label encoding and on hot encoding

June 20, 2023

```
[3]: import pandas as pd
     df=pd.read_csv(r"C:\Users\Rakesh\Downloads\homeprice.csv")
     df
```

```
[3]:     Unnamed: 0            town  area   price
     0            0  monroe township  2600  550000
     1            1  monroe township  3000  565000
     2            2  monroe township  3200  610000
     3            3  monroe township  3600  680000
     4            4  monroe township  4000  725000
     5            5     west windsor  2600  585000
     6            6     west windsor  2800  615000
     7            7     west windsor  3300  650000
     8            8     west windsor  3600  710000
     9            9      robinsville  2600  575000
     10          10      robinsville  2900  600000
     11          11      robinsville  3100  620000
     12          12      robinsville  3600  695000
```

```
[5]: df1=df.drop(['Unnamed: 0'],axis=1)
```

```
[6]: df1
```

```
[6]:             town  area   price
     0   monroe township  2600  550000
     1   monroe township  3000  565000
     2   monroe township  3200  610000
     3   monroe township  3600  680000
     4   monroe township  4000  725000
     5      west windsor  2600  585000
     6      west windsor  2800  615000
     7      west windsor  3300  650000
     8      west windsor  3600  710000
     9       robinsville  2600  575000
     10      robinsville  2900  600000
     11      robinsville  3100  620000
     12      robinsville  3600  695000
```

```python
[7]: from sklearn.preprocessing import LabelEncoder
     le=LabelEncoder()
```

```python
[9]: df1.town=le.fit_transform(df1.town)
```

```python
[11]: df1
```

```
[11]:      town  area   price
     0       0  2600  550000
     1       0  3000  565000
     2       0  3200  610000
     3       0  3600  680000
     4       0  4000  725000
     5       2  2600  585000
     6       2  2800  615000
     7       2  3300  650000
     8       2  3600  710000
     9       1  2600  575000
     10      1  2900  600000
     11      1  3100  620000
     12      1  3600  695000
```

```python
[66]: x=df1.drop(['price'],axis=1).values
      x
```

```
[66]: array([[   0, 2600],
             [   0, 3000],
             [   0, 3200],
             [   0, 3600],
             [   0, 4000],
             [   2, 2600],
             [   2, 2800],
             [   2, 3300],
             [   2, 3600],
             [   1, 2600],
             [   1, 2900],
             [   1, 3100],
             [   1, 3600]], dtype=int64)
```

```python
[67]: y=df1['price'].values
      y
```

```
[67]: array([550000, 565000, 610000, 680000, 725000, 585000, 615000, 650000,
             710000, 575000, 600000, 620000, 695000], dtype=int64)
```

```python
[68]: #now we use one hot encoder
```

```python
[69]: from sklearn.preprocessing import OneHotEncoder
      OHE=OneHotEncoder()
      from sklearn.compose import ColumnTransformer
      ct = ColumnTransformer([('town', OneHotEncoder(), [0])], remainder =␣
        ↪'passthrough')
```

```python
[70]: x = ct.fit_transform(x)
      x
```

```
[70]: array([[1.0e+00, 0.0e+00, 0.0e+00, 2.6e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 3.0e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 3.2e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 3.6e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 4.0e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 2.6e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 2.8e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 3.3e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 3.6e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 2.6e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 2.9e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 3.1e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 3.6e+03]])
```

```python
[ ]: 
```

```python
[71]: x
```

```
[71]: array([[1.0e+00, 0.0e+00, 0.0e+00, 2.6e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 3.0e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 3.2e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 3.6e+03],
             [1.0e+00, 0.0e+00, 0.0e+00, 4.0e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 2.6e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 2.8e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 3.3e+03],
             [0.0e+00, 0.0e+00, 1.0e+00, 3.6e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 2.6e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 2.9e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 3.1e+03],
             [0.0e+00, 1.0e+00, 0.0e+00, 3.6e+03]])
```

```python
[72]: x=x[:,1:]
      x
```

```
[72]: array([[0.0e+00, 0.0e+00, 2.6e+03],
             [0.0e+00, 0.0e+00, 3.0e+03],
             [0.0e+00, 0.0e+00, 3.2e+03],
             [0.0e+00, 0.0e+00, 3.6e+03],
```

```
       [0.0e+00, 0.0e+00, 4.0e+03],
       [0.0e+00, 1.0e+00, 2.6e+03],
       [0.0e+00, 1.0e+00, 2.8e+03],
       [0.0e+00, 1.0e+00, 3.3e+03],
       [0.0e+00, 1.0e+00, 3.6e+03],
       [1.0e+00, 0.0e+00, 2.6e+03],
       [1.0e+00, 0.0e+00, 2.9e+03],
       [1.0e+00, 0.0e+00, 3.1e+03],
       [1.0e+00, 0.0e+00, 3.6e+03]])
```

[ ]:

[73]:
```python
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

[74]:
```python
model.fit(x,y)
```

[74]: LinearRegression()

[75]:
```python
model.predict([[0,1,3400]])
```

[75]: array([681241.6684584])

[ ]: