In [7]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

In [8]:
```
# Generating sample data
df = pd.read_csv('housing.csv')
```

In [9]:
```
print(df.head())
```

```
      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

  hotwaterheating airconditioning  parking prefarea furnishingstat
us
0              no             yes        2      yes        furnish
ed
1              no             yes        3       no        furnish
ed
2              no              no        2      yes   semi-furnish
ed
3              no             yes        3      yes        furnish
ed
4              no             yes        2       no        furnish
ed
```

In [10]:
```
print(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
None
```

In [11]:
```
print(df.columns)
```

```
Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainr
oad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditionin
g',
       'parking', 'prefarea', 'furnishingstatus'],
      dtype='object')
```

In [12]:
```
print(df.describe())
```

```
              price          area    bedrooms   bathrooms     stor
ies  \
count  5.450000e+02    545.000000  545.000000  545.000000  545.000
000
mean   4.766729e+06   5150.541284    2.965138    1.286239    1.805
505
std    1.870440e+06   2170.141023    0.738064    0.502470    0.867
492
min    1.750000e+06   1650.000000    1.000000    1.000000    1.000
000
25%    3.430000e+06   3600.000000    2.000000    1.000000    1.000
000
50%    4.340000e+06   4600.000000    3.000000    1.000000    2.000
000
75%    5.740000e+06   6360.000000    3.000000    2.000000    2.000
000
max    1.330000e+07  16200.000000    6.000000    4.000000    4.000
000

          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
```

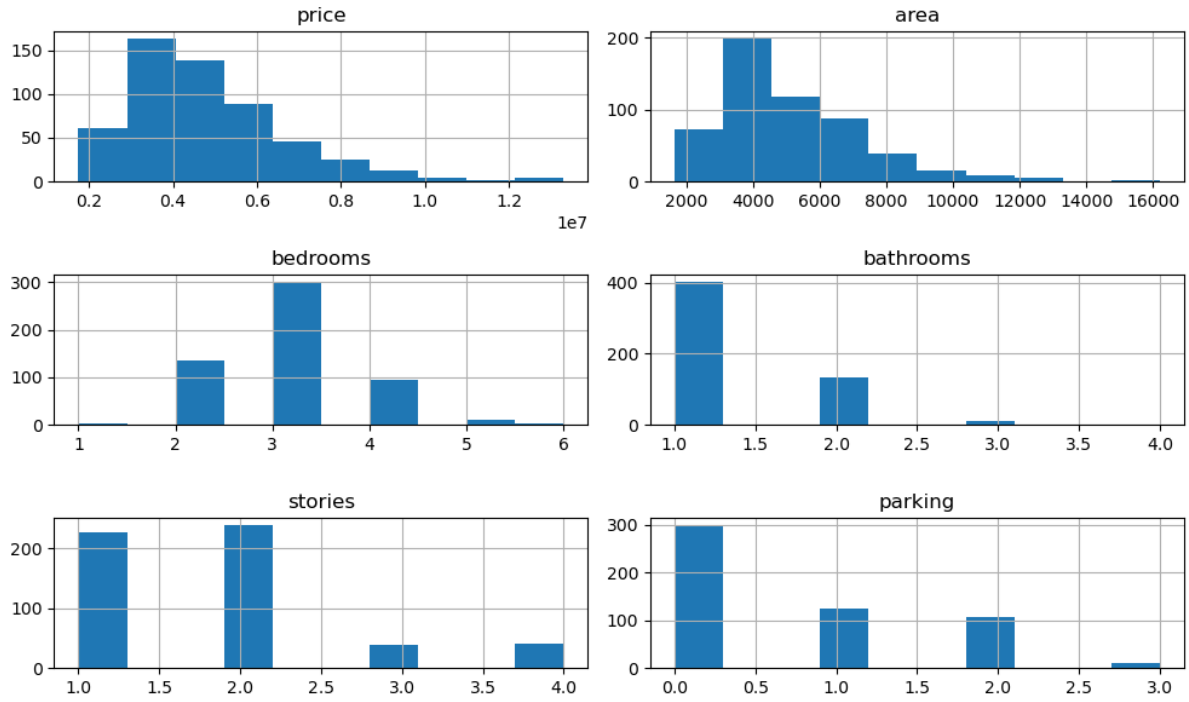
In [13]:
```
print(df.isna().sum())
```

```
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
```

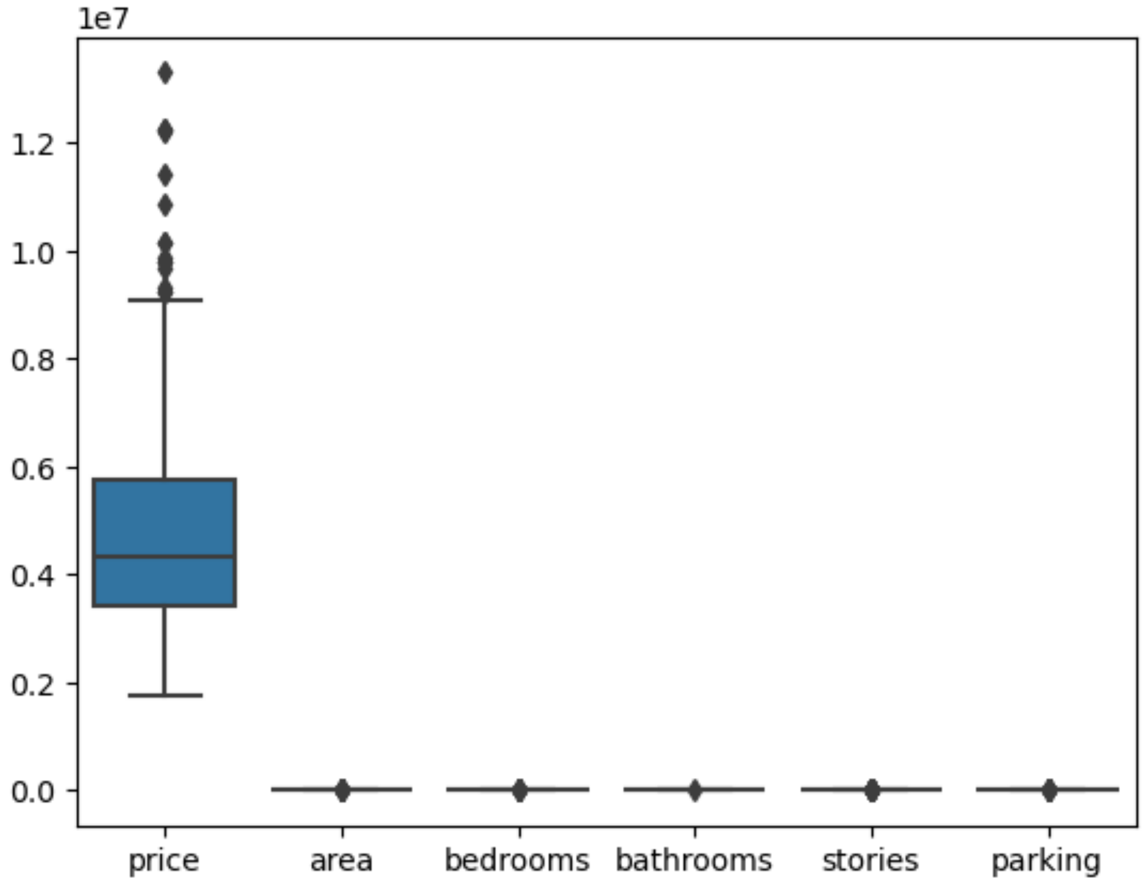
In [14]:
```
# Summary statistics
summary = df.describe()
print(summary)
```

```
              price          area    bedrooms   bathrooms     stor
ies  \
count  5.450000e+02    545.000000  545.000000  545.000000  545.000
000
mean   4.766729e+06   5150.541284    2.965138    1.286239    1.805
505
std    1.870440e+06   2170.141023    0.738064    0.502470    0.867
492
min    1.750000e+06   1650.000000    1.000000    1.000000    1.000
000
25%    3.430000e+06   3600.000000    2.000000    1.000000    1.000
000
50%    4.340000e+06   4600.000000    3.000000    1.000000    2.000
000
75%    5.740000e+06   6360.000000    3.000000    2.000000    2.000
000
max    1.330000e+07  16200.000000    6.000000    4.000000    4.000
000

          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
```

In [15]:
```
# Histograms
df.hist(figsize=(10, 6))
plt.tight_layout()
plt.show()
```



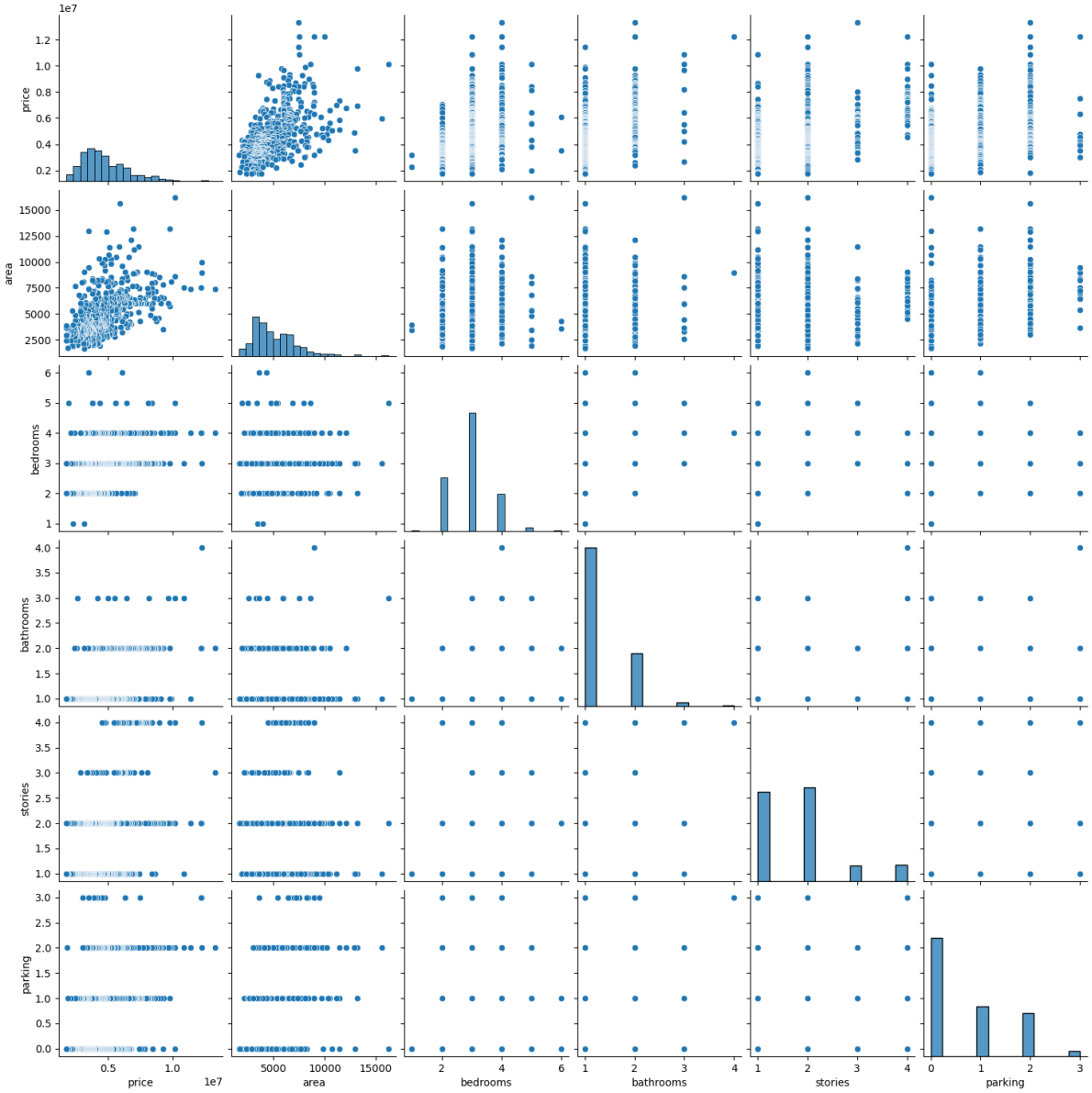
In [16]:
```
# Boxplot
sns.boxplot(data=df)
plt.show()
```



In [17]:
```
# Scatter plot
sns.pairplot(df)
plt.show()
```

```
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:11
19: FutureWarning: use_inf_as_na option is deprecated and will be
removed in a future version. Convert inf values to NaN before oper
ating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:11
19: FutureWarning: use_inf_as_na option is deprecated and will be
removed in a future version. Convert inf values to NaN before oper
ating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:11
19: FutureWarning: use_inf_as_na option is deprecated and will be
removed in a future version. Convert inf values to NaN before oper
ating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:11
19: FutureWarning: use_inf_as_na option is deprecated and will be
removed in a future version. Convert inf values to NaN before oper
ating instead.
```

```
  with pd.option_context('mode.use_inf_as_na', True):
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



In [18]:
```
df = df.apply(pd.to_numeric, errors='coerce')
corr_matrix = df.corr()
print(corr_matrix)
```

```
                price      area  bedrooms  bathrooms   storie
s  mainroad  \
price           1.000000  0.535997  0.366494   0.517545  0.42071
2       NaN
```

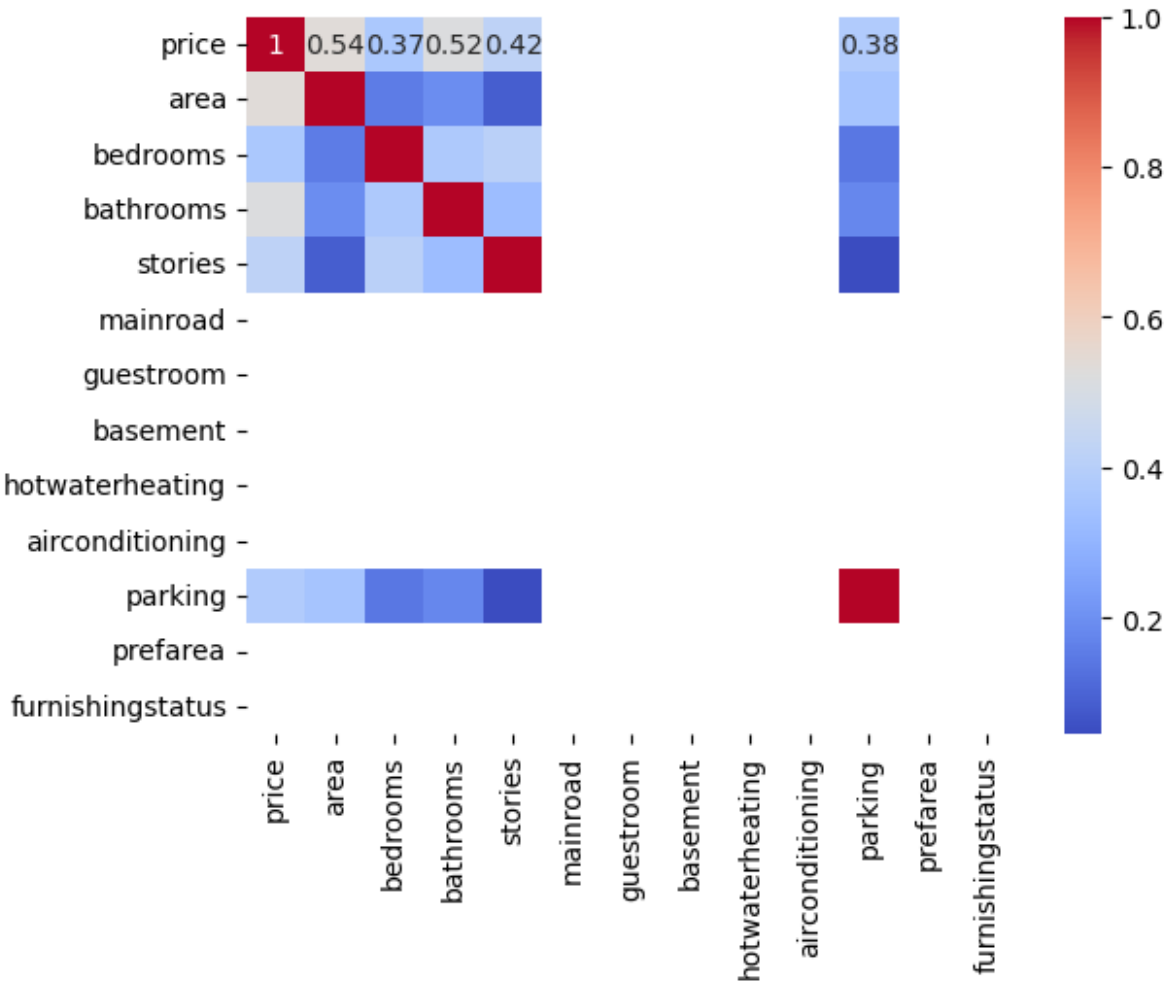
```
area              0.535997  1.000000  0.151858   0.193820  0.08399
6       NaN
bedrooms          0.366494  0.151858  1.000000   0.373930  0.40856
4       NaN
bathrooms         0.517545  0.193820  0.373930   1.000000  0.32616
5       NaN
stories           0.420712  0.083996  0.408564   0.326165  1.00000
0       NaN
mainroad               NaN       NaN       NaN        NaN       Na
N       NaN
guestroom              NaN       NaN       NaN        NaN       Na
N       NaN
basement               NaN       NaN       NaN        NaN       Na
N       NaN
hotwaterheating        NaN       NaN       NaN        NaN       Na
N       NaN
airconditioning        NaN       NaN       NaN        NaN       Na
N       NaN
parking           0.384394  0.352980  0.139270   0.177496  0.04554
7       NaN
prefarea               NaN       NaN       NaN        NaN       Na
N       NaN
furnishingstatus       NaN       NaN       NaN        NaN       Na
N       NaN

                  guestroom  basement  hotwaterheating  airconditi
oning  \
price                   NaN       NaN              NaN
NaN
area                    NaN       NaN              NaN
NaN
bedrooms                NaN       NaN              NaN
NaN
bathrooms               NaN       NaN              NaN
NaN
stories                 NaN       NaN              NaN
NaN
mainroad                NaN       NaN              NaN
NaN
guestroom               NaN       NaN              NaN
NaN
basement                NaN       NaN              NaN
NaN
hotwaterheating         NaN       NaN              NaN
NaN
airconditioning         NaN       NaN              NaN
NaN
parking                 NaN       NaN              NaN
NaN
prefarea                NaN       NaN              NaN
NaN
furnishingstatus        NaN       NaN              NaN
NaN
```

	parking	prefarea	furnishingstatus
price	0.384394	NaN	NaN
area	0.352980	NaN	NaN
bedrooms	0.139270	NaN	NaN
bathrooms	0.177496	NaN	NaN
stories	0.045547	NaN	NaN
mainroad	NaN	NaN	NaN
guestroom	NaN	NaN	NaN
basement	NaN	NaN	NaN
hotwaterheating	NaN	NaN	NaN
airconditioning	NaN	NaN	NaN
parking	1.000000	NaN	NaN
prefarea	NaN	NaN	NaN
furnishingstatus	NaN	NaN	NaN

In [19]:
```
sns.heatmap(df.corr(),annot=True,cmap='coolwarm')
```

```
/opt/anaconda3/lib/python3.11/site-packages/seaborn/matrix.py:260:
FutureWarning: Format strings passed to MaskedConstant are ignore
d, but in future may error or produce different behavior
  annotation = ("{:" + self.fmt + "}").format(val)
```

Out[19]: <Axes: >



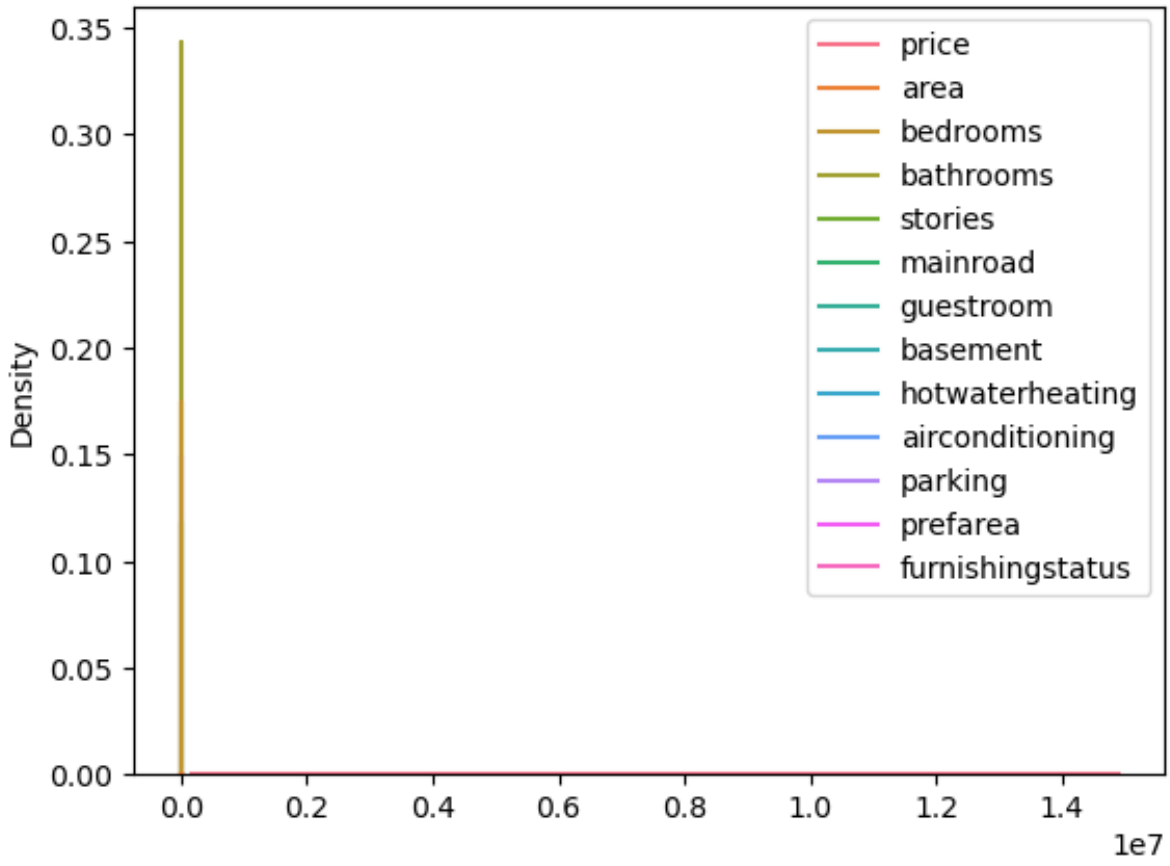
In [20]:
```
!pip install matplotlib
```

```
Requirement already satisfied: matplotlib in /opt/anaconda3/lib/py
thon3.11/site-packages (3.8.0)
Requirement already satisfied: contourpy>=1.0.1 in /opt/anaconda3/
lib/python3.11/site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in /opt/anaconda3/lib/
python3.11/site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /opt/anaconda
3/lib/python3.11/site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/anaconda
3/lib/python3.11/site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy<2,>=1.21 in /opt/anaconda3/li
b/python3.11/site-packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /opt/anaconda3/l
ib/python3.11/site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in /opt/anaconda3/li
b/python3.11/site-packages (from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in /opt/anaconda3/
lib/python3.11/site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in /opt/anacon
da3/lib/python3.11/site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in /opt/anaconda3/lib/pyth
on3.11/site-packages (from python-dateutil>=2.7->matplotlib) (1.1
6.0)
```

In [21]:
```
import numpy as np
import matplotlib.pyplot as plt
```

In [33]:
```
sns.kdeplot(df)
plt.show()
```

```
/opt/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



In [ ]: