Function Transformer

A part of handling skewed or non-linear data in machine learning preprocessing.

➢ What is a Function Transformer?

A **FunctionTransformer** in sklearn lets you apply any custom or mathematical transformation (like log, square root, reciprocal) directly on your data **as part of a pipeline**. It's useful when your data isn't linear or is skewed, and you want to make it behave more "normally."

➢ What are Mathematical Transformations?

These are functions we apply to **transform the distribution** of features (columns), often to:

- Reduce **skewness**
- Improve **model performance**
- Help algorithms that assume normality (like linear regression)

➢ Common Transformations Explained:

Transformation	Formula	Use Case
Log Transform	log(x)	Handles **right-skewed** data
Reciprocal Transform	1/x	Useful when large values dominate
Square Transform	x^2	Can help **left-skewed** data
Square Root Transform	$\sqrt{x}$	Less aggressive than log

These require positive values—cannot be applied directly to negative or zero values!

➢ How to Know if Data is Normal?

You can check if your data follows a **normal (bell curve) distribution** using:

- **Histograms**
- **QQ Plots** (Quantile-Quantile plots) — they compare your data's distribution to a normal distribution

---

➢ When to Use FunctionTransformer?

- As a **step in your pipeline**
- When **custom transformations** are needed
- When you're preprocessing **specific columns**

➢ Example from Titanic Dataset:

In that dataset:

- You applied transformations on columns like *Fare* (highly skewed)
- Used log and reciprocal transforms to make the distribution closer to normal
- This helps models make better predictions

➢ Final Thoughts:

The **FunctionTransformer** is powerful when your data needs mathematical tweaking. It's fully compatible with pipelines and ColumnTransformer, helping your models work better with “cleaner” data distributions.
