quartile) and Q3 (Third Quartile):**
   - Q1 (25th percentile) = 7.5
   - Q3 (75th percentile) = 12.75
2. **Compute IQR:**

$$IQR = Q3 - Q1 = 12.75 - 7.5 = 5.25$$

3. **Find the lower and upper bounds:**
   - Lower bound: $Q1 - 1.5 \times IQR = 7.5 - 1.5(5.25) = -0.375$
   - Upper bound: $Q3 + 1.5 \times IQR = 12.75 + 1.5(5.25) = 20.625$
4. **Identify Outliers:**
   - Any value below **-0.375** or above **20.625** is an outlier.
   - Since **100** is greater than **20.625**, it is an **outlier**.
   - **5** is within range, so it's **not an outlier**.

Conclusion:

- **Outlier** in this dataset: **100**
- **Min (5) and Max (100) alone don't determine outliers**, but the max value (100) is an outlier in this case.

4.Skewness & Kurtosis

```
Univariate(dataset,quan)
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108.0	67.303395	66.334744	66.358558	72.100558	62.278186	277648.648649
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
99%	212.86	87.0	91.129	83.86	97.0	76.1142	NaN
Q4:100%	215.0	89.4	91.15	88.5	98.0	77.89	390000.0
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	-106.0	37.95	42.75	44.5	24.75	45.48	150000.0
Greater	322.0	98.35	91.15	88.5	118.75	78.72	390000.0
Min	1	40.89	42.75	50.0	50.0	51.21	200000.0
Max	215	89.4	91.15	88.5	98.0	77.89	390000.0
kurtosis	-1.2	-0.60751	0.086901	-0.09749	-1.08858	-0.470723	-0.239837
skew	0.0	-0.132649	0.162611	0.204164	0.282308	0.313576	0.8067