In [43]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
```

In [44]:
```
df = pd.read_csv('Real-Estate dataset.csv')
df.head()
```

Out[44]:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwate
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	

In [45]:
```
df.isna().sum()
```

Out[45]:
```
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 [46]:
```
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
```

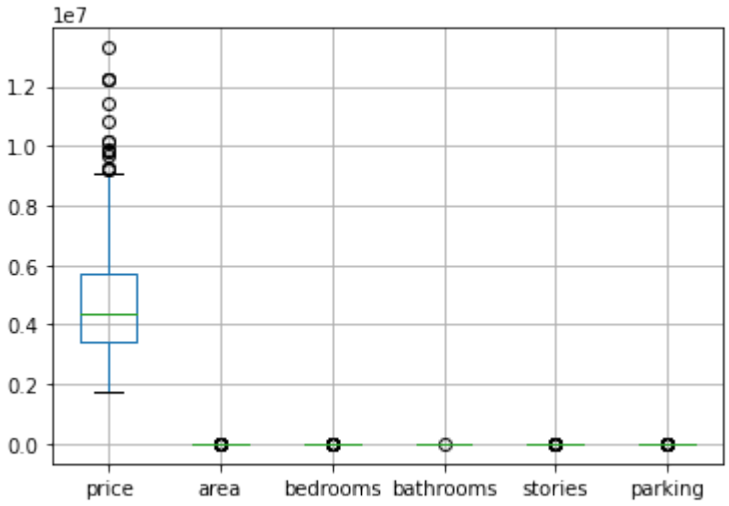
In [47]:
```
df.describe()
```

Out[47]:

	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 [48]:
```
df.boxplot()
```

Out[48]:
```
<AxesSubplot:>
```



In [49]:
```
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[49]:

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

In [50]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 530 entries, 15 to 544
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   price             530 non-null    int64 
 1   area              530 non-null    int64 
 2   bedrooms          530 non-null    int64 
 3   bathrooms         530 non-null    int64 
 4   stories           530 non-null    int64 
 5   mainroad          530 non-null    object
 6   guestroom         530 non-null    object
 7   basement          530 non-null    object
 8   hotwaterheating   530 non-null    object
 9   airconditioning   530 non-null    object
 10  parking           530 non-null    int64 
 11  prefarea          530 non-null    object
 12  furnishingstatus  530 non-null    object
dtypes: int64(6), object(7)
memory usage: 58.0+ KB
```

In [51]:
```
df.columns
```

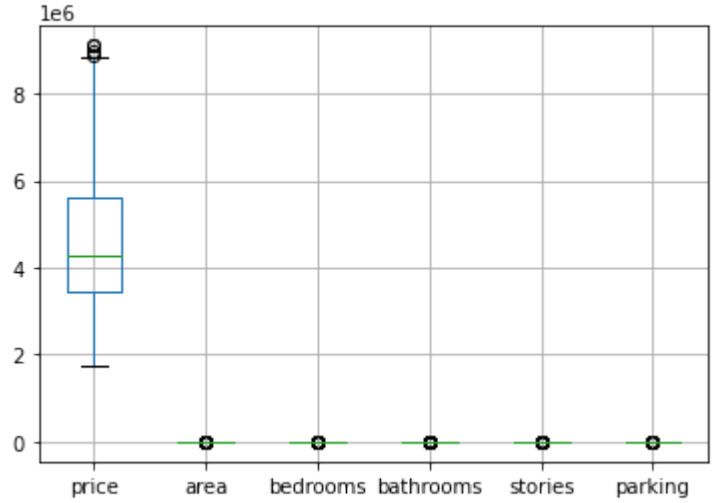
Out[51]:
```
Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
       'parking', 'prefarea', 'furnishingstatus'],
      dtype='object')
```

In [52]:
```
df.dtypes
```

Out[52]:
```
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 [53]:
```
df.boxplot()
```

Out[53]: <AxesSubplot:>



In [54]:
```
df.head()
```

Out[54]:

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

In [56]:
```
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['parking'] = le.fit_transform(df['parking'])
df['prefarea'] = le.fit_transform(df['prefarea'])

df.head()
```

Out[56]:

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

In [58]:
```
df = pd.get_dummies(df)
```

In [59]:
```
df.head()
```

Out[59]:

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

In [ ]:
```
```