

```
import pandas as pd
a=pd.read_csv("/housing.csv")
print(a)
```

```

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]

```
print("Datatype of each column:")
print(a.dtypes)
print("\nShape of the DataFrame:")
print(a.shape)
```

```
Datatype of each column:
longitude      float64
latitude      float64
housing_median_age  float64
total_rooms    float64
total_bedrooms  float64
population     float64
households     float64
median_income  float64
median_house_value  float64
ocean_proximity  object
dtype: object
```

```
Shape of the DataFrame:
(20640, 10)
```

```
null_values = a.isnull().sum()
print("Columns with null values and their counts:")
print(null_values[null_values > 0])
```

```
Columns with null values and their counts:
total_bedrooms    207
dtype: int64
```

```
b=a.fillna(0)
null_values = b.isnull().sum()
print("Columns with null values and their counts:")
print(null_values[null_values > 0])
```

```
Columns with null values and their counts:
Series([], dtype: int64)
```

```
print(b.head())
print(b.columns)
target_variable = 'median_house_value'
features = b.columns[b.columns != target_variable]
print('Target Variable:', target_variable)
print('Features:', features)

      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

      population  households  median_income  median_house_value  ocean_proximity
0         322.0         126.0          8.3252         452600.0      NEAR BAY
1        2401.0        1138.0          8.3014        358500.0      NEAR BAY
2         496.0         177.0          7.2574        352100.0      NEAR BAY
3         558.0         219.0          5.6431        341300.0      NEAR BAY
4         565.0         259.0          3.8462        342200.0      NEAR BAY
Index(['longitude', 'latitude', 'housing_median_age', 'total_rooms',
      'total_bedrooms', 'population', 'households', 'median_income',
      'median_house_value', 'ocean_proximity'],
      dtype='object')
Target Variable: median_house_value
Features: Index(['longitude', 'latitude', 'housing_median_age', 'total_rooms',
      'total_bedrooms', 'population', 'households', 'median_income',
      'ocean_proximity'],
      dtype='object')
```

```
y=b['median_house_value']
y

0         452600.0
1         358500.0
2         352100.0
3         341300.0
4         342200.0
...
20635        78100.0
20636        77100.0
20637        92300.0
20638        84700.0
20639        89400.0
Name: median_house_value, Length: 20640, dtype: float64
```

```
X=a.drop('median_house_value',axis=1)
X
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	NEAR BAY
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	NEAR BAY
...
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	INLAND
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	INLAND
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	INLAND
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	INLAND
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	INLAND

20640 rows x 9 columns

Next steps:

 [View recommended plots](#)

```
X['ocean_proximity'] = X['ocean_proximity'].replace({'NEAR BAY': 0, '<1H OCEAN': 1, 'INLAND':2, 'NEAR OCEAN':3, 'ISLAND':4})
X
```

3/7/24, 11:49 AM2203a52210(ass4).ipynb - Colaboratory

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	0
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	0
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	0
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	0
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	0
...
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	2
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	2
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	2
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	2
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	2

20640 rows × 9 columns

Next steps: [View recommended plots](#)

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3)

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
print("\nScaled data:")
print(pd.DataFrame(X_train_scaled, columns=X_train.columns).head())
```

Scaled data:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	\
0	0.589641	0.185972	0.960784	0.029888	0.040968	
1	0.601594	0.149841	0.823529	0.046218	0.055400	
2	0.711155	0.030818	0.529412	0.082286	0.082557	
3	0.666335	0.112646	0.450980	0.042682	0.035847	
4	0.561753	0.187035	0.372549	0.040189	0.028864	

	population	households	median_income	ocean_proximity
0	0.033437	0.044072	0.186866	0.25
1	0.029765	0.056076	0.191094	0.25
2	0.037305	0.084197	0.253838	0.75
3	0.021806	0.037658	0.439084	0.25
4	0.019507	0.034205	0.417125	0.25