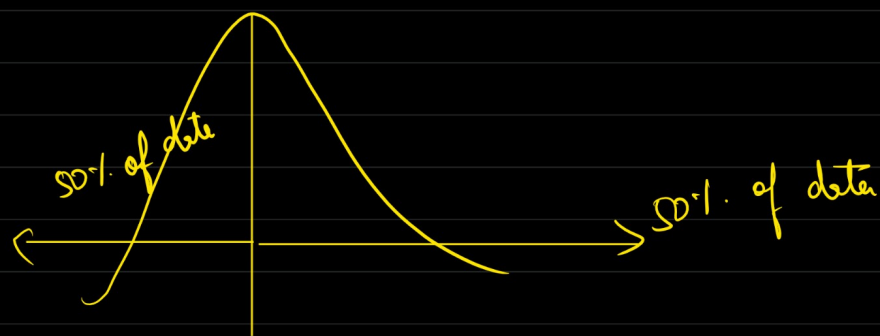
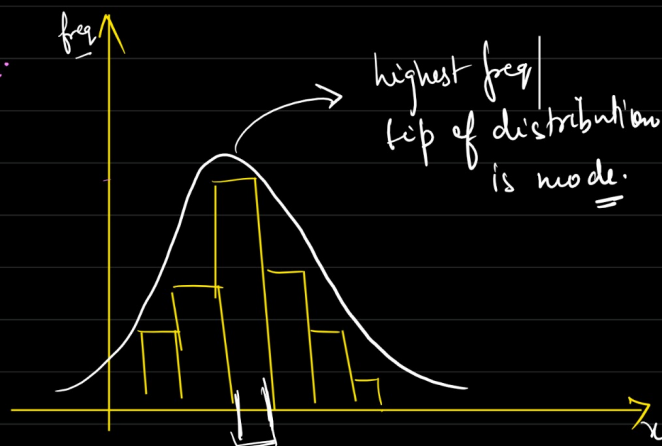


* Measures of Symmetry

↓
Shape of data.

→ If anything is exactly toward the left and right.



Skewness → measure of symmetry.

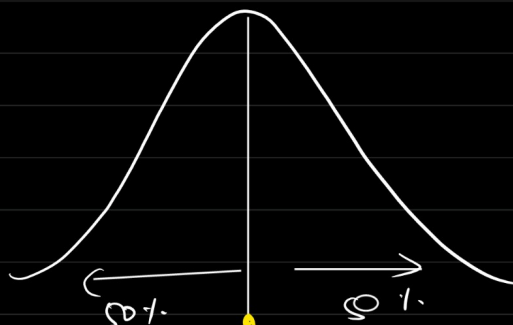
⇓

Inclined to some side.

① No Skewness (no inclination | symmetric)

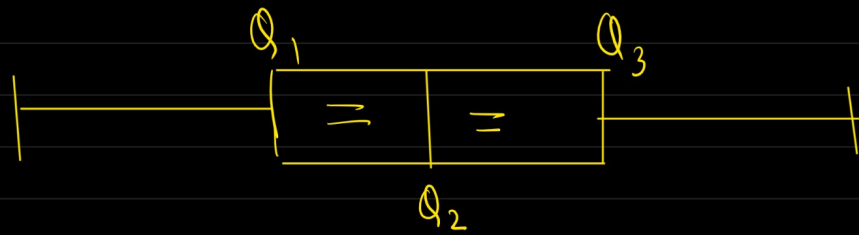
Skewness = 0

$$\text{Skewness} = \frac{\sum (x_i - \mu)^3}{n\sigma^3}$$



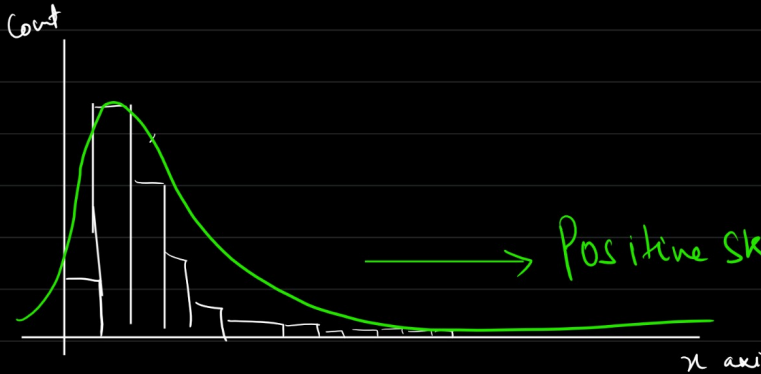
mode = median = mean

eg 1 2 3 4 5
 └─┬─┘ └─┬─┘
 ↑
 mean = median = mode



$$Q_3 - Q_2 \approx Q_2 - Q_1$$

②

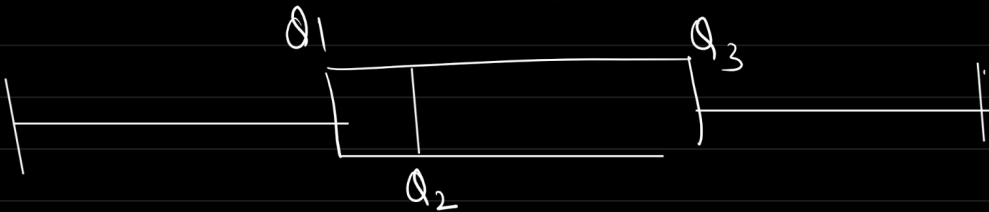


→ Positive skewed data

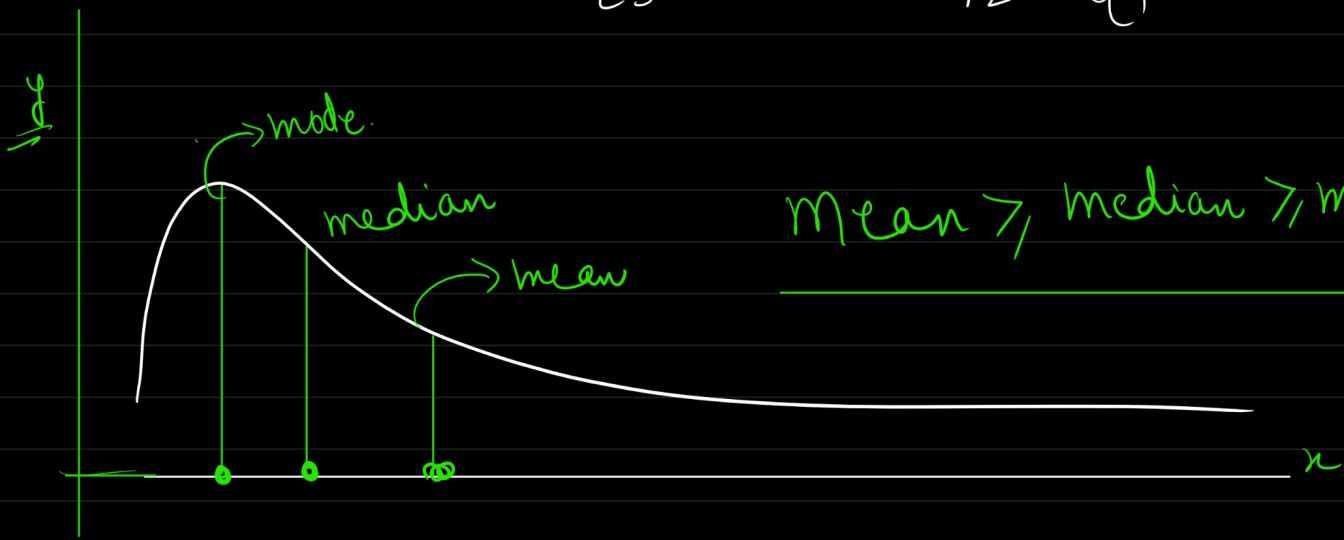
→ tail is on right side of distribution

→ Most of the data in the distribution lies on right side

ex. Log-Normal distⁿ

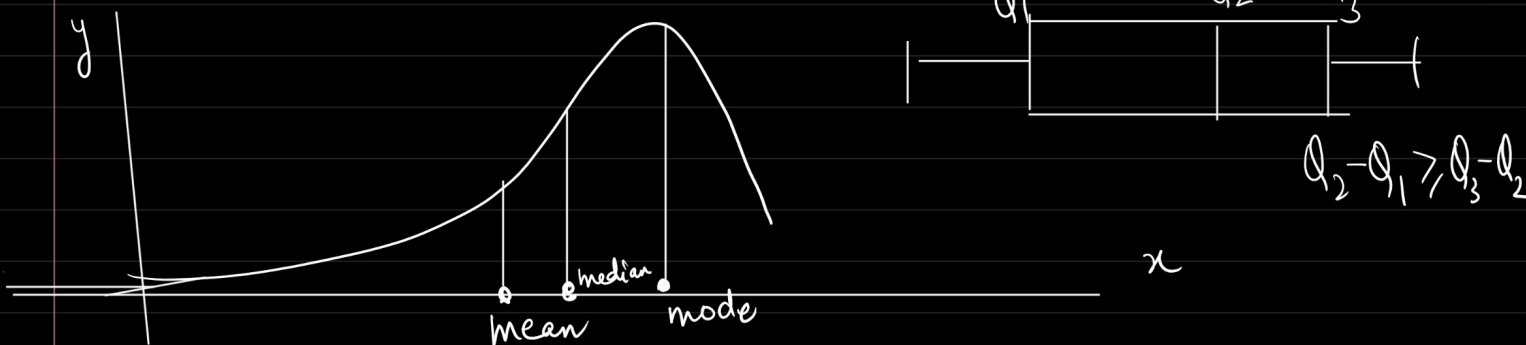


$$Q_3 - Q_2 > Q_2 - Q_1$$



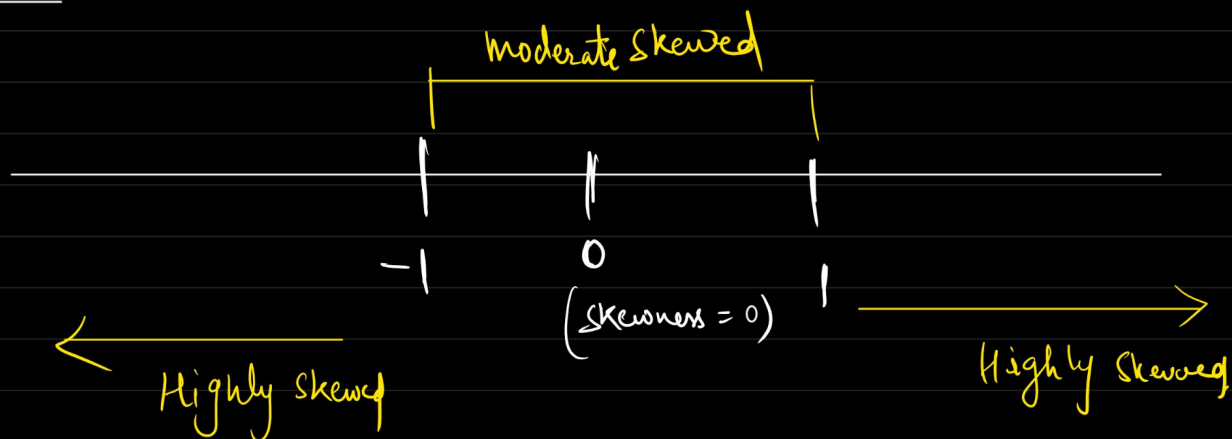
$$\text{Mean} > \text{median} > \text{mode}$$

③ Left skewed distribution / negative skewed data.



$$\text{mode} > \text{median} > \text{mean}$$

Use Case



To know if data is skewed or not?

- Visualisation (distribution plot)
- Skewness.

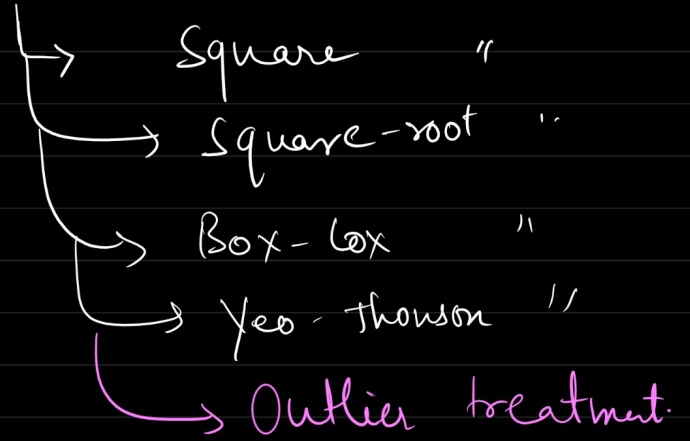
Use Case

Some of ML algorithms requires symmetric data to build the ML model.

if not symmetric :-

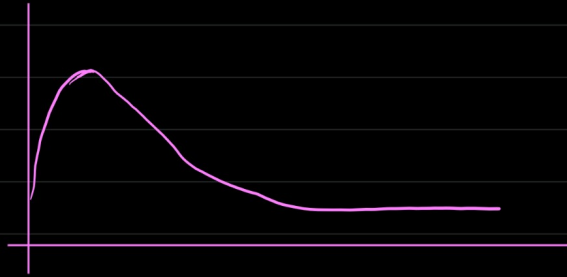
- * Transformation
 - Log transformation
 - Exponential transformation
 - reciprocal "

Why data is skewed



→ Nature of data

eg. tickets sold after a movie release.



→ Presence of outlier → To accomodate outlier, the distribution becomes skewed.

