COLUMN TRANSFORMER

June 21, 2023

[53]:

```
import pandas as pd
import numpy as np
```

[54]:

```
df=pd.read_csv(r"C:\Users\Rakesh\Downloads\covid.csv")
df
```

[54]:

```
    age  gender  fever   cough       city has_covid
0    60    Male  103.0    Mild    Kolkata        No
1    27    Male  100.0    Mild      Delhi       Yes
2    42    Male  101.0    Mild      Delhi        No
3    31  Female   98.0    Mild    Kolkata        No
4    65  Female  101.0    Mild     Mumbai        No
..    …      …     …      …          …       …
95   12  Female  104.0    Mild  Bangalore        No
96   51  Female  101.0  Strong    Kolkata       Yes
97   20  Female  101.0    Mild  Bangalore        No
98    5  Female   98.0  Strong     Mumbai        No
99   10  Female   98.0  Strong    Kolkata       Yes

[100 rows x 6 columns]
```

[55]:

```
df.isnull().sum()
```

[55]:

```
age           0
gender        0
fever        10
cough         0
city          0
has_covid     0
dtype: int64
```

[56]:

```
#gender-one hot encoding
#cough-ordinal encoding
#fever-simple inputer
#age-min max scaler(but not applying here in this problem)
```

[57]:

```
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import OrdinalEncoder
```

[58]:
```
si=SimpleImputer()
OHE=OneHotEncoder()
OE=OrdinalEncoder()
```

[59]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(df.
 ↪drop(['has_covid'],axis=1),df['has_covid'],test_size=0.2)
x_train
```

[59]:
```
    age  gender  fever   cough       city
61   81  Female   98.0  Strong     Mumbai
17   40  Female   98.0  Strong      Delhi
30   15    Male  101.0    Mild      Delhi
24   13  Female  100.0  Strong    Kolkata
80   14  Female   99.0    Mild     Mumbai
..  …       …     …      …         …
23   80  Female   98.0    Mild      Delhi
25   23    Male    NaN    Mild     Mumbai
77    8  Female  101.0    Mild    Kolkata
27   33  Female  102.0  Strong      Delhi
47   18  Female  104.0    Mild  Bangalore

[80 rows x 5 columns]
```

[60]:
```
x_train.shape
```

[60]:
```
(80, 5)
```

[61]:
```
from sklearn.compose import ColumnTransformer
ct = ColumnTransformer(transformers=[
    ('n1',SimpleImputer(),['fever']),
    ('n2',OneHotEncoder(sparse=False,drop='first'),['gender','city']),
    ('n3',OrdinalEncoder(categories=[['Mild','Strong']]),['cough'])
],remainder='passthrough')
```

[62]:
```
ct.fit_transform(x_train).shape
```

[62]:
```
(80, 7)
```

[63]:
```
x_trainF1=ct.fit_transform(x_train)
x_trainF1
```

[63]:
```
array([[ 98.        ,  0.        ,  0.        ,  0.        ,
          1.        ,  1.        , 81.        ],
       [ 98.        ,  0.        ,  1.        ,  0.        ,
          0.        ,  1.        , 40.        ],
       [101.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  15.        ],
       [100.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   1.        ,  13.        ],
       [ 99.        ,   0.        ,   0.        ,   0.        ,
          1.        ,   0.        ,  14.        ],
       [101.04166667,   1.        ,   0.        ,   1.        ,
          0.        ,   1.        ,  71.        ],
       [101.04166667,   0.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  42.        ],
       [100.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,   5.        ],
       [ 99.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  65.        ],
       [101.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  47.        ],
       [100.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  27.        ],
       [101.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  14.        ],
       [ 98.        ,   1.        ,   0.        ,   0.        ,
          1.        ,   1.        ,  23.        ],
       [101.04166667,   0.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  84.        ],
       [ 99.        ,   0.        ,   1.        ,   0.        ,
          0.        ,   1.        ,  59.        ],
       [101.04166667,   0.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  75.        ],
       [103.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  69.        ],
       [101.        ,   0.        ,   0.        ,   0.        ,
          1.        ,   0.        ,  81.        ],
       [100.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  80.        ],
       [101.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  42.        ],
       [100.        ,   0.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  19.        ],
       [104.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   1.        ,  54.        ],
       [104.        ,   0.        ,   1.        ,   0.        ,
          0.        ,   1.        ,  34.        ],
       [101.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  83.        ],
       [103.        ,   1.        ,   0.        ,   1.        ,
          0.        ,   1.        ,  70.        ],
       [100.        ,   1.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  27.        ],
       [ 99.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  66.        ],
       [ 99.        ,   0.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  22.        ],
       [101.04166667,   0.        ,   0.        ,   0.        ,
          1.        ,   1.        ,  20.        ],
       [100.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  27.        ],
       [103.        ,   1.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  83.        ],
       [ 98.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  26.        ],
       [104.        ,   1.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  16.        ],
       [100.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  11.        ],
       [ 99.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  65.        ],
       [104.        ,   1.        ,   0.        ,   0.        ,
          1.        ,   0.        ,  44.        ],
       [102.        ,   0.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  69.        ],
       [101.        ,   0.        ,   1.        ,   0.        ,
          0.        ,   1.        ,  68.        ],
       [ 99.        ,   0.        ,   0.        ,   0.        ,
          1.        ,   0.        ,  60.        ],
       [104.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  51.        ],
       [100.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  19.        ],
       [103.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  50.        ],
       [101.        ,   0.        ,   0.        ,   0.        ,
          1.        ,   0.        ,  19.        ],
       [101.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  19.        ],
       [102.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   1.        ,  82.        ],
       [104.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,  17.        ],
       [100.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   1.        ,  11.        ],
       [104.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  25.        ],
       [104.        ,   0.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  56.        ],
       [101.        ,   0.        ,   1.        ,   0.        ,
         0.         ,   0.        ,  49.        ],
       [104.        ,   0.        ,   1.        ,   0.        ,
         0.         ,   1.        ,  75.        ],
       [103.        ,   0.        ,   1.        ,   0.        ,
         0.         ,   0.        ,  73.        ],
       [ 98.        ,   1.        ,   0.        ,   0.        ,
         0.         ,   0.        ,  73.        ],
       [101.        ,   0.        ,   1.        ,   0.        ,
         0.         ,   1.        ,  34.        ],
       [101.        ,   0.        ,   0.        ,   0.        ,
         0.         ,   0.        ,  38.        ],
       [ 98.        ,   1.        ,   1.        ,   0.        ,
         0.         ,   0.        ,  83.        ],
       [100.        ,   0.        ,   0.        ,   0.        ,
         0.         ,   1.        ,  47.        ],
       [ 99.        ,   0.        ,   0.        ,   1.        ,
         0.         ,   1.        ,  25.        ],
       [103.        ,   0.        ,   0.        ,   1.        ,
         0.         ,   0.        ,  48.        ],
       [103.        ,   0.        ,   0.        ,   0.        ,
         0.         ,   0.        ,  16.        ],
       [ 98.        ,   1.        ,   0.        ,   0.        ,
         0.         ,   1.        ,  12.        ],
       [102.        ,   0.        ,   0.        ,   0.        ,
         0.         ,   1.        ,  82.        ],
       [101.04166667,   1.        ,   1.        ,   0.        ,
         0.         ,   0.        ,  38.        ],
       [ 98.        ,   0.        ,   0.        ,   0.        ,
         1.         ,   0.        ,  65.        ],
       [104.        ,   1.        ,   0.        ,   1.        ,
         0.         ,   0.        ,  51.        ],
       [ 98.        ,   1.        ,   0.        ,   1.        ,
         0.         ,   1.        ,  34.        ],
       [ 98.        ,   0.        ,   0.        ,   0.        ,
         0.         ,   0.        ,  64.        ],
       [101.        ,   0.        ,   0.        ,   0.        ,
         1.         ,   0.        ,  65.        ],
       [101.04166667,   1.        ,   0.        ,   1.        ,
         0.         ,   0.        ,  82.        ],
       [104.        ,   1.        ,   0.        ,   0.        ,
         1.         ,   0.        ,  42.        ],
       [101.        ,   0.        ,   0.        ,   1.        ,
         0.         ,   1.        ,  51.        ],
       [104.        ,   0.        ,   0.        ,   0.        ,
         0.         ,   0.        ,  12.        ],
       [102.        ,   1.        ,   1.        ,   0.        ,
          0.        ,   1.        ,  20.        ],
       [103.        ,   1.        ,   0.        ,   0.        ,
          0.        ,   1.        ,  46.        ],
       [104.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,   6.        ],
       [ 98.        ,   0.        ,   1.        ,   0.        ,
          0.        ,   0.        ,  80.        ],
       [101.04166667,   1.        ,   0.        ,   0.        ,
          1.        ,   0.        ,  23.        ],
       [101.        ,   0.        ,   0.        ,   1.        ,
          0.        ,   0.        ,   8.        ],
       [102.        ,   0.        ,   1.        ,   0.        ,
          0.        ,   1.        ,  33.        ],
       [104.        ,   0.        ,   0.        ,   0.        ,
          0.        ,   0.        ,  18.        ]])
```

[ ]:

[ ]: