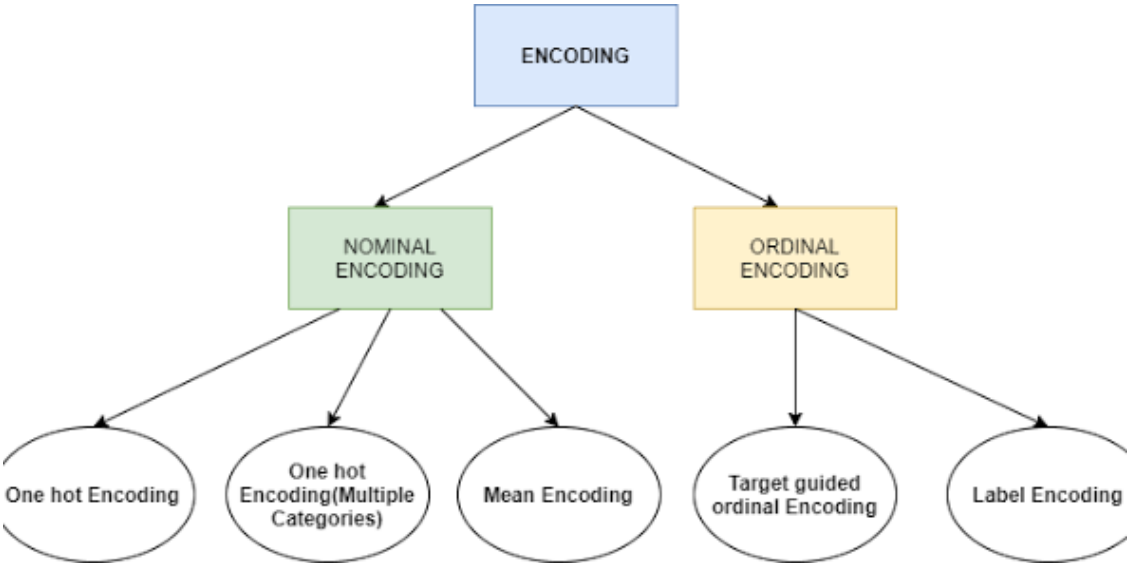
# onehot

February 9, 2026




[1]:
```
import pandas as pd
```

[ ]:
```

```

[2]:
```
df=pd.read_csv("home_prices.csv")
```

[3]:
```
df
```

[3]:

| | locality | area_sqr_ft | price_lakhs | bedrooms |
|---|---|---|---|---|
| 0 | Kollur | 656 | 39.0 | 2 |
| 1 | Kollur | 1260 | 83.2 | 2 |
| 2 | Kollur | 1057 | 86.6 | 3 |
| 3 | Kollur | 1259 | 59.0 | 2 |
| 4 | Kollur | 1800 | 140.0 | 3 |
| 5 | Kollur | 1325 | 80.1 | 2 |
| 6 | Kollur | 1085 | 116.0 | 3 |
| 7 | Kollur | 1110 | 45.0 | 2 |
| 8 | Kollur | 1700 | 100.0 | 3 |
| 9 | Banjara Hills | 1650 | 200.0 | 3 |
| 10 | Banjara Hills | 2438 | 316.0 | 3 |
| 11 | Banjara Hills | 2115 | 220.0 | 2 |

```
12  Banjara Hills         1600        150.0         3
13  Banjara Hills         2400        300.0         3
14  Banjara Hills         1100         85.0         2
15  Banjara Hills         2600        400.0         3
16        Mankhal         1100         54.0         2
17        Mankhal         1125         64.0         2
18        Mankhal         1008         50.0         2
19        Mankhal         1266         78.0         2
20        Mankhal         1540         94.0         3
21        Mankhal         1200         85.0         3
```

[4]:
```
pd.get_dummies(df,columns=["locality"])
```

[4]:
```
    area_sqr_ft  price_lakhs  bedrooms  locality_Banjara Hills  \
0           656         39.0         2                   False
1          1260         83.2         2                   False
2          1057         86.6         3                   False
3          1259         59.0         2                   False
4          1800        140.0         3                   False
5          1325         80.1         2                   False
6          1085        116.0         3                   False
7          1110         45.0         2                   False
8          1700        100.0         3                   False
9          1650        200.0         3                    True
10         2438        316.0         3                    True
11         2115        220.0         2                    True
12         1600        150.0         3                    True
13         2400        300.0         3                    True
14         1100         85.0         2                    True
15         2600        400.0         3                    True
16         1100         54.0         2                   False
17         1125         64.0         2                   False
18         1008         50.0         2                   False
19         1266         78.0         2                   False
20         1540         94.0         3                   False
21         1200         85.0         3                   False

    locality_Kollur  locality_Mankhal
0              True             False
1              True             False
2              True             False
3              True             False
4              True             False
5              True             False
6              True             False
7              True             False
8              True             False
```

```
9             False             False
10            False             False
11            False             False
12            False             False
13            False             False
14            False             False
15            False             False
16            False              True
17            False              True
18            False              True
19            False              True
20            False              True
21            False              True
```

[5]:
```
df_encoded=pd.get_dummies(df,columns=["locality"],drop_first=True)
```

[6]:
```
df_encoded
```

[6]:

| | area_sqr_ft | price_lakhs | bedrooms | locality_Kollur | locality_Mankhal |
|---|---|---|---|---|---|
| 0 | 656 | 39.0 | 2 | True | False |
| 1 | 1260 | 83.2 | 2 | True | False |
| 2 | 1057 | 86.6 | 3 | True | False |
| 3 | 1259 | 59.0 | 2 | True | False |
| 4 | 1800 | 140.0 | 3 | True | False |
| 5 | 1325 | 80.1 | 2 | True | False |
| 6 | 1085 | 116.0 | 3 | True | False |
| 7 | 1110 | 45.0 | 2 | True | False |
| 8 | 1700 | 100.0 | 3 | True | False |
| 9 | 1650 | 200.0 | 3 | False | False |
| 10 | 2438 | 316.0 | 3 | False | False |
| 11 | 2115 | 220.0 | 2 | False | False |
| 12 | 1600 | 150.0 | 3 | False | False |
| 13 | 2400 | 300.0 | 3 | False | False |
| 14 | 1100 | 85.0 | 2 | False | False |
| 15 | 2600 | 400.0 | 3 | False | False |
| 16 | 1100 | 54.0 | 2 | False | True |
| 17 | 1125 | 64.0 | 2 | False | True |
| 18 | 1008 | 50.0 | 2 | False | True |
| 19 | 1266 | 78.0 | 2 | False | True |
| 20 | 1540 | 94.0 | 3 | False | True |
| 21 | 1200 | 85.0 | 3 | False | True |

[7]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

[8]:
```
X=df_encoded.drop("price_lakhs",axis=1)
y= df_encoded["price_lakhs"]
```

[9]:
```
# train- test - split
X_train,X_test,y_train,y_test= train_test_split(X,y,test_size=0.
 ↪2,random_state=42)
model=LinearRegression()
model.fit(X_train,y_train)
```

[9]:
```
LinearRegression()
```

[14]:
```
model.score(X_test,y_test)
```

[14]:
```
0.855890526315538
```

[10]:
```
test = pd.DataFrame([
    {'area_sqr_ft': 1600, "bedrooms": 2, "locality_Kollur": False,␣
 ↪"locality_Mankhal": False},
    {'area_sqr_ft': 1600, "bedrooms": 2, "locality_Kollur": False,␣
 ↪"locality_Mankhal": True},
])

model.predict(test)
```

[10]:
```
array([157.03383393, 109.25104283])
```

[ ]:
```

```