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

import warnings
warnings.filterwarnings('ignore')
```

In [3]:
```
data=pd.read_csv('Housing.csv')
data
```

Out[3]:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating
0	13300000	7420	4	2	3	yes	no	no	no
1	12250000	8960	4	4	4	yes	no	no	no
2	12250000	9960	3	2	2	yes	no	yes	no
3	12215000	7500	4	2	2	yes	no	yes	no
4	11410000	7420	4	1	2	yes	yes	yes	no
5	10850000	7500	3	3	1	yes	no	yes	no
6	10150000	8580	4	3	4	yes	no	no	no
7	10150000	16200	5	3	2	yes	no	no	no
8	9870000	8100	4	1	2	yes	yes	yes	no
9	9800000	5750	3	2	4	yes	yes	no	no
10	9800000	13200	3	1	2	yes	no	yes	no
11	9681000	6000	4	3	2	yes	yes	yes	yes
12	9310000	6550	4	2	2	yes	no	no	no
13	9240000	3500	4	2	2	yes	no	no	yes
14	9240000	7800	3	2	2	yes	no	no	no
15	9100000	6000	4	1	2	yes	no	yes	no
16	9100000	6600	4	2	2	yes	yes	yes	no
17	8960000	8500	3	2	4	yes	no	no	no
18	8890000	4600	3	2	2	yes	yes	no	no
19	8855000	6420	3	2	2	yes	no	no	no
20	8750000	4320	3	1	2	yes	no	yes	yes
21	8680000	7155	3	2	1	yes	yes	yes	no
22	8645000	8050	3	1	1	yes	yes	yes	no
23	8645000	4560	3	2	2	yes	yes	yes	no
24	8575000	8800	3	2	2	yes	no	no	no
25	8540000	6540	4	2	2	yes	yes	yes	no
26	8463000	6000	3	2	4	yes	yes	yes	no
27	8400000	8875	3	1	1	yes	no	no	no
28	8400000	7950	5	2	2	yes	no	yes	yes
29	8400000	5500	4	2	2	yes	no	yes	no
...	...	...	...	...	...	...	...	...	...
515	2450000	3210	3	1	2	yes	no	yes	no
516	2450000	3240	2	1	1	no	yes	no	no
517	2450000	3000	2	1	1	yes	no	no	no
518	2450000	3500	2	1	1	yes	yes	no	no
519	2450000	4840	2	1	2	yes	no	no	no
520	2450000	7700	2	1	1	yes	no	no	no

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating
521	2408000	3635	2	1	1	no	no	no	no
522	2380000	2475	3	1	2	yes	no	no	no
523	2380000	2787	4	2	2	yes	no	no	no
524	2380000	3264	2	1	1	yes	no	no	no
525	2345000	3640	2	1	1	yes	no	no	no
526	2310000	3180	2	1	1	yes	no	no	no
527	2275000	1836	2	1	1	no	no	yes	no
528	2275000	3970	1	1	1	no	no	no	no
529	2275000	3970	3	1	2	yes	no	yes	no
530	2240000	1950	3	1	1	no	no	no	yes
531	2233000	5300	3	1	1	no	no	no	no
532	2135000	3000	2	1	1	no	no	no	no
533	2100000	2400	3	1	2	yes	no	no	no
534	2100000	3000	4	1	2	yes	no	no	no
535	2100000	3360	2	1	1	yes	no	no	no
536	1960000	3420	5	1	2	no	no	no	no
537	1890000	1700	3	1	2	yes	no	no	no
538	1890000	3649	2	1	1	yes	no	no	no
539	1855000	2990	2	1	1	no	no	no	no
540	1820000	3000	2	1	1	yes	no	yes	no
541	1767150	2400	3	1	1	no	no	no	no
542	1750000	3620	2	1	1	yes	no	no	no
543	1750000	2910	3	1	1	no	no	no	no
544	1750000	3850	3	1	2	yes	no	no	no

545 rows × 13 columns

In [9]:
```
data.shape
```

Out[9]:
```
(545, 13)
```

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

Out[10]:

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

In [11]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
price               545 non-null int64
area                545 non-null int64
bedrooms            545 non-null int64
bathrooms           545 non-null int64
stories             545 non-null int64
mainroad            545 non-null object
guestroom           545 non-null object
basement            545 non-null object
hotwaterheating     545 non-null object
airconditioning     545 non-null object
parking             545 non-null int64
prefarea            545 non-null object
furnishingstatus    545 non-null object
dtypes: int64(6), object(7)
memory usage: 55.4+ KB
```

In [12]:
```
data.describe()
```

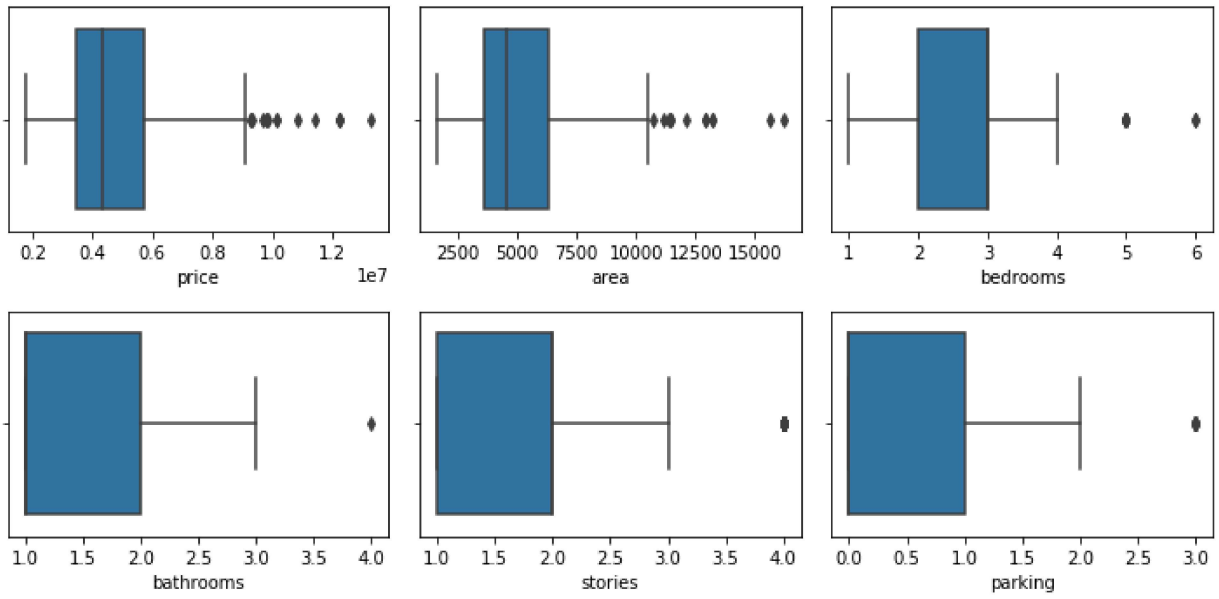
Out[12]:

	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

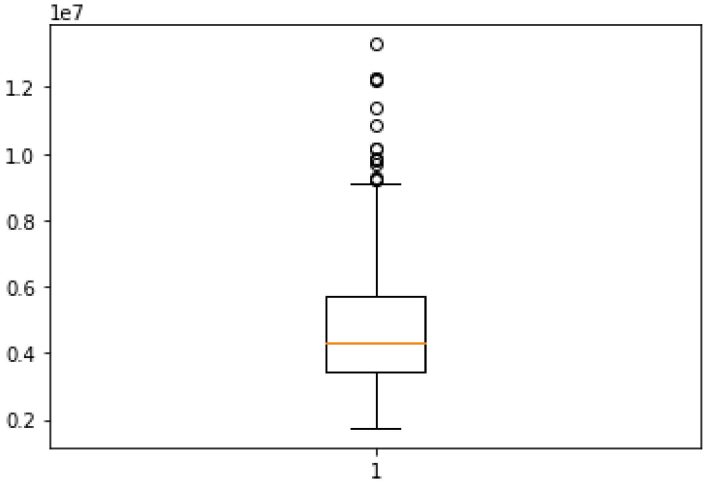
In [13]:
```
data.isnull().sum()*100/data.shape[0]
```

Out[13]:
```
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
prefarea          0.0
furnishingstatus  0.0
dtype: float64
```

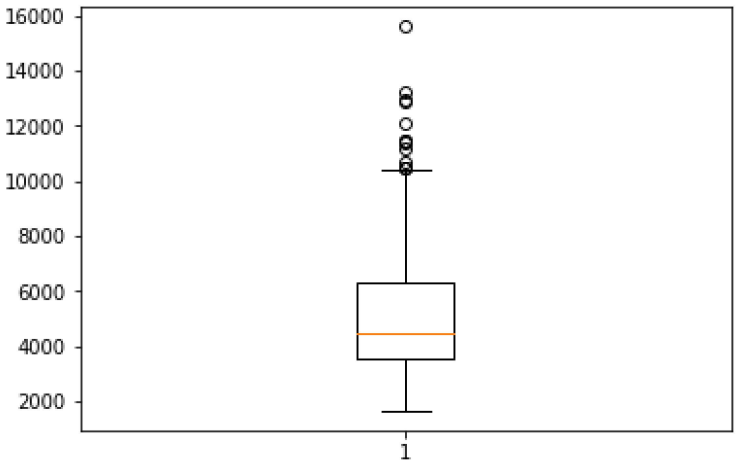
In [14]:
```
fig, axs = plt.subplots(2,3, figsize = (10,5))
plt1 = sns.boxplot(data['price'], ax = axs[0,0])
plt2 = sns.boxplot(data['area'], ax = axs[0,1])
plt3 = sns.boxplot(data['bedrooms'], ax = axs[0,2])
plt1 = sns.boxplot(data['bathrooms'], ax = axs[1,0])
plt2 = sns.boxplot(data['stories'], ax = axs[1,1])
plt3 = sns.boxplot(data['parking'], ax = axs[1,2])

plt.tight_layout()
```



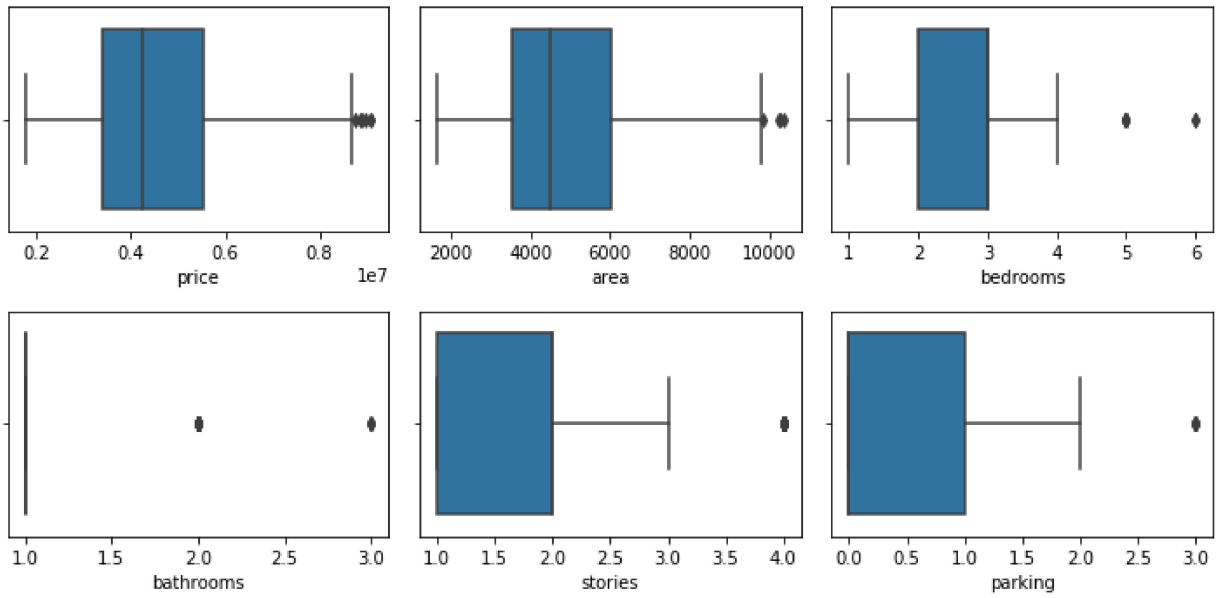
In [15]:
```
plt.boxplot(data.price)
Q1 = data.price.quantile(0.25)
Q3 = data.price.quantile(0.75)
IQR = Q3 - Q1
data = data[(data.price >= Q1 - 1.5*IQR) & (data.price <= Q3 + 1.5*IQR)]
```



In [16]:
```
plt.boxplot(data.area)
Q1 = data.area.quantile(0.25)
Q3 = data.area.quantile(0.75)
IQR = Q3 - Q1
data = data[(data.area >= Q1 - 1.5*IQR) & (data.area <= Q3 + 1.5*IQR)]
```

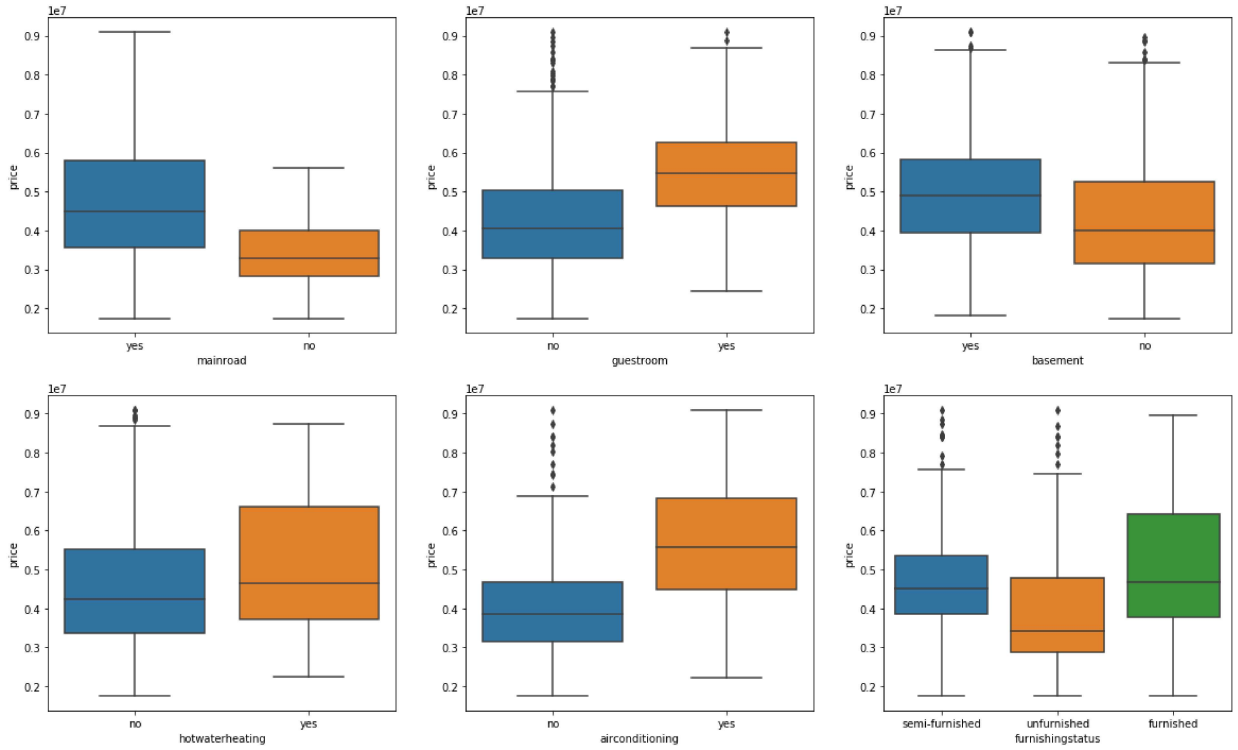


In [17]:
```
fig, axs = plt.subplots(2,3, figsize = (10,5))
plt1 = sns.boxplot(data['price'], ax = axs[0,0])
plt2 = sns.boxplot(data['area'], ax = axs[0,1])
plt3 = sns.boxplot(data['bedrooms'], ax = axs[0,2])
plt1 = sns.boxplot(data['bathrooms'], ax = axs[1,0])
plt2 = sns.boxplot(data['stories'], ax = axs[1,1])
plt3 = sns.boxplot(data['parking'], ax = axs[1,2])

plt.tight_layout()
```

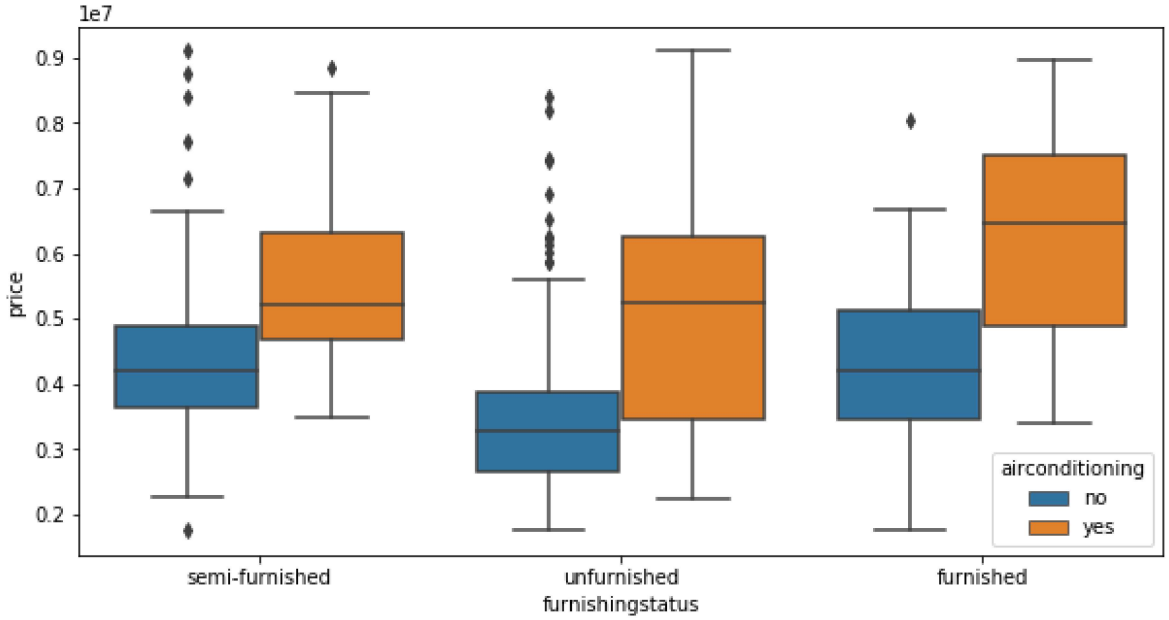


In [19]:
```
sns.pairplot(data)
plt.show()
```

In [20]:
```
plt.figure(figsize=(20, 12))
plt.subplot(2,3,1)
sns.boxplot(x = 'mainroad', y = 'price', data = data)
plt.subplot(2,3,2)
sns.boxplot(x = 'guestroom', y = 'price', data = data)
plt.subplot(2,3,3)
sns.boxplot(x = 'basement', y = 'price', data = data)
plt.subplot(2,3,4)
sns.boxplot(x = 'hotwaterheating', y = 'price', data = data)
plt.subplot(2,3,5)
sns.boxplot(x = 'airconditioning', y = 'price', data = data)
plt.subplot(2,3,6)
sns.boxplot(x = 'furnishingstatus', y = 'price', data = data)
plt.show()
```



In [21]:
```
plt.figure(figsize = (10, 5))
sns.boxplot(x = 'furnishingstatus', y = 'price', hue = 'airconditioning', data = dat
plt.show()
```



In [ ]: