Kurtosis and Skewness Report – Placement Dataset

In this analysis, I studied the shape of the data distribution for each numerical column in the placement dataset. To do this, I calculated two statistical measures:

- **Skewness**: This tells us whether the data is symmetrical or if it leans more to one side (left or right).
- **Kurtosis**: This tells us whether the data is peaked or flat compared to a normal bell curve.

Code Used

To calculate skewness and kurtosis for each column, I used the following Python code inside a loop:

```
descriptive[columnName]["Kurtosis"] = dataset[columnName].kurtosis()

descriptive[columnName]["skew"] = dataset[columnName].skew()
```

This added both skewness and kurtosis values to my descriptive statistics table.

Table

Statistic	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Skew	0.000	−0.133	0.163	0.204	0.282	0.314	0.807
Kurtosis	−1.200	−0.608	0.087	−0.097	−1.089	−0.471	−0.240

What These Results Mean

- Skewness shows if the values are spread equally or more on one side:
  - Skew close to 0 means the distribution is balanced.
  - Skew > 0 means the tail is longer on the right (more lower values, few high ones).
  - Skew < 0 means the tail is on the left (more higher values, few low ones).
- Kurtosis tells if the graph is pointy or flat:
  - If kurtosis < 0 → the curve is flat (platykurtic)
  - If kurtosis ≈ 0 → the curve is normal
  - If kurtosis > 0 → the curve is peaked (leptokurtic)

My Observations

- Most columns had kurtosis less than zero, which means their distributions are more flat than normal.
- Columns like ssc_p, degree_p, and mba_p are almost symmetrical with low skewness.
- salary had the highest skew (0.807) → This means a few students had very high salaries, which pulls the graph to the right.

- etest_p was also moderately right-skewed and had a very flat shape, which means students scored across a wide range.
- sl_no had a perfect skew of 0, which makes sense because it's just a serial number and not a real variable.

Conclusion

By calculating skewness and kurtosis with Python, I was able to understand how each column is distributed without needing graphs. These statistics helped me spot:

- Columns that are symmetrical (like sl_no)
- Columns that have a long right tail (like salary)
- Columns that are spread out more than normal (flat-shaped)

This is a helpful step in data analysis, especially before building models or doing data transformations.