Kurtosis

Kurtosis tells us how much of the data is in the tails of a distribution — in other words, how many extreme values (outliers) are present.

There are **three types of kurtosis**:

1. **Mesokurtic (Normal Kurtosis = 3.0)**

- This is the standard shape — like a normal bell curve.
- The data is evenly spread, with a moderate number of outliers.
- Example: Heights of adults in a population.

2. **Leptokurtic (Kurtosis > 3.0)**

- The curve is taller and sharper in the middle, with **more data in the tails**.
- This means **more outliers** — values far from the average.
- Example: Stock market returns (lots of extreme highs and lows).

3. **Platykurtic (Kurtosis < 3.0)**

- The curve is flatter and wider.
- **Fewer outliers** — most data is close to the average.
- Example: Test scores where most students scored similarly.

If our dataset:
From ssc_p to salary, all columns have kurtosis **less than 3**, so they are **Platykurtic** — meaning the data is spread out with **few extreme values**.

Skewness

Skewness tells us whether the data is **symmetrical** or **tilted** to one side.

There are **three types of skewness**:

1. **Positive Skewness**

- The **right tail** is longer — more high values.
- Mean and median are **greater than** the mode.
- Example: Income distribution — a few people earn a lot more than the rest.

2. **Negative Skewness**

- The **left tail** is longer — more low values.
- Mean and median are **less than** the mode.
- Example: Age at retirement — most retire around 60, but a few retire much earlier.

3. **Zero Skewness**

- The distribution is **perfectly symmetrical**.
- Mean = Median = Mode.
- Example: Idealized normal distribution.

If our dataset:

- ssc_p has a skewness of **-0.13** → this is **slightly negatively skewed**, meaning the data leans a bit to the left.
- All other columns have skewness close to **0**, which means they are **normally distributed** — balanced and symmetrical.

Histogram

A **histogram** is a type of graph that shows how your data is spread across different value ranges. It helps you understand the overall shape of the data — whether it's balanced, tilted to one side, or has extreme values.

In a histogram:

- The **horizontal axis** (x-axis) shows value ranges, like scores from 0 to 100 split into groups.
- The **vertical axis** (y-axis) shows how many data points fall into each group.
- Each **bar** represents how common values are in that range — taller bars mean more data in that group.

You can use a histogram to quickly see:

- Where most of the data is concentrated.
- Whether the data is symmetrical or skewed.
- If there are any gaps or extreme values (outliers).

If our Dataset:

- If the histogram for ssc_p shows more values on the right and a longer tail on the left, it matches the **negative skewness** you observed.
- If other columns have bars that are evenly spread and centered, it supports the idea of a **normal distribution**.
- If the bars are wide and flat, it matches the **platykurtic** pattern — meaning fewer extreme values and more data near the average.