■ Notes on Categorical Encoding

■ Step 1: Identify the Type of Categorical Variable

Type	Meaning	Example
Nominal (no order)	Categories are names without ranking	Fuel Type: petrol, diesel, electric, hybrid
Ordinal (has order)	Categories have a natural order	Size: small, medium, large

■ Step 2: Choose Encoding Method Based on Type & Situation

Situation	Best Encoding	Reas
Nominal (no order)	One Hot Encoding (OHE)	Treats all categories equally
Ordinal (has order)	Label Encoding / Ordinal Encoding	Preserves the natu
Many categories (>10 or >20)	Target / Frequency Encoding	Prevents too many
Tree-based models (RandomForest, XGBoost)	Label Encoding	Trees can handle c
Linear models (Linear / Logistic Regression)	One Hot Encoding	Avoids false numeri

■ Step 3: Apply It to Real Cases

Case A: No Natural Order
Example: Engine Fuel Type → ['regular unleaded', 'premium unleaded', 'diesel', 'electric']
Use: One Hot Encoding
Example code:

```
pd.get_dummies(df, columns=['Engine Fuel Type'])
```

Case B: Has Natural Order
Example: Size → ['Small', 'Medium', 'Large', 'Extra Large']
Use: Label or Ordinal Encoding
Example code:

```
from sklearn.preprocessing import OrdinalEncoder
encoder = OrdinalEncoder(categories=[['Small', 'Medium', 'Large', 'Extra
Large']])
df['Size_encoded'] = encoder.fit_transform(df[['Size']])
```

■ Step 4: Summary Table

No. of Unique Values	Variable Type	Encoding Type	Example Use
2	Nominal	Binary / Label Encoding	Gender
3–10	Nominal	One Hot Encoding	Fuel Type
>10	Nominal	Target / Frequency Encoding	Country Names
Ordered values	Ordinal	Label / Ordinal Encoding	Size, Rating

■ Final Answer for Your Case

Since your column has 10 unique, non-ordered categories, use One Hot Encoding (OHE). It’s safe, interpretable, and avoids adding false order.