

## encoding

February 9, 2026

```
[7]: import pandas as pd  
from sklearn.preprocessing import OneHotEncoder
```

```
[8]: df = pd.read_csv("home_prices.csv")  
print(df.head())
```

	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

```
[9]: cat_cols = ['locality']
```

```
[10]: encoder = OneHotEncoder(sparse_output=False, handle_unknown='ignore')
```

```
[11]: encoded_array = encoder.fit_transform(df[cat_cols])
```

```
[13]: encoded_df = pd.DataFrame(  
    encoded_array,  
    columns=encoder.get_feature_names_out(cat_cols))
```

```
[14]: encoded_df.index = df.index
```

```
[15]: df = df.drop(cat_cols, axis=1)
```

```
[16]: df = pd.concat([df, encoded_df], axis=1)
```

```
[19]: df.head()
```

```
[19]: area_sqr_ft  price_lakhs  bedrooms  locality_Banjara Hills  \  
 0          656        39.0       2            0.0  
 1         1260        83.2       2            0.0  
 2         1057        86.6       3            0.0  
 3         1259        59.0       2            0.0  
 4         1800       140.0       3            0.0  
  
locality_Kollur  locality_Mankhal
```

```
0          1.0      0.0  
1          1.0      0.0  
2          1.0      0.0  
3          1.0      0.0  
4          1.0      0.0
```

```
[21]: df.drop(columns=["locality_Banjara Hills"], inplace=True)
```

```
[22]: df.head()
```

```
[22]:   area_sqr_ft  price_lakhs  bedrooms  locality_Kollur  locality_Mankhal  
0          656        39.0       2           1.0          0.0  
1         1260        83.2       2           1.0          0.0  
2         1057        86.6       3           1.0          0.0  
3         1259        59.0       2           1.0          0.0  
4         1800       140.0       3           1.0          0.0
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```