LONDON HOUSING DATASET

This dataset is primarly centred around the housing in london.It contains a lot of additional relevant data

1.Yearly number of houses prices.

2.Yearly numbers of houses sold.

3.Monthly numbers of crimes commited

HERE ARE RANDOM SETS OF QUESTIONS

In [1]:

```
#importing modules
import pandas as pd
```

In [5]:

```
data=pd.read_csv(r"5. London Housing Data.csv")
#r is to remove the unicode error
```

In [8]:

```
data.head(5)
```

Out[8]:

	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

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

exploringdata

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

Out[11]:
```
date              13549
area              13549
average_price     13549
code              13549
houses_sold       13455
no_of_crimes       7439
dtype: int64
```

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

Out[12]:

	date	area	average_price	code	houses_sold	no_of_crimes
0	False	False	False	False	False	True
1	False	False	False	False	False	True
2	False	False	False	False	False	True
3	False	False	False	False	False	True
4	False	False	False	False	False	True
...	...	...	...	...	...	...
13544	False	False	False	False	False	True
13545	False	False	False	False	False	True
13546	False	False	False	False	False	True
13547	False	False	False	False	True	True
13548	False	False	False	False	True	True

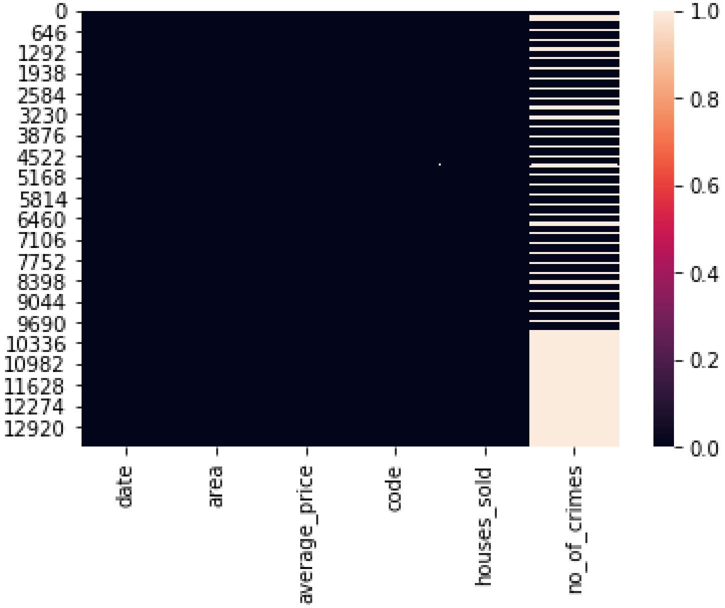
13549 rows × 6 columns

In [13]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

In [17]:
```
sns.heatmap(data.isnull())
```

Out[17]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a260578d0>
```



1.Convert the datatype of 'Date' column to Date-time format

In [18]:
```
data.head()
```

Out[18]:

	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

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	date	area	average_price	code	houses_sold	no_of_crimes
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 [21]:

```
data['date']=pd.to_datetime(data['date'])
```

In [22]:

```
data.dtypes
```

Out[22]:

```
date              datetime64[ns]
area                      object
average_price              int64
code                      object
houses_sold              float64
no_of_crimes             float64
dtype: object
```

2.Add a new column year

In [25]:

```
data['Year']=data.date.dt.year
```

In [26]:

```
data
```

Out[26]:

	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
...	...	...	...	...	...	...	...
13544	2019-09-01	england	249942	E92000001	64605.0	NaN	2019

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	date	area	average_price	code	houses_sold	no_of_crimes	Year
13545	2019-10-01	england	249376	E92000001	68677.0	NaN	2019
13546	2019-11-01	england	248515	E92000001	67814.0	NaN	2019
13547	2019-12-01	england	250410	E92000001	NaN	NaN	2019
13548	2020-01-01	england	247355	E92000001	NaN	NaN	2020

13549 rows × 7 columns

In [27]:

```
#2.1 Add month as 2nd column
```

In [29]:

```
data.insert(1,'month',data.date.dt.month)
```

In [30]:

```
data
```

Out[30]:

	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
...	...	...	...	...	...	...	...	...
13544	2019-09-01	9	england	249942	E92000001	64605.0	NaN	2019
13545	2019-10-01	10	england	249376	E92000001	68677.0	NaN	2019

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	date	month	area	average_price	code	houses_sold	no_of_crimes	Year
13546	2019-11-01	11	england	248515	E92000001	67814.0	NaN	2019
13547	2019-12-01	12	england	250410	E92000001	NaN	NaN	2019
13548	2020-01-01	1	england	247355	E92000001	NaN	NaN	2020

13549 rows × 8 columns

In [31]:
```
data.head()
```

Out[31]:

	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

3.column removal

In [35]:
```
data.drop(['month','Year'],axis=1,inplace=True)
```

In [36]:
```
data
```

Out[36]:

	date	area	average_price	code	houses_sold	no_of_crimes

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	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
...	...	...	...	...	...	...
13544	2019-09-01	england	249942	E92000001	64605.0	NaN
13545	2019-10-01	england	249376	E92000001	68677.0	NaN
13546	2019-11-01	england	248515	E92000001	67814.0	NaN
13547	2019-12-01	england	250410	E92000001	NaN	NaN
13548	2020-01-01	england	247355	E92000001	NaN	NaN

13549 rows × 6 columns

4. Show all the records where No of Crimes is 0.How many such records are there?

```
In [39]: len(data[data.no_of_crimes == 0])
```

```
Out[39]: 104
```

```
In [40]: (data[data.no_of_crimes == 0])
```

```
Out[40]:
```

	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

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	date	area	average_price	code	houses_sold	no_of_crimes
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

5.what is maximum and minimum average_price per year in england?

In [42]:

```
data['year']=data.date.dt.year
```

In [43]:

```
data
```

Out[43]:

	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
...	...	...	...	...	...	...	...

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	date	area	average_price	code	houses_sold	no_of_crimes	year
13544	2019-09-01	england	249942	E92000001	64605.0	NaN	2019
13545	2019-10-01	england	249376	E92000001	68677.0	NaN	2019
13546	2019-11-01	england	248515	E92000001	67814.0	NaN	2019
13547	2019-12-01	england	250410	E92000001	NaN	NaN	2019
13548	2020-01-01	england	247355	E92000001	NaN	NaN	2020

13549 rows × 7 columns

In [53]:

```
df1=data[data.area == 'england']
```

In [54]:

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

Out[54]:

```
year
1995     52788
1996     52333
1997     55789
1998     61659
1999     65522
2000     75219
2001     84245
2002     96215
2003    121610
2004    139719
2005    158572
2006    166544
2007    181824
2008    165795
2009    159340
2010    174458
2011    173046
2012    174161
2013    176816
2014    188265
2015    202856
2016    220361
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
2017    231593
2018    240428
2019    243281
2020    247355
Name: average_price, dtype: int64
```

In [46]:

```
data
```

Out[46]:

	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
...	...	...	...	...	...	...	...
13544	2019-09-01	england	249942	E92000001	64605.0	NaN	2019
13545	2019-10-01	england	249376	E92000001	68677.0	NaN	2019
13546	2019-11-01	england	248515	E92000001	67814.0	NaN	2019
13547	2019-12-01	england	250410	E92000001	NaN	NaN	2019
13548	2020-01-01	england	247355	E92000001	NaN	NaN	2020

13549 rows × 7 columns

What is max and min number of crime rates?

In [56]:

```
data.groupby('area').no_of_crimes.max()
```

Out[56]:

```
area
barking and dagenham      2049.0
barnet                    2893.0
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
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
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
wandsworth                3051.0
west midlands                NaN
westminster               7461.0
yorks and the humber         NaN
Name: no_of_crimes, dtype: float64
```

In [57]:

```
data.groupby('area').no_of_crimes.min()
```

Out[57]:

```
area
barking and dagenham      1217.0
barnet                    1703.0
bexley                     860.0
brent                     1850.0
bromley                   1441.0
camden                    2079.0
city of london               0.0
croydon                   2031.0
ealing                    1871.0
east midlands                NaN
east of england              NaN
enfield                   1635.0
england                      NaN
greenwich                 1513.0
hackney                   1870.0
hammersmith and fulham    1323.0
haringey                  1536.0
harrow                     937.0
havering                  1130.0
hillingdon                1445.0
hounslow                  1529.0
inner london                 NaN
islington                 1871.0
kensington and chelsea    1347.0
kingston upon thames       692.0
lambeth                   2381.0
lewisham                  1675.0
london                       NaN
merton                     819.0
newham                    2130.0
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
north east                   NaN
north west                   NaN
outer london                 NaN
redbridge                 1487.0
richmond upon thames       700.0
south east                   NaN
south west                   NaN
southwark                 2267.0
sutton                     787.0
tower hamlets             1646.0
waltham forest            1575.0
wandsworth                1582.0
west midlands                NaN
westminster               3504.0
yorks and the humber         NaN
Name: no_of_crimes, dtype: float64
```

In [59]:
```
#show the total count of records of each area ,where the average price
 is less than 1Lakh
data[data.average_price < 100000].area.value_counts()
```

Out[59]:
```
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
bexley                   64
waltham forest           64
lewisham                 62
havering                 60
south east               59
greenwich                59
croydon                  57
enfield                  54
sutton                   54
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
sutton                   54
hackney                  53
redbridge                52
southwark                48
tower hamlets            47
outer london             46
hillingdon               44
hounslow                 41
lambeth                  41
brent                    40
london                   39
merton                   35
bromley                  33
haringey                 33
ealing                   31
inner london             31
harrow                   30
kingston upon thames     30
wandsworth               26
barnet                   25
islington                19
city of london           11
Name: area, dtype: int64
```

In [ ]:

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD