```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

data=pd.read_csv("Housing.csv")
data
```

```
        price  area  bedrooms  bathrooms  stories mainroad guestroom
basement  \
0    13300000  7420         4          2        3      yes        no
no
1    12250000  8960         4          4        4      yes        no
no
2    12250000  9960         3          2        2      yes        no
yes
3    12215000  7500         4          2        2      yes        no
yes
4    11410000  7420         4          1        2      yes       yes
yes
..        ...   ...       ...        ...      ...      ...       ...
...
540   1820000  3000         2          1        1      yes        no
yes
541   1767150  2400         3          1        1       no        no
no
542   1750000  3620         2          1        1      yes        no
no
543   1750000  2910         3          1        1       no        no
no
544   1750000  3850         3          1        2      yes        no
no

    hotwaterheating airconditioning  parking prefarea furnishingstatus

0                no             yes        2      yes        furnished

1                no             yes        3       no        furnished

2                no              no        2      yes   semi-furnished

3                no             yes        3      yes        furnished

4                no             yes        2       no        furnished

..              ...             ...      ...      ...              ...

540              no              no        2       no      unfurnished

541              no              no        0       no   semi-furnished
```

```
542               no               no         0       no      unfurnished

543               no               no         0       no        furnished

544               no               no         0       no      unfurnished


[545 rows x 13 columns]
```

```
print("\nSample Data")
print(data.head(10))
```

```
Sample Data
      price   area  bedrooms  bathrooms  stories mainroad guestroom
basement  \
0  13300000   7420         4          2        3      yes        no
no
1  12250000   8960         4          4        4      yes        no
no
2  12250000   9960         3          2        2      yes        no
yes
3  12215000   7500         4          2        2      yes        no
yes
4  11410000   7420         4          1        2      yes       yes
yes
5  10850000   7500         3          3        1      yes        no
yes
6  10150000   8580         4          3        4      yes        no
no
7  10150000  16200         5          3        2      yes        no
no
8   9870000   8100         4          1        2      yes       yes
yes
9   9800000   5750         3          2        4      yes       yes
no

  hotwaterheating airconditioning  parking prefarea furnishingstatus
0              no              yes        2      yes        furnished
1              no              yes        3       no        furnished
2              no               no        2      yes   semi-furnished
3              no              yes        3      yes        furnished
4              no              yes        2       no        furnished
5              no              yes        2      yes   semi-furnished
6              no              yes        2      yes   semi-furnished
7              no               no        0       no      unfurnished
8              no              yes        2      yes        furnished
9              no              yes        1      yes      unfurnished
```

```
print("\nSample Data")
print(data.tail(10))
```

```
Sample Data
       price  area  bedrooms  bathrooms  stories mainroad guestroom
basement  \
535  2100000  3360         2          1        1      yes        no
no
536  1960000  3420         5          1        2       no        no
no
537  1890000  1700         3          1        2      yes        no
no
538  1890000  3649         2          1        1      yes        no
no
539  1855000  2990         2          1        1       no        no
no
540  1820000  3000         2          1        1      yes        no
yes
541  1767150  2400         3          1        1       no        no
no
542  1750000  3620         2          1        1      yes        no
no
543  1750000  2910         3          1        1       no        no
no
544  1750000  3850         3          1        2      yes        no
no

    hotwaterheating airconditioning  parking prefarea furnishingstatus

535              no              no        1       no      unfurnished

536              no              no        0       no      unfurnished

537              no              no        0       no      unfurnished

538              no              no        0       no      unfurnished

539              no              no        1       no      unfurnished

540              no              no        2       no      unfurnished

541              no              no        0       no   semi-furnished

542              no              no        0       no      unfurnished

543              no              no        0       no        furnished

544              no              no        0       no      unfurnished


print("Basic Information")
print(data.info())
```

```
Basic Information
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   price             545 non-null    int64
 1   area              545 non-null    int64
 2   bedrooms          545 non-null    int64
 3   bathrooms         545 non-null    int64
 4   stories           545 non-null    int64
 5   mainroad          545 non-null    object
 6   guestroom         545 non-null    object
 7   basement          545 non-null    object
 8   hotwaterheating   545 non-null    object
 9   airconditioning   545 non-null    object
 10  parking           545 non-null    int64
 11  prefarea          545 non-null    object
 12  furnishingstatus  545 non-null    object
dtypes: int64(6), object(7)
memory usage: 55.5+ KB
None
```

```
data.dtypes
```

```
price                int64
area                 int64
bedrooms             int64
bathrooms            int64
stories              int64
mainroad            object
guestroom           object
basement            object
hotwaterheating     object
airconditioning     object
parking              int64
prefarea            object
furnishingstatus    object
dtype: object
```

```
data.drop(columns='prefarea',inplace=True)
```

```
data.columns
```

```
Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories',
'mainroad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
       'parking', 'furnishingstatus'],
      dtype='object')
```

```
print("\nSummary Statistics")
print(data.describe())
```

```
Summary Statistics
              price          area    bedrooms   bathrooms     stories
\
count  5.450000e+02    545.000000  545.000000  545.000000  545.000000

mean   4.766729e+06   5150.541284    2.965138    1.286239    1.805505

std    1.870440e+06   2170.141023    0.738064    0.502470    0.867492

min    1.750000e+06   1650.000000    1.000000    1.000000    1.000000

25%    3.430000e+06   3600.000000    2.000000    1.000000    1.000000

50%    4.340000e+06   4600.000000    3.000000    1.000000    2.000000

75%    5.740000e+06   6360.000000    3.000000    2.000000    2.000000

max    1.330000e+07  16200.000000    6.000000    4.000000    4.000000


          parking
count  545.000000
mean     0.693578
std      0.861586
min      0.000000
25%      0.000000
50%      0.000000
75%      1.000000
max      3.000000
```

```
data.isnull().sum()
```

```
price               0
area                0
bedrooms            0
bathrooms           0
stories             0
mainroad            0
guestroom           0
basement            0
hotwaterheating     0
airconditioning     0
parking             0
furnishingstatus    0
dtype: int64
```

```
data.dropna(inplace=True)
```

```
data.count()
```

```
price               545
area                545
bedrooms            545
bathrooms           545
stories             545
mainroad            545
guestroom           545
basement            545
hotwaterheating     545
airconditioning     545
parking             545
furnishingstatus    545
dtype: int64
```

```
duplicate_rows_df=data[data.duplicated()]
print("Number of duplicate rows: ",duplicate_rows_df.shape)
```

```
Number of duplicate rows:  (0, 12)
```

```
data.count()
```

```
price               545
area                545
bedrooms            545
bathrooms           545
stories             545
mainroad            545
guestroom           545
basement            545
hotwaterheating     545
airconditioning     545
parking             545
furnishingstatus    545
dtype: int64
```

```
data=data.drop_duplicates()
data.count()
```

```
price               545
area                545
bedrooms            545
bathrooms           545
stories             545
mainroad            545
guestroom           545
basement            545
hotwaterheating     545
airconditioning     545
parking             545
```

```
furnishingstatus    545
dtype: int64
```

```
features=data.columns
features
```

```
Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories',
'mainroad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
       'parking', 'furnishingstatus'],
      dtype='object')
```

```
zero_vals_cols=(data[features]==0).sum()
zero_vals_cols
```

```
price                 0
area                  0
bedrooms              0
bathrooms             0
stories               0
mainroad              0
guestroom             0
basement              0
hotwaterheating       0
airconditioning       0
parking             299
furnishingstatus      0
dtype: int64
```

```
data.isnull().sum()/len(data)*100
```

```
price               0.0
area                0.0
bedrooms            0.0
bathrooms           0.0
stories             0.0
mainroad            0.0
guestroom           0.0
basement            0.0
hotwaterheating     0.0
airconditioning     0.0
parking             0.0
furnishingstatus    0.0
dtype: float64
```

```
data[['price','parking']]=data[['price','parking']].replace(0,np.NaN)
```

```
data.area.fillna(data.area.median(),inplace=True)
```

```
one_hot_encoded=pd.get_dummies(data,columns=['mainroad'],prefix=['main
road'])
```

```
print("One-Hot Encoded Data: ")
print(one_hot_encoded)
```

```
One-Hot Encoded Data:
        price  area  bedrooms  bathrooms  stories guestroom
basement  \
0    13300000  7420         4          2        3        no        no

1    12250000  8960         4          4        4        no        no

2    12250000  9960         3          2        2        no       yes

3    12215000  7500         4          2        2        no       yes

4    11410000  7420         4          1        2       yes       yes

..        ...   ...       ...        ...      ...       ...       ...

540   1820000  3000         2          1        1        no       yes

541   1767150  2400         3          1        1        no        no

542   1750000  3620         2          1        1        no        no

543   1750000  2910         3          1        1        no        no

544   1750000  3850         3          1        2        no        no


    hotwaterheating airconditioning  parking furnishingstatus
mainroad_no  \
0                no             yes      2.0        furnished
0
1                no             yes      3.0        furnished
0
2                no              no      2.0   semi-furnished
0
3                no             yes      3.0        furnished
0
4                no             yes      2.0        furnished
0
..              ...             ...      ...              ...
...
540              no              no      2.0      unfurnished
0
541              no              no      NaN   semi-furnished
1
542              no              no      NaN      unfurnished
0
543              no              no      NaN        furnished
1
```

```
544               no               no      NaN      unfurnished
0


     mainroad_yes
0               1
1               1
2               1
3               1
4               1
..            ...
540             1
541             0
542             1
543             0
544             1

[545 rows x 13 columns]
```

```
from sklearn.preprocessing import LabelEncoder
label_encoder=LabelEncoder()
data['guestroom_LabelEncoded']=label_encoder.fit_transform(data['guest
room'])
print("\nLabel Encoded Data: ")
print(data)
```

```
Label Encoded Data:
        price  area  bedrooms  bathrooms  stories mainroad guestroom
basement  \
0    13300000  7420         4          2        3      yes        no
no
1    12250000  8960         4          4        4      yes        no
no
2    12250000  9960         3          2        2      yes        no
yes
3    12215000  7500         4          2        2      yes        no
yes
4    11410000  7420         4          1        2      yes       yes
yes
..        ...   ...       ...        ...      ...      ...       ...
...
540   1820000  3000         2          1        1      yes        no
yes
541   1767150  2400         3          1        1       no        no
no
542   1750000  3620         2          1        1      yes        no
no
543   1750000  2910         3          1        1       no        no
no
544   1750000  3850         3          1        2      yes        no
```

```
no

    hotwaterheating airconditioning  parking furnishingstatus  \
0                no             yes      2.0        furnished
1                no             yes      3.0        furnished
2                no              no      2.0   semi-furnished
3                no             yes      3.0        furnished
4                no             yes      2.0        furnished
..              ...             ...      ...              ...
540              no              no      2.0      unfurnished
541              no              no      NaN   semi-furnished
542              no              no      NaN      unfurnished
543              no              no      NaN        furnished
544              no              no      NaN      unfurnished

     guestroom_LabelEncoded
0                         0
1                         0
2                         0
3                         0
4                         1
..                      ...
540                       0
541                       0
542                       0
543                       0
544                       0

[545 rows x 13 columns]
```