

# DATA SCIENCE – UNIVARIATE ANALYSIS

## Histogram, Skewness, Kurtosis

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
<b>Mean</b>	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
<b>Median</b>	67.0	65.0	66.0	71.0	62.0	265000.0
<b>Mode</b>	62.0	63.0	65.0	60.0	56.7	300000.0
<b>Q1:25%</b>	60.6	60.9	61.0	60.0	57.945	240000.0
<b>Q2:50%</b>	67.0	65.0	66.0	71.0	62.0	265000.0
<b>Q3:75%</b>	75.7	73.0	72.0	83.5	66.255	300000.0
<b>Q4:100%</b>	89.4	97.7	91.0	98.0	77.89	940000.0
<b>IQR</b>	15.1	12.1	11.0	23.5	8.31	60000.0
<b>1.5Rule</b>	22.65	18.15	16.5	35.25	12.465	90000.0
<b>Lesser</b>	37.95	42.75	44.5	24.75	45.48	150000.0
<b>Greater</b>	98.35	91.15	88.5	118.75	78.72	390000.0
<b>Min</b>	40.89	37.0	50.0	50.0	51.21	200000.0
<b>Max</b>	89.4	97.7	91.0	98.0	77.89	940000.0
<b>kurtosis</b>	-0.60751	0.450765	0.052143	-1.08858	-0.470723	18.544273
<b>skew</b>	-0.132649	0.163639	0.244917	0.282308	0.313576	3.569747
<b>Q5:99%</b>	87.0	91.86	83.86	97.0	76.1142	671200.0

# **UNDERSTANDING HISTOGRAM, SKEWNESS, AND KURTOSIS:-**

## **HISTOGRAM:**

**A histogram visually represents the distribution of a dataset by grouping values into bins. It helps to see if the data is normally distributed, skewed, or has outliers.**

## **SKEWNESS:**

**Skewness measures the asymmetry of data distribution. If skewness is near zero, the data is symmetric. Positive skewness indicates a longer right tail, while negative skewness means a longer left tail.**

## **KURTOSIS:**

Kurtosis measures the "tailedness" of the distribution. A high kurtosis indicates heavy tails (more outliers), whereas low kurtosis suggests a flatter distribution.

## **Kurtosis Analysis:**

1. The kurtosis value for salary (18.544273) is significantly high, indicating a heavy-tailed distribution with extreme values (outliers).
2. The other variables (ssc\_p, hsc\_p, degree\_p, etest\_p, mba\_p) have kurtosis values close to zero, meaning they follow a near-normal distribution.

3. The etest\_p (-1.08858) and mba\_p (-0.470723) have negative kurtosis, indicating a platykurtic (flat) distribution with fewer extreme values.
4. A comparison shows that salary has the most variation in data, likely due to differences in experience and job roles.

### **Skewness Analysis:**

1. Salary (3.569747) has the highest skewness, indicating a highly right-skewed distribution, meaning most salaries are concentrated at the lower end, with a few extreme high values.

2. All other variables have skewness values close to zero, suggesting a nearly symmetrical distribution of academic percentages.
3. Etest\_p (0.282308) and mba\_p (0.313576) have a slight positive skew, indicating a minor right tail but still close to normal distribution.
4. Ssc\_p (-0.132649) has a small negative skew, suggesting slightly more values on the higher end but not significantly.

## **Conclusion:**

- The salary column has the highest kurtosis and skewness, meaning it has many extreme

values and is highly right-skewed. The academic scores have kurtosis near zero, showing a normal distribution, with slight skewness in some variables. Etest\_p and mba\_p show mild positive skewness, while ssc\_p is slightly negatively skewed. The presence of high kurtosis in salary suggests that a few individuals earn significantly higher than the average, creating outliers. Understanding these statistical properties helps analyze distribution patterns and detect anomalies in the dataset.

