2203a51521-ml1

March 4, 2024

[2]:

```python
import pandas as pd
housing=pd.read_csv('/content/housing.csv')
print(housing)
```

```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
0        -122.23     37.88                41.0        880.0           129.0
1        -122.22     37.86                21.0       7099.0          1106.0
2        -122.24     37.85                52.0       1467.0           190.0
3        -122.25     37.85                52.0       1274.0           235.0
4        -122.25     37.85                52.0       1627.0           280.0
…            …         …                   …            …               …
20635    -121.09     39.48                25.0       1665.0           374.0
20636    -121.21     39.49                18.0        697.0           150.0
20637    -121.22     39.43                17.0       2254.0           485.0
20638    -121.32     39.43                18.0       1860.0           409.0
20639    -121.24     39.37                16.0       2785.0           616.0

       population  households  median_income  median_house_value  \
0           322.0       126.0         8.3252            452600.0
1          2401.0      1138.0         8.3014            358500.0
2           496.0       177.0         7.2574            352100.0
3           558.0       219.0         5.6431            341300.0
4           565.0       259.0         3.8462            342200.0
…             …           …             …                   …
20635       845.0       330.0         1.5603             78100.0
20636       356.0       114.0         2.5568             77100.0
20637      1007.0       433.0         1.7000             92300.0
20638       741.0       349.0         1.8672             84700.0
20639      1387.0       530.0         2.3886             89400.0

      ocean_proximity
0            NEAR BAY
1            NEAR BAY
2            NEAR BAY
3            NEAR BAY
4            NEAR BAY
…                 …
20635          INLAND
```

```
20636         INLAND
20637         INLAND
20638         INLAND
20639         INLAND

[20640 rows x 10 columns]
```

[4]:
```
housing.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   longitude           20640 non-null  float64
 1   latitude            20640 non-null  float64
 2   housing_median_age  20640 non-null  float64
 3   total_rooms         20640 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20640 non-null  float64
 6   households          20640 non-null  float64
 7   median_income       20640 non-null  float64
 8   median_house_value  20640 non-null  float64
 9   ocean_proximity     20640 non-null  object
dtypes: float64(9), object(1)
memory usage: 1.6+ MB
```

[5]:
```
housing.describe()
```

[5]:
```
          longitude      latitude  housing_median_age   total_rooms  \
count  20640.000000  20640.000000        20640.000000  20640.000000
mean    -119.569704     35.631861           28.639486   2635.763081
std        2.003532      2.135952           12.585558   2181.615252
min     -124.350000     32.540000            1.000000      2.000000
25%     -121.800000     33.930000           18.000000   1447.750000
50%     -118.490000     34.260000           29.000000   2127.000000
75%     -118.010000     37.710000           37.000000   3148.000000
max     -114.310000     41.950000           52.000000  39320.000000

       total_bedrooms    population    households  median_income  \
count    20433.000000  20640.000000  20640.000000   20640.000000
mean       537.870553   1425.476744    499.539680       3.870671
std        421.385070   1132.462122    382.329753       1.899822
min          1.000000      3.000000      1.000000       0.499900
25%        296.000000    787.000000    280.000000       2.563400
50%        435.000000   1166.000000    409.000000       3.534800
75%        647.000000   1725.000000    605.000000       4.743250
```

```
max       6445.000000  35682.000000   6082.000000      15.000100

       median_house_value
count         20640.000000
mean         206855.816909
std          115395.615874
min           14999.000000
25%          119600.000000
50%          179700.000000
75%          264725.000000
max          500001.000000
```

[6]:
```python
housing['households'].shape
```

[6]: (20640,)

[22]:
```python
z=housing.drop('ocean_proximity',axis=1)
print(z)
```

```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
0        -122.23     37.88                41.0        880.0           129.0
1        -122.22     37.86                21.0       7099.0          1106.0
2        -122.24     37.85                52.0       1467.0           190.0
3        -122.25     37.85                52.0       1274.0           235.0
4        -122.25     37.85                52.0       1627.0           280.0
…            …         …                   …            …              …
20635    -121.09     39.48                25.0       1665.0           374.0
20636    -121.21     39.49                18.0        697.0           150.0
20637    -121.22     39.43                17.0       2254.0           485.0
20638    -121.32     39.43                18.0       1860.0           409.0
20639    -121.24     39.37                16.0       2785.0           616.0

       population  households  median_income  median_house_value
0           322.0       126.0         8.3252            452600.0
1          2401.0      1138.0         8.3014            358500.0
2           496.0       177.0         7.2574            352100.0
3           558.0       219.0         5.6431            341300.0
4           565.0       259.0         3.8462            342200.0
…             …           …             …                  …
20635       845.0       330.0         1.5603             78100.0
20636       356.0       114.0         2.5568             77100.0
20637      1007.0       433.0         1.7000             92300.0
20638       741.0       349.0         1.8672             84700.0
20639      1387.0       530.0         2.3886             89400.0

[20640 rows x 9 columns]
```

[24]:
```python
print(housing.isnull().sum())
```

```
longitude               0
latitude                0
housing_median_age      0
total_rooms             0
total_bedrooms        207
population              0
households              0
median_income           0
median_house_value      0
ocean_proximity         0
dtype: int64
```

[12]:
```python
housing['total_bedrooms'].isnull().sum()/housing.shape[0]
```

[12]: 0.01002906976744186

[25]:
```python
y=housing['median_house_value']
x=housing.drop('median_house_value',axis=1)

print(x)
```

```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
0        -122.23     37.88                41.0        880.0           129.0
1        -122.22     37.86                21.0       7099.0          1106.0
2        -122.24     37.85                52.0       1467.0           190.0
3        -122.25     37.85                52.0       1274.0           235.0
4        -122.25     37.85                52.0       1627.0           280.0
…            …         …                   …            …              …
20635    -121.09     39.48                25.0       1665.0           374.0
20636    -121.21     39.49                18.0        697.0           150.0
20637    -121.22     39.43                17.0       2254.0           485.0
20638    -121.32     39.43                18.0       1860.0           409.0
20639    -121.24     39.37                16.0       2785.0           616.0

       population  households  median_income ocean_proximity
0           322.0       126.0         8.3252        NEAR BAY
1          2401.0      1138.0         8.3014        NEAR BAY
2           496.0       177.0         7.2574        NEAR BAY
3           558.0       219.0         5.6431        NEAR BAY
4           565.0       259.0         3.8462        NEAR BAY
…             …           …             …             …
20635       845.0       330.0         1.5603          INLAND
20636       356.0       114.0         2.5568          INLAND
20637      1007.0       433.0         1.7000          INLAND
20638       741.0       349.0         1.8672          INLAND
20639      1387.0       530.0         2.3886          INLAND
```

```
[20640 rows x 9 columns]
```

[15]:

```
print(x.shape)
print(y.shape)
```

```
(20640, 9)
(20640,)
```

[16]:

```
from sklearn.model_selection import train_test_split
x_train,y_train,x_test,y_test=train_test_split(x,y,test_size=0.
 ↪30,random_state=40)
print(x_train)
```

```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
14827    -117.09     32.66                37.0       1232.0           330.0
12640    -121.45     38.53                34.0       1893.0           415.0
6016     -117.78     34.06                33.0       1056.0           272.0
13650    -117.31     34.08                37.0        953.0           231.0
13275    -117.63     34.10                15.0       4799.0          1209.0
…              …       …                   …            …               …
11532    -118.09     33.77                26.0       5359.0          1508.0
16065    -122.48     37.75                49.0       2203.0           407.0
14501    -117.23     32.86                16.0       1200.0           468.0
14555    -117.13     32.96                15.0       2267.0           292.0
11590    -118.01     33.78                26.0       2343.0           377.0

       population  households  median_income ocean_proximity
14827      1086.0       330.0         1.6389      NEAR OCEAN
12640       884.0       395.0         2.1679          INLAND
6016        964.0       300.0         2.4464          INLAND
13650       611.0       230.0         1.9926          INLAND
13275      2554.0      1057.0         2.6582          INLAND
…             …           …             …               …
11532      1829.0      1393.0         1.7675       <1H OCEAN
16065      1052.0       405.0         4.4375        NEAR BAY
14501       648.0       443.0         3.0450      NEAR OCEAN
14555      1180.0       289.0         6.7120       <1H OCEAN
11590      1166.0       373.0         6.0000       <1H OCEAN

[14448 rows x 9 columns]
```

[17]:

```
print(x_test)
```

```
14827    114300.0
12640     75400.0
6016     128700.0
13650     81500.0
```

```
13275    122800.0
           …
11532     61300.0
16065    329200.0
14501    100000.0
14555    240200.0
11590    233100.0
Name: median_house_value, Length: 14448, dtype: float64
```

[18]:
```python
print(y_train)
```

```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
6607     -118.14     34.18                47.0       3457.0           622.0
884      -121.97     37.54                31.0       1949.0           344.0
9457     -123.84     39.83                19.0       1461.0           340.0
18681    -121.82     36.86                17.0       1573.0           272.0
4654     -118.33     34.05                48.0       2405.0           527.0
…            …       …                  …            …              …
3606     -118.48     34.23                29.0       3354.0           707.0
11607    -118.01     33.80                16.0       4021.0           701.0
12610    -121.51     38.54                34.0       2815.0           479.0
8561     -118.41     33.93                22.0       2514.0           605.0
203      -122.23     37.78                52.0        986.0           258.0

       population  households  median_income ocean_proximity
6607       1700.0       579.0         3.5164       <1H OCEAN
884         986.0       322.0         4.6349       <1H OCEAN
9457        515.0       227.0         1.5278      NEAR OCEAN
18681       142.0        55.0         2.1719      NEAR OCEAN
4654       1868.0       502.0         3.3750       <1H OCEAN
…             …           …             …              …
3606       1752.0       650.0         4.5484       <1H OCEAN
11607      1488.0       650.0         5.3200       <1H OCEAN
12610      1075.0       471.0         3.9792          INLAND
8561       1225.0       568.0         4.1818       <1H OCEAN
203        1008.0       255.0         1.4844        NEAR BAY

[6192 rows x 9 columns]
```

[19]:
```python
print(y_test)
```

```
6607     226500.0
884      196200.0
9457     145800.0
18681    420000.0
4654     257800.0
           …
3606     239900.0
```

```
11607    219500.0
12610    164800.0
8561     339700.0
203      119400.0
Name: median_house_value, Length: 6192, dtype: float64
```

[20]:

```
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)
```

```
(14448, 9)
(6192, 9)
(14448,)
(6192,)
```

[23]:

```
from sklearn.preprocessing import MinMaxScaler
d=MinMaxScaler()
dd=d.fit_transform(z)
print(dd)
```

```
[[0.21115538 0.5674814  0.78431373 … 0.02055583 0.53966842 0.90226638]
 [0.21215139 0.565356   0.39215686 … 0.18697583 0.53802706 0.70824656]
 [0.21015936 0.5642933  1.         … 0.02894261 0.46602805 0.69505074]
 …
 [0.31175299 0.73219979 0.31372549 … 0.07104095 0.08276438 0.15938285]
 [0.30179283 0.73219979 0.33333333 … 0.05722743 0.09429525 0.14371281]
 [0.30976096 0.72582359 0.29411765 … 0.08699227 0.13025338 0.15340349]]
```