In [1]:

```
import pandas as pd
```

In [3]:

```
data=pd.read_csv(r"C:\Users\Bhavya's Surface\Documents\pandas\LondonHousingData.csv")
```

In [4]:

```
data
```

Out[4]:

	date	area	average_price	code	houses_sold	no_of_crimes
0	1/1/1995	city of london	91449	E09000001	17.0	NaN
1	2/1/1995	city of london	82203	E09000001	7.0	NaN
2	3/1/1995	city of london	79121	E09000001	14.0	NaN
3	4/1/1995	city of london	77101	E09000001	7.0	NaN
4	5/1/1995	city of london	84409	E09000001	10.0	NaN
...	...	...	...	...	...	...
13544	9/1/2019	england	249942	E92000001	64605.0	NaN
13545	10/1/2019	england	249376	E92000001	68677.0	NaN
13546	11/1/2019	england	248515	E92000001	67814.0	NaN
13547	12/1/2019	england	250410	E92000001	NaN	NaN
13548	1/1/2020	england	247355	E92000001	NaN	NaN

13549 rows × 6 columns

In [5]:

```
data.count()
```

Out[5]:

```
date             13549
area             13549
average_price    13549
code             13549
houses_sold      13455
no_of_crimes      7439
dtype: int64
```

In [10]:

```
data.isnull().sum()
```

Out[10]:

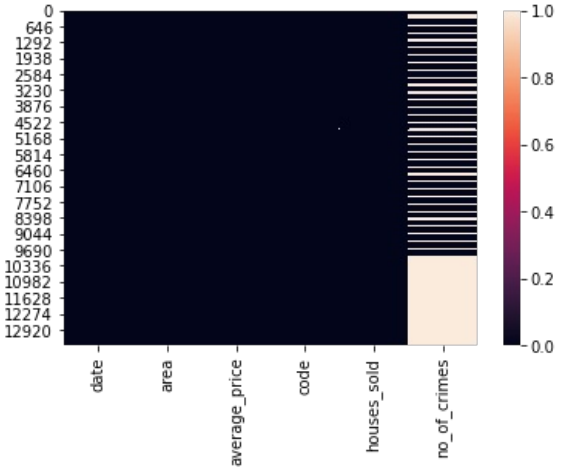
```
date                0
area                0
average_price       0
code                0
houses_sold        94
no_of_crimes     6110
dtype: int64
```

In [12]:

```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [14]:

```
#heatmap of no. of crimes
sns.heatmap(data.isnull())
plt.show()
```



In [15]:

```
data.head()
```

Out[15]:

	date	area	average_price	code	houses_sold	no_of_crimes
0	1/1/1995	city of london	91449	E09000001	17.0	NaN
1	2/1/1995	city of london	82203	E09000001	7.0	NaN
2	3/1/1995	city of london	79121	E09000001	14.0	NaN
3	4/1/1995	city of london	77101	E09000001	7.0	NaN
4	5/1/1995	city of london	84409	E09000001	10.0	NaN

In [16]:

```
data.dtypes
```

Out[16]:

```
date              object
area              object
average_price      int64
code              object
houses_sold      float64
no_of_crimes     float64
dtype: object
```

In [19]:

```
#for conversion of date into date time format
data.date=pd.to_datetime(data.date)
```

In [20]:

```
data.head()
```

Out[20]:

	date	area	average_price	code	houses_sold	no_of_crimes
0	1995-01-01	city of london	91449	E09000001	17.0	NaN
1	1995-02-01	city of london	82203	E09000001	7.0	NaN
2	1995-03-01	city of london	79121	E09000001	14.0	NaN
3	1995-04-01	city of london	77101	E09000001	7.0	NaN
4	1995-05-01	city of london	84409	E09000001	10.0	NaN

In [22]:

```
#adding of date column
data['Year']= data.date.dt.year
```

In [23]:

```
data.head()
```

Out[23]:

	date	area	average_price	code	houses_sold	no_of_crimes	Year
0	1995-01-01	city of london	91449	E09000001	17.0	NaN	1995
1	1995-02-01	city of london	82203	E09000001	7.0	NaN	1995
2	1995-03-01	city of london	79121	E09000001	14.0	NaN	1995
3	1995-04-01	city of london	77101	E09000001	7.0	NaN	1995
4	1995-05-01	city of london	84409	E09000001	10.0	NaN	1995

In [24]:

```
#adding of the month as 2nd column
data.insert(1,'month', data.date.dt.month)
```

In [25]:

```
data.head()
```

Out[25]:

	date	month	area	average_price	code	houses_sold	no_of_crimes	Year
0	1995-01-01	1	city of london	91449	E09000001	17.0	NaN	1995
1	1995-02-01	2	city of london	82203	E09000001	7.0	NaN	1995
2	1995-03-01	3	city of london	79121	E09000001	14.0	NaN	1995
3	1995-04-01	4	city of london	77101	E09000001	7.0	NaN	1995
4	1995-05-01	5	city of london	84409	E09000001	10.0	NaN	1995

In [27]:

```
#remove the columns of year and month
data.drop(['month','Year'],axis=1, inplace=True)
```

In [28]:

```
data.head()
```

Out[28]:

	date	area	average_price	code	houses_sold	no_of_crimes
0	1995-01-01	city of london	91449	E09000001	17.0	NaN
1	1995-02-01	city of london	82203	E09000001	7.0	NaN
2	1995-03-01	city of london	79121	E09000001	14.0	NaN
3	1995-04-01	city of london	77101	E09000001	7.0	NaN
4	1995-05-01	city of london	84409	E09000001	10.0	NaN

In [29]:

```
#show no. of records where crime is 0 and total no. of those records
data.head()
```

Out[29]:

	date	area	average_price	code	houses_sold	no_of_crimes
0	1995-01-01	city of london	91449	E09000001	17.0	NaN
1	1995-02-01	city of london	82203	E09000001	7.0	NaN
2	1995-03-01	city of london	79121	E09000001	14.0	NaN
3	1995-04-01	city of london	77101	E09000001	7.0	NaN
4	1995-05-01	city of london	84409	E09000001	10.0	NaN

In [33]:

```
data[data.no_of_crimes==0]
```

Out[33]:

	date	area	average_price	code	houses_sold	no_of_crimes
72	2001-01-01	city of london	284262	E09000001	24.0	0.0
73	2001-02-01	city of london	198137	E09000001	37.0	0.0
74	2001-03-01	city of london	189033	E09000001	44.0	0.0
75	2001-04-01	city of london	205494	E09000001	38.0	0.0
76	2001-05-01	city of london	223459	E09000001	30.0	0.0
...	...	...	...	...	...	...
178	2009-11-01	city of london	397909	E09000001	11.0	0.0
179	2009-12-01	city of london	411955	E09000001	16.0	0.0
180	2010-01-01	city of london	464436	E09000001	20.0	0.0
181	2010-02-01	city of london	490525	E09000001	9.0	0.0
182	2010-03-01	city of london	498241	E09000001	15.0	0.0

104 rows × 6 columns

In [35]:

```
len(data[data.no_of_crimes==0])
```

Out[35]:

```
104
```

In [38]:

```
#max and min houses_sold in england
df1=data[data.area=='england']
```

In [41]:

```
df1.groupby('houses_sold').average_price.max()
```

Out[41]:

```
houses_sold
25782.0     162673
26660.0     160956
34383.0     174458
34857.0     159340
35029.0     168230
              ...
124849.0     93331
126360.0    112832
126475.0    110001
126778.0    156730
132163.0    104458
Name: average_price, Length: 297, dtype: int64
```

In [43]:

```
#max & min no_of_crimes
data.groupby('area').no_of_crimes.max()
```

Out[43]:

```
area
barking and dagenham      2049.0
barnet                    2893.0
bexley                    1914.0
brent                     2937.0
bromley                   2637.0
camden                    4558.0
city of london              10.0
croydon                   3263.0
ealing                    3401.0
east midlands                NaN
east of england              NaN
enfield                   2798.0
england                      NaN
greenwich                 2853.0
hackney                   3466.0
hammersmith and fulham    2645.0
haringey                  3199.0
harrow                    1763.0
havering                  1956.0
hillingdon                2819.0
hounslow                  2817.0
inner london                 NaN
islington                 3384.0
kensington and chelsea    2778.0
kingston upon thames      1379.0
lambeth                   4701.0
lewisham                  2813.0
london                       NaN
merton                    1623.0
newham                    3668.0
north east                   NaN
north west                   NaN
outer london                 NaN
redbridge                 2560.0
richmond upon thames      1551.0
south east                   NaN
south west                   NaN
southwark                 3821.0
sutton                    1425.0
tower hamlets             3316.0
waltham forest            2941.0
wandsworth                3051.0
west midlands                NaN
westminster               7461.0
yorks and the humber         NaN
Name: no_of_crimes, dtype: float64
```

In [46]:

```
data.groupby('area').no_of_crimes.min().sort_values(ascending=True)
```

Out[46]:

```
area
city of london               0.0
kingston upon thames       692.0
richmond upon thames       700.0
sutton                     787.0
merton                     819.0
bexley                     860.0
harrow                     937.0
havering                  1130.0
barking and dagenham      1217.0
hammersmith and fulham    1323.0
kensington and chelsea    1347.0
bromley                   1441.0
hillingdon                1445.0
redbridge                 1487.0
greenwich                 1513.0
hounslow                  1529.0
haringey                  1536.0
waltham forest            1575.0
wandsworth                1582.0
enfield                   1635.0
tower hamlets             1646.0
lewisham                  1675.0
barnet                    1703.0
brent                     1850.0
hackney                   1870.0
ealing                    1871.0
islington                 1871.0
croydon                   2031.0
camden                    2079.0
newham                    2130.0
southwark                 2267.0
lambeth                   2381.0
westminster               3504.0
east midlands                NaN
east of england              NaN
england                      NaN
inner london                 NaN
london                       NaN
north east                   NaN
north west                   NaN
outer london                 NaN
south east                   NaN
south west                   NaN
west midlands                NaN
yorks and the humber         NaN
Name: no_of_crimes, dtype: float64
```

In [48]:

```
#total count of records of each area whr avg price < 100000
data[data.average_price<100000].area.value_counts()
```

Out[48]:

```
north east              112
north west              111
yorks and the humber    110
east midlands            96
west midlands            94
england                  87
barking and dagenham     85
south west               78
east of england          76
newham                   72
waltham forest           64
bexley                   64
lewisham                 62
havering                 60
south east               59
greenwich                59
croydon                  57
enfield                  54
sutton                   54
hackney                  53
redbridge                52
southwark                48
tower hamlets            47
outer london             46
hillingdon               44
lambeth                  41
hounslow                 41
brent                    40
london                   39
merton                   35
bromley                  33
haringey                 33
inner london             31
ealing                   31
harrow                   30
kingston upon thames     30
wandsworth               26
barnet                   25
islington                19
city of london           11
Name: area, dtype: int64
```

In [ ]:

```

```