```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        from sklearn.preprocessing import LabelEncoder

        df = pd.read_csv("Housing.csv")
        df.head()
```

Out[1]:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	f
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	

```
In [2]: df.isna().sum()
```

```
Out[2]: price               0
        area                0
        bedrooms            0
        bathrooms           0
        stories             0
        mainroad            0
        guestroom           0
        basement            0
        hotwaterheating     0
        airconditioning     0
        parking             0
        prefarea            0
        furnishingstatus    0
        dtype: int64
```

```
In [3]: df.info()
```

```
<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
```
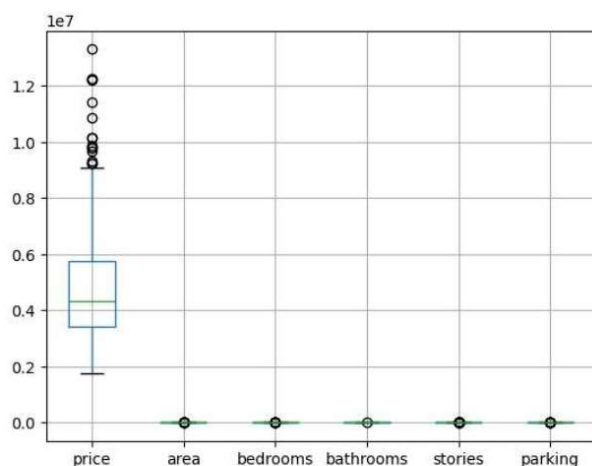
```
In [4]: df.describe()
```

Out[4]:

	price	area	bedrooms	bathrooms	stories	parking
count	5.450000e+02	545.000000	545.000000	545.000000	545.000000	545.000000
mean	4.766729e+06	5150.541284	2.965138	1.286239	1.805505	0.693578
std	1.870440e+06	2170.141023	0.738064	0.502470	0.867492	0.861586
min	1.750000e+06	1650.000000	1.000000	1.000000	1.000000	0.000000
25%	3.430000e+06	3600.000000	2.000000	1.000000	1.000000	0.000000
50%	4.340000e+06	4600.000000	3.000000	1.000000	2.000000	0.000000
75%	5.740000e+06	6360.000000	3.000000	2.000000	2.000000	1.000000
max	1.330000e+07	16200.000000	6.000000	4.000000	4.000000	3.000000

In [5]: `df.boxplot()`

Out[5]: <Axes: >

In [6]:
```python
Q1 = df['price'].quantile(0.25)
Q3 = df['price'].quantile(0.75)
iqr = Q3 - Q1
minm= Q1 - (1.5*iqr)
maxm = Q3 + (1.5*iqr)
df=df[(df['price']>minm) & (df['price']<maxm)]
df.head()
```

Out[6]:

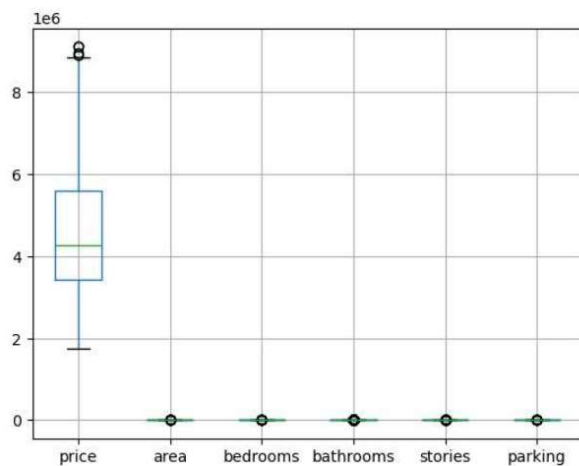|	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	f
15	9100000	6000	4	1	2	yes	no	yes	no	no	2	no	
16	9100000	6600	4	2	2	yes	yes	yes	no	yes	1	yes	
17	8960000	8500	3	2	4	yes	no	no	no	yes	2	no	
18	8890000	4600	3	2	2	yes	yes	no	no	yes	2	no	
19	8855000	6420	3	2	2	yes	no	no	no	yes	1	yes	

In [7]: `df.columns`

Out[7]:
```
Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
       'parking', 'prefarea', 'furnishingstatus'],
      dtype='object')
```

In [8]: `df.dtypes`

Out[8]:
```
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
```

```
In [9]: df.boxplot()
        plt.show()
```



```
In [10]: df.head()
```

Out[10]:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	fu
15	9100000	6000	4	1	2	yes	no	yes	no	no	2	no	
16	9100000	6600	4	2	2	yes	yes	yes	no	yes	1	yes	
17	8960000	8500	3	2	4	yes	no	no	no	yes	2	no	
18	8890000	4600	3	2	2	yes	yes	no	no	yes	2	no	
19	8855000	6420	3	2	2	yes	no	no	no	yes	1	yes	

```
In [11]: le = LabelEncoder()
         df['mainroad'] = le.fit_transform(df['mainroad'])
         df['guestroom'] = le.fit_transform(df['guestroom'])
         df['basement'] = le.fit_transform(df['basement'])
         df['hotwaterheating'] = le.fit_transform(df['hotwaterheating'])
         df['airconditioning'] = le.fit_transform(df['airconditioning'])
         df['furnishingstatus'] = le.fit_transform(df['furnishingstatus'])
         df['prefarea'] = le.fit_transform(df['prefarea'])

         df.head()
```

Out[11]:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	fu
15	9100000	6000	4	1	2	1	0	1	0	0	2	0	
16	9100000	6600	4	2	2	1	1	1	0	1	1	1	
17	8960000	8500	3	2	4	1	0	0	0	1	2	0	
18	8890000	4600	3	2	2	1	1	0	0	1	2	0	
19	8855000	6420	3	2	2	1	0	0	0	1	1	1	

```
In [ ]:
```