Target guided ordinal encoding:

Colab:

https://colab.research.google.com/drive/1Ok6FKIck5eO3uvHXI5NI_Vwa0azXMPkr#scrollTo=PIfHuQZhBBx2

Why do we need Target-Guided Ordinal Encoding?

So far you've seen:

- **Label Encoding** → arbitrary numbers ❌ (can create false order)
- **Ordinal Encoding** → order is **manually defined** ✅

But what if:

- The feature **does not have a natural order**, AND
- One-Hot Encoding creates **too many columns**, AND
- We want the **order to come from the data itself**?

👉 That's where **Target-Guided Ordinal Encoding** comes in.

---

2. Core idea (in one sentence)

Categories are ordered based on their relationship with the target variable.

The encoding is **guided by the target (y)**, not by human judgment.

---

3. When is it useful?

Use Target-Guided Ordinal Encoding when:

- The feature is **categorical**
- There is **no obvious natural order**
- The feature has **many unique categories**
- You are doing **supervised learning**

Typical examples:

- City vs house price

- Job title vs salary
- Education institute vs admission probability

---

4. Simple example (from scratch)

Dataset

City	Price (Target)
Delhi	100
Mumbai	200
Delhi	120
Chennai	90
Mumbai	220
Chennai	95

---

5. Step-by-step process 🧠

Step 1: Compute target statistic per category

(usually **mean**, sometimes median)

City	Mean Price
Chennai	92.5
Delhi	110
Mumbai	210

Step 2: Sort categories by target value

Chennai < Delhi < Mumbai

This order is **learned from data**, not assumed.

Step 3: Assign ordinal numbers

City	Encoded
Chennai	0

City	Encoded
Delhi	1
Mumbai	2

Now the numbers **carry meaning**.