Analysis of PLACEMENT DATASET

logic to separate Quan & Qual - Automation logic using Function

```
def quanQual(dataset):
    quan=[]
    qual=[]
    for columnName in dataset.columns:
        if (dataset[columnName].dtypes == "O"):   # "O" - Object Data type
            qual.append(columnName)
        else:
            quan.append(columnName)
    return quan,qual

quan,qual = quanQual(dataset)  #calling quanQual() function
print("quan:",quan)  # numerical data
print("qual:",qual)  # categorical data
```

```
quan: ['ssc_p', 'hsc_p', 'degree_p', 'etest_p', 'mba_p', 'salary']
qual: ['gender', 'ssc_b', 'hsc_b', 'hsc_s', 'degree_t', 'workex', 'specialisation', 'status']
```

MEASURE OF CENTRAL TENDENCY

```
for columnName in quan:
    univariate_analysis.loc["Mean",columnName]=dataset[columnName].mean()
    univariate_analysis.loc["Median",columnName]=dataset[columnName].median()
    univariate_analysis.loc["Mode",columnName]=dataset[columnName].mode()[0]

univariate_analysis
```

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
Median	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	62.0	63.0	65.0	60.0	56.7	300000.0

MY ANALYSIS IN BUSINEES POV:

Analysis of Placement Dataset Using Measures of Central Tendency

I analyzed the placement dataset by first separating the **quantitative** and **qualitative** columns. For the quantitative variables, I applied measures of central tendency—**mean, median, and mode**—to understand the overall academic and placement performance of the students.

🔷 Mean Analysis

The mean values provide an overview of the average performance of students across different stages of education:

- The **mean SSLC pass percentage** is **67%**, indicating that students have, on average, performed at an **average level** in their SSLC board examinations.
- The **mean HSC pass percentage** is **66.3%**, which also suggests an **average performance** among students.
- The **mean degree pass percentage** is **66.3%**, showing that the overall degree performance of the batch is **average**.
- The **mean E-test score** is **72.1%**, which reflects a **good performance** in the entrance test.
- The **mean MBA pass percentage** is **62%**, indicating that students have performed at an **average level** during their MBA.
- The **mean salary offered during placements** is **₹2,88,655**, which gives an idea of the average compensation received by students.

From the mean analysis, it can be inferred that most students demonstrate **average academic performance**, with relatively better results in the E-test.

🔷 Median Analysis

The median values help understand the central tendency **without the influence of outliers**:

- The **median SSLC percentage** is **67%**
- The **median HSC percentage** is **65%**
- The **median degree percentage** is **66%**

These values indicate that the majority of students have performed at an **average level** academically.

- The **median E-test score** is **71%**, suggesting that a significant group of students performed **well** in the test.
- The **median MBA percentage** is **62%**, again reflecting an **average performance**.
- The **median salary** is **₹2,65,000**, which is lower than the mean salary, indicating the presence of **higher salary outliers** in the dataset.

Since the median is less affected by extreme values, it provides a more **realistic representation** of the typical student's performance and salary.

🔷 Mode Analysis

The mode identifies the most frequently occurring values:

- **SSLC mode**: 62%

- **HSC mode**: 63%
- **Degree mode**: 65%
- **E-test mode**: 60%
- **MBA mode**: 56.7%
- **Salary mode**: ₹3,00,000

This indicates that most students tend to score around these values, and the most common salary package offered is **₹3 LPA**.

🔷 Overall Conclusion

From the analysis using measures of central tendency:

- The dataset represents a group of students who have generally **average academic performance**.
- **Mean** includes all values, including extreme scores (outliers).
- **Median** provides a better estimate of typical performance by excluding the influence of outliers.
- **Mode** helps identify the most common scores and salary offered.

Overall, this analysis shows that the batch consists predominantly of **average-performing students**, with comparatively better performance in the E-test and a commonly offered salary package of **₹3 LPA**.

MEASURE OF LOCATION OF DATA

Percentile – Placement dataset Percentile analysis report:

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
Median	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	62.0	63.0	65.0	60.0	56.7	300000.0
Q1:25%	60.6	60.9	61.0	60.0	57.945	240000.0
Q2:50%	67.0	65.0	66.0	71.0	62.0	265000.0
Q3:75%	75.7	73.0	72.0	83.5	66.255	300000.0
99%	87.0	91.86	83.86	97.0	76.1142	NaN
Q4:100%	89.4	97.7	91.0	98.0	77.89	940000.0

1. SSLC (Secondary School) Marks Analysis

The SSLC marks distribution shows a **moderate and consistent performance trend** among students. Most students scored between **60 and 75 marks**, indicating average academic performance at the school level. The top-performing students scored between **87 and 89 marks**, and there is no evidence of extreme outliers. Overall, SSLC scores demonstrate stable and balanced student performance.

2. HSC (Higher Secondary) Marks Analysis

HSC marks display a **wider performance variation** compared to SSLC marks. While the majority of students scored between **60 and 73 marks**, a smaller group of students achieved significantly higher scores ranging from **91 to 97 marks**. This indicates the presence of high-performing students and a stronger differentiation in performance levels.

3. Degree Marks Analysis

Degree marks show a **uniform and stable performance distribution**. Most students scored between **61 and 83 marks**, with very few students scoring as high as **91 marks**. The data spread indicates consistent academic evaluation and minimal extreme variations across students.

4. Employability Test (E-test) Analysis

The E-test marks demonstrate a **larger performance spread**, suggesting varied employability skill levels among students. Most students scored between **60 and 83 marks**, while top performers scored between **97 and 98 marks**. The close grouping among top scorers indicates strong competition and skill similarity at higher performance levels.

5. MBA Marks Analysis

MBA marks exhibit a **tightly clustered distribution**, with most students scoring between **57 and 76 marks**. The smaller data spread suggests relatively uniform performance among students, indicating standardized evaluation patterns or similar academic outcomes.

6. Salary Distribution Analysis

The salary distribution is **highly skewed**, with the majority of students receiving placement packages between **2 to 3 LPA**. A small number of students secured premium salary packages up to **9.4 LPA**, indicating the presence of high-salary outliers. This suggests that while most placements fall within a moderate salary range, only a few students achieve significantly higher compensation.

Overall Insights

- Academic performance across SSLC, Degree, and MBA levels shows relatively consistent and moderate distribution.
- HSC and E-test results highlight stronger performance differentiation among students.
- Salary distribution indicates that most students receive moderate placement packages, with only a few achieving premium placements.
- Employability test scores suggest notable variation in practical and skill-based competencies among students.

Inter Quartile Range (IQR Method) – Placement dataset Outlier analysis report:

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
Median	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	62.0	63.0	65.0	60.0	56.7	300000.0
Q1:25%	60.6	60.9	61.0	60.0	57.945	240000.0
Q2:50%	67.0	65.0	66.0	71.0	62.0	265000.0
Q3:75%	75.7	73.0	72.0	83.5	66.255	300000.0
99%	87.0	91.86	83.86	97.0	76.1142	NaN
Q4:100%	89.4	97.7	91.0	98.0	77.89	940000.0
IQR	15.1	12.1	11.0	23.5	8.31	60000.0
1.5_rule	22.65	18.15	16.5	35.25	12.465	90000.0
least_outlier	37.95	42.75	44.5	24.75	45.48	150000.0
greatest_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

To identify outliers in the placement dataset, the Interquartile Range (IQR) method was applied. The lower and upper outlier boundaries were calculated using the 1.5 × IQR rule.

SSLC Marks

The calculated lower boundary is **37.95** and the upper boundary is **98.35**.
The minimum score (40.89) and maximum score (89.4) fall within this range.

Therefore, no outliers are present in SSLC marks.

HSC Marks

The lower boundary is **42.75** and the upper boundary is **91.15**.
The minimum score (37) is below the lower boundary and the maximum score (97) exceeds the upper boundary.

This indicates the presence of both lower and upper outliers in HSC marks.

Degree Marks

The lower boundary is **44.5** and the upper boundary is **88.5**.
The minimum score (50) lies within the boundary, while the maximum score (91) exceeds the upper boundary.

Degree marks contain an upper outlier.

Employability Test (E-Test) Marks

The lower boundary is **24.75** and the upper boundary is **118.75**.
Both the minimum (50) and maximum (98) values fall within the boundary.

No outliers are detected in E-test marks.

MBA Marks

The lower boundary is **45.48** and the upper boundary is **78.72**.
Both minimum (51) and maximum (77) values fall within the acceptable range.

MBA marks do not contain outliers.

Salary

The lower boundary is **1,50,000** and the upper boundary is **3,90,000**.
The maximum salary value of **9,40,000** exceeds the upper boundary.

Salary column contains significant upper outliers, indicating the presence of high-paying placement offers received by a small number of students.

Overall Conclusion

Most academic performance columns show stable score distributions with minimal outliers. However, HSC and Degree marks contain limited extreme values. The salary distribution contains significant upper outliers, suggesting unequal salary distribution where a small proportion of students receive premium placement packages.

BEFORE REPLACING OUTLIER

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
Median	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	62.0	63.0	65.0	60.0	56.7	300000.0
Q1:25%	60.6	60.9	61.0	60.0	57.945	240000.0
Q2:50%	67.0	65.0	66.0	71.0	62.0	265000.0
Q3:75%	75.7	73.0	72.0	83.5	66.255	300000.0
99%	87.0	91.86	83.86	97.0	76.1142	NaN
Q4:100%	89.4	97.7	91.0	98.0	77.89	940000.0
IQR	15.1	12.1	11.0	23.5	8.31	60000.0
1.5_rule	22.65	18.15	16.5	35.25	12.465	90000.0
lower_bound	37.95	42.75	44.5	24.75	45.48	150000.0
upper_bound	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

AFTER REPLACING OUTLIER

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.303395	66.334744	66.358558	72.100558	62.278186	277648.648649
Median	67.0	65.0	66.0	71.0	62.0	265000.0
Mode	62.0	63.0	65.0	60.0	56.7	300000.0
Q1:25%	60.6	60.9	61.0	60.0	57.945	240000.0
Q2:50%	67.0	65.0	66.0	71.0	62.0	265000.0
Q3:75%	75.7	73.0	72.0	83.5	66.255	300000.0
99%	87.0	91.129	83.86	97.0	76.1142	NaN
Q4:100%	89.4	91.15	88.5	98.0	77.89	390000.0
IQR	15.1	12.1	11.0	23.5	8.31	60000.0
1.5_rule	22.65	18.15	16.5	35.25	12.465	90000.0
lower_bound	37.95	42.75	44.5	24.75	45.48	150000.0
upper_bound	98.35	91.15	88.5	118.75	78.72	390000.0
min	40.89	42.75	50.0	50.0	51.21	200000.0
max	89.4	91.15	88.5	98.0	77.89	390000.0

Outliers were treated using the IQR method. In the HSC pass marks column, both lower and upper outliers were handled. In the degree pass marks and salary columns, only upper outliers were treated. As a result of this outlier treatment, minor variations were observed across the percentile distribution (99^{th} & 100^{th} percentile).