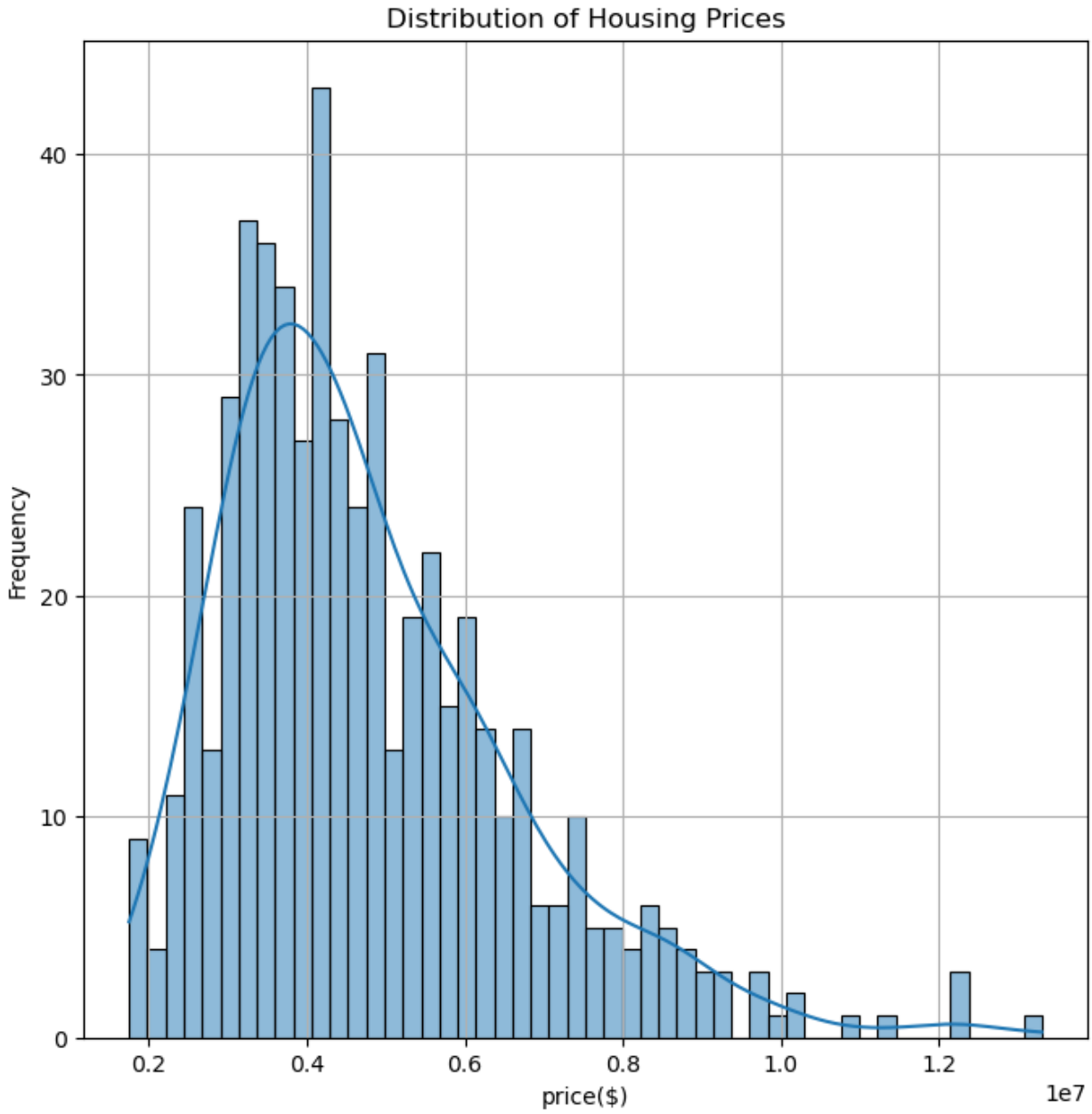
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

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

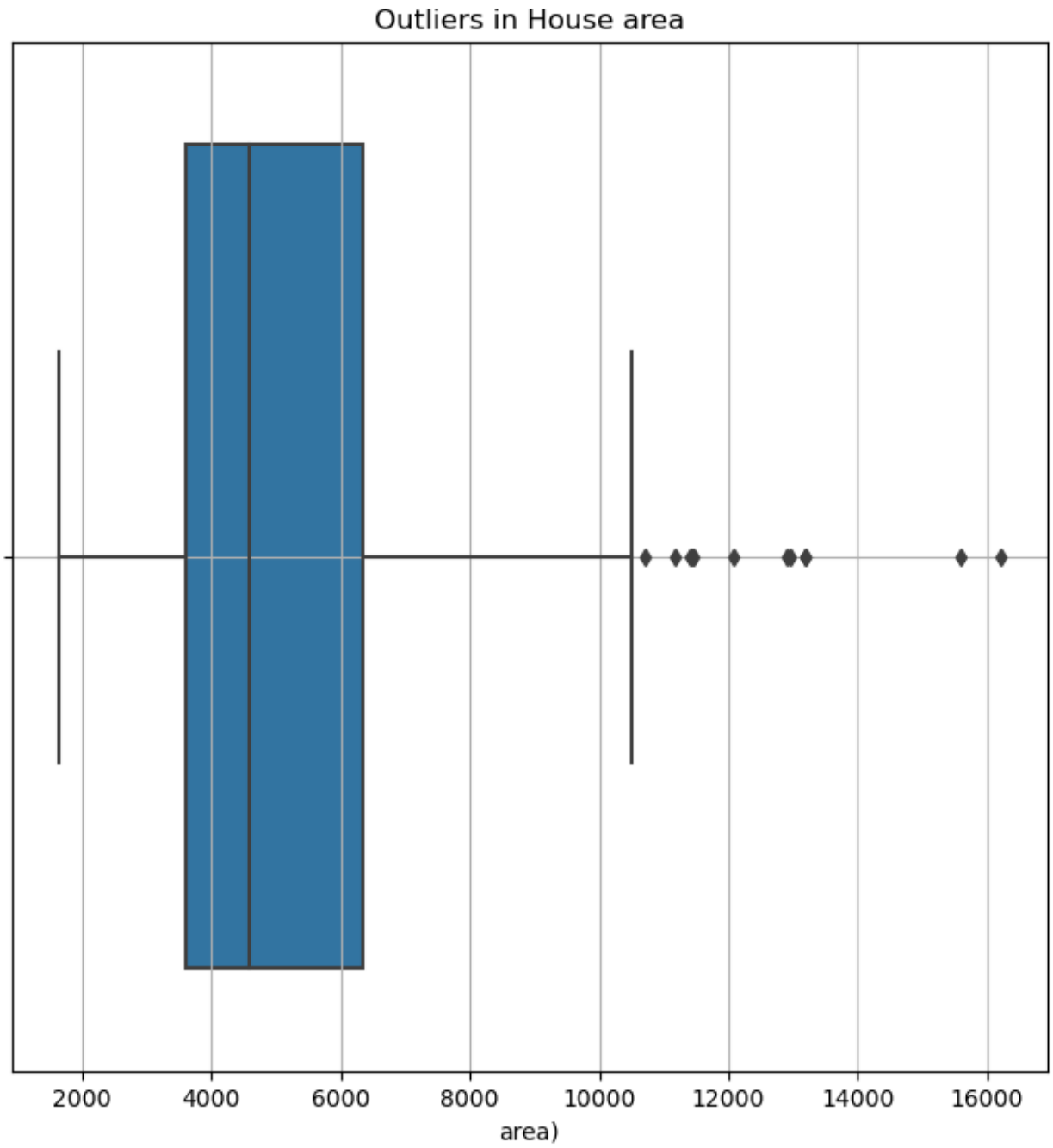
```
        price  area  bedrooms  bathrooms  stories mainroad guestroom
basement  \
0    13300000  7420         4          2        3      yes        no
no
1    12250000  8960         4          4        4      yes        no
no
2    12250000  9960         3          2        2      yes        no
yes
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]
```

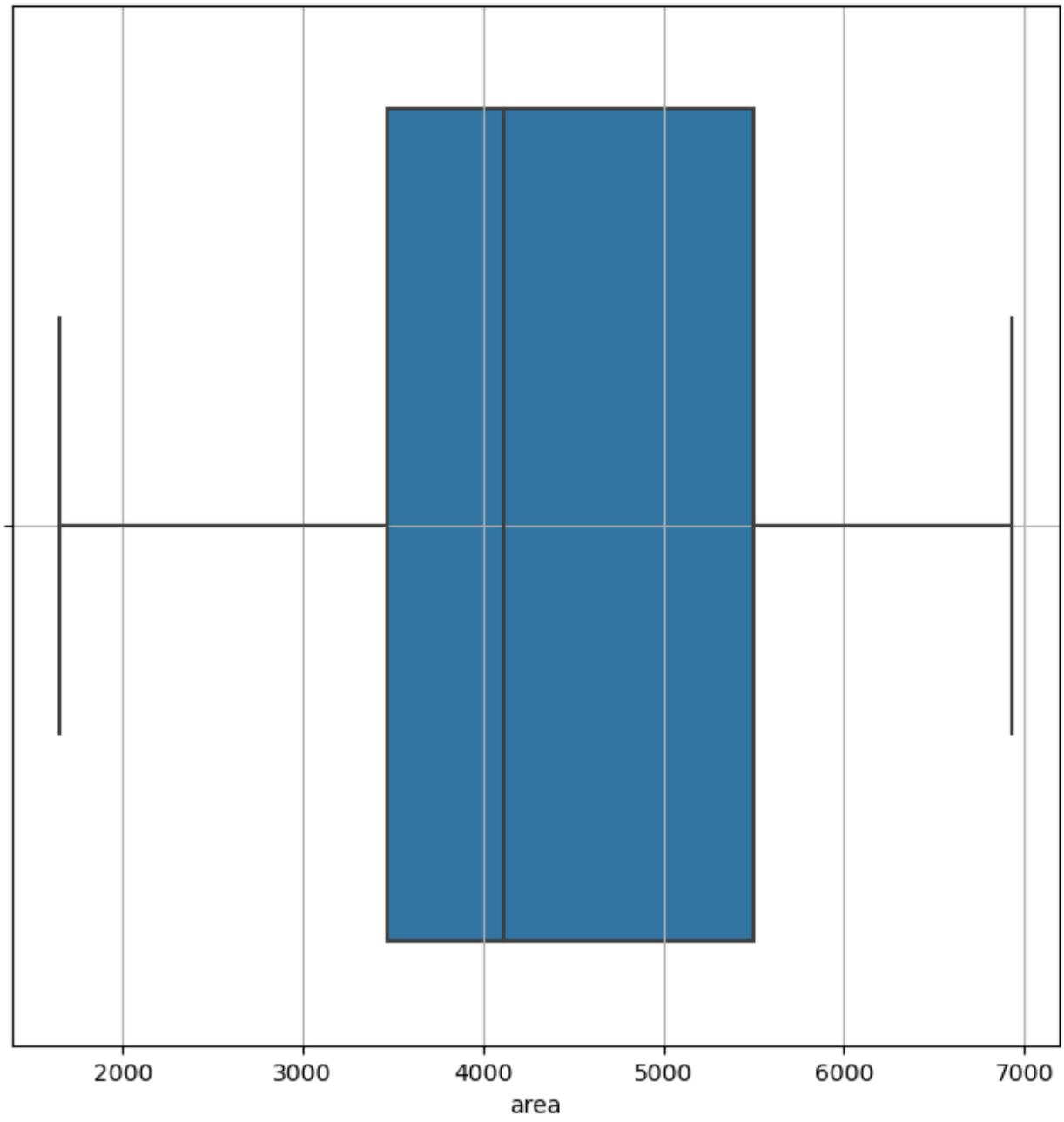
```
plt.figure(figsize=(8,8))
sns.histplot(data['price'],bins=50,kde=True)
plt.title("Distribution of Housing Prices")
plt.xlabel("price($)")
plt.ylabel("Frequency")
plt.grid(True)
plt.show()
```




```
plt.figure(figsize=(8,8))
sns.boxplot(data=data,x='area')
plt.title("Outliers in House area")
plt.xlabel("area)")
plt.grid(True)
plt.show()
```




```
data=data[data['area']<7000]
plt.figure(figsize=(8,8))
sns.boxplot(data=data,x='area')
plt.xlabel("area")
plt.grid(True)
plt.show()
```




```
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.pop('mainroad')
```

```
9      yes
11     yes
12     yes
13     yes
15     yes
      ...
540    yes
541     no
542    yes
543     no
544    yes
Name: mainroad, Length: 454, dtype: object
```

```
data.pop('guestroom')
```

```
9      yes
11     yes
12      no
13      no
15      no
      ...
540     no
541     no
542     no
543     no
544     no
Name: guestroom, Length: 454, dtype: object
```

```
data.pop('basement')
```

```
9       no
11     yes
12      no
13      no
15     yes
      ...
540    yes
541     no
542     no
543     no
544     no
Name: basement, Length: 454, dtype: object
```

```
data.pop('hotwaterheating')
```

```
9       no
11     yes
12      no
13     yes
15      no
      ...
540     no
541     no
542     no
543     no
544     no
Name: hotwaterheating, Length: 454, dtype: object
```

```
data.pop('airconditioning')
```

```
9      yes
11      no
12     yes
13      no
15      no
      ...
540     no
541     no
542     no
543     no
544     no
Name: airconditioning, Length: 454, dtype: object
```

```
data.pop('furnishingstatus')
```

```
9         unfurnished
11     semi-furnished
12     semi-furnished
13          furnished
15     semi-furnished
           ...
540       unfurnished
541    semi-furnished
542       unfurnished
543         furnished
544       unfurnished
Name: furnishingstatus, Length: 454, dtype: object
```

```
data.pop('prefarea')
```

```
9      yes
11      no
12     yes
13      no
```

```
15      no
      ...
540     no
541     no
542     no
543     no
544     no
Name: prefarea, Length: 454, dtype: object
```

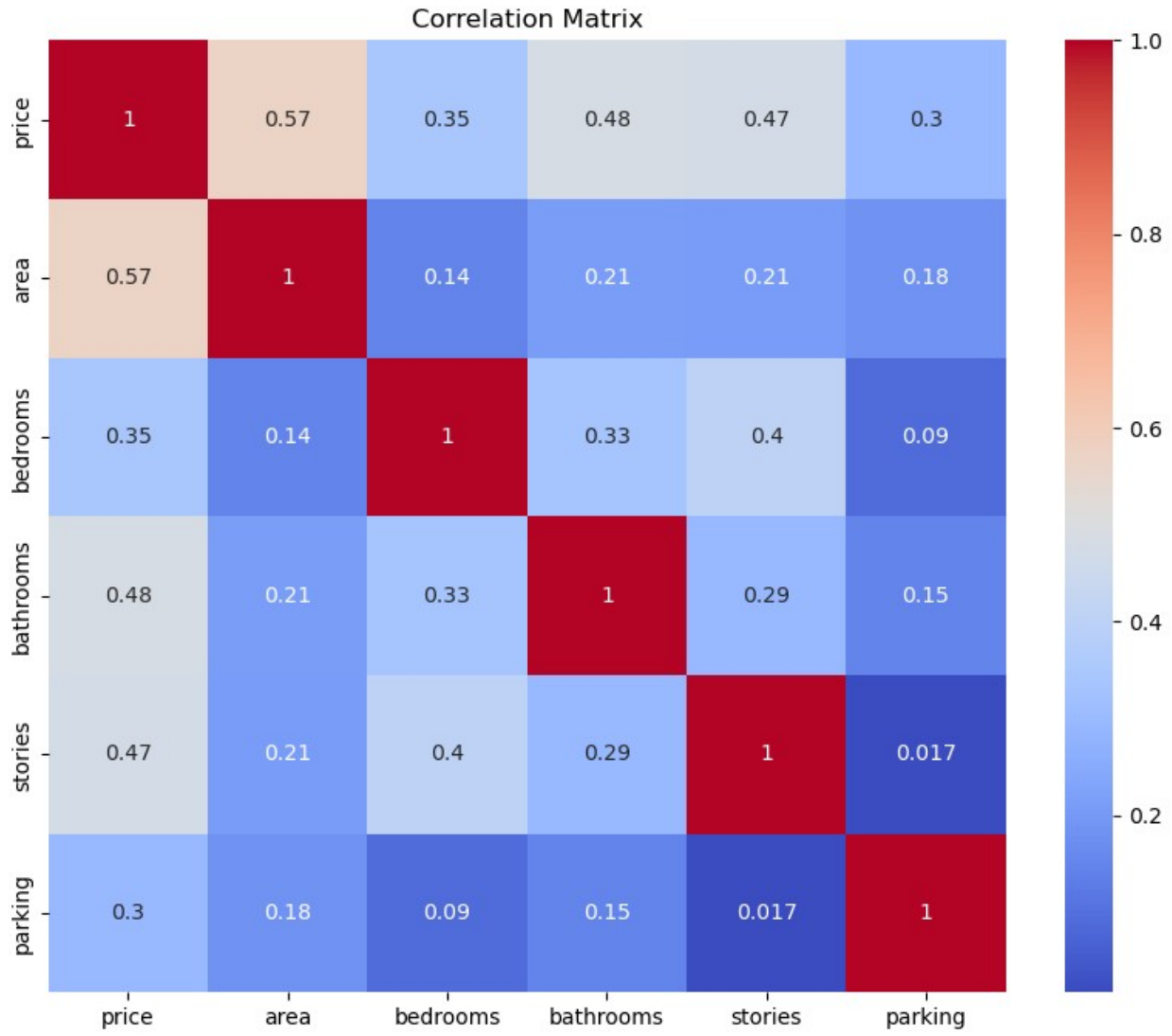
```
data
```

	price	area	bedrooms	bathrooms	stories	parking
9	9800000	5750	3	2	4	1
11	9681000	6000	4	3	2	2
12	9310000	6550	4	2	2	1
13	9240000	3500	4	2	2	2
15	9100000	6000	4	1	2	2
..	...	...	...	...	...	...
540	1820000	3000	2	1	1	2
541	1767150	2400	3	1	1	0
542	1750000	3620	2	1	1	0
543	1750000	2910	3	1	1	0
544	1750000	3850	3	1	2	0

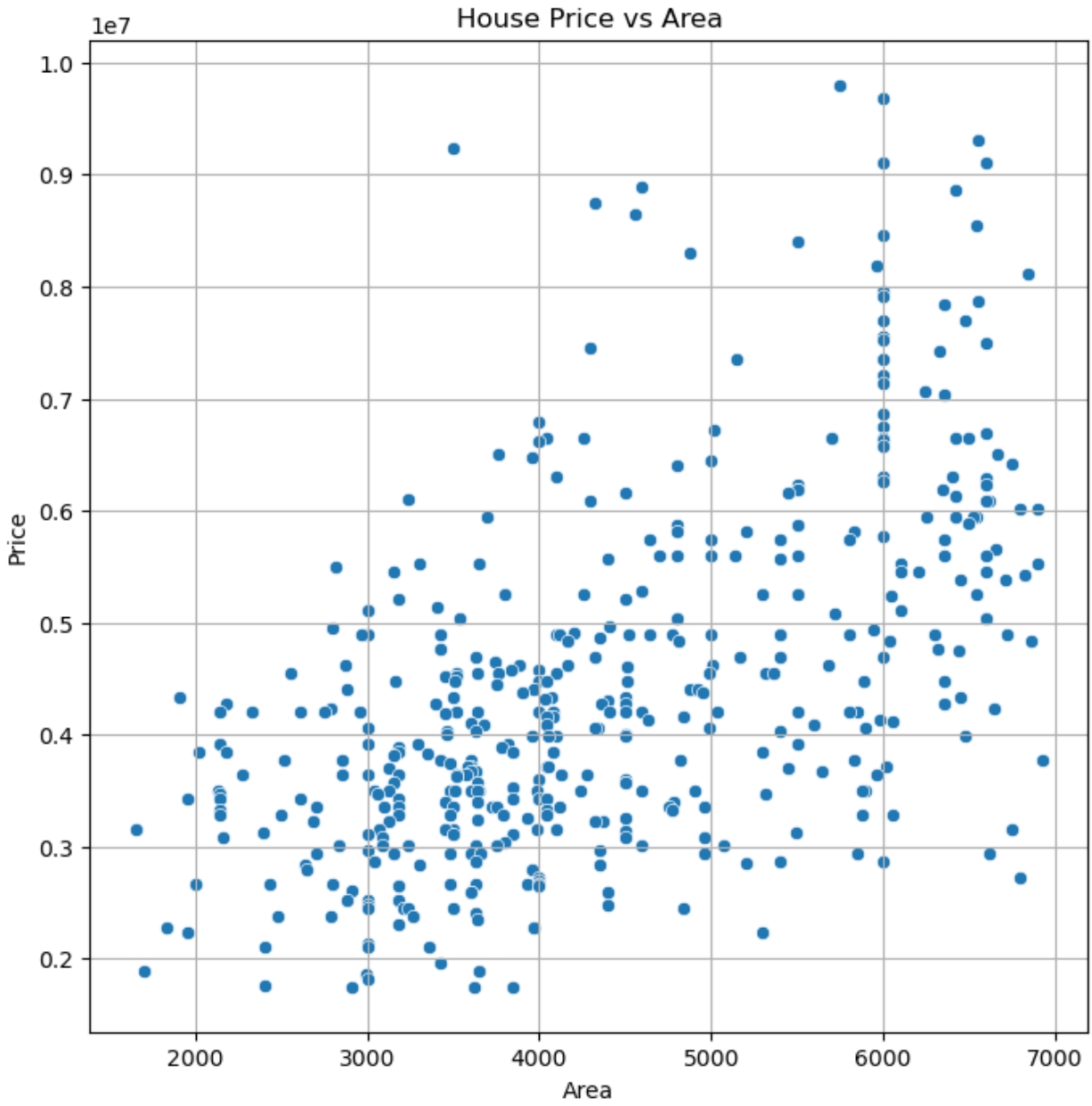
```
[454 rows x 6 columns]
```

```
correlation_matrix=data.corr()
```

```
plt.figure(figsize=(10,8))
sns.heatmap(correlation_matrix,annot=True,cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()
```




```
plt.figure(figsize=(8,8))
sns.scatterplot(data=data,x='area',y='price')
plt.title("House Price vs Area")
plt.xlabel("Area")
plt.ylabel("Price")
plt.grid(True)
plt.show()
```


House Price vs Area
1e7
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Price
2000
3000
4000
5000
6000
7000
Area