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

lhd = pd.read_csv(r"C:\Users\Ashish\OneDrive\Pyhton Project file ALL\
Project 5 London Housing Data Set.csv")

lhd

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

       no_of_crimes  
0               NaN  
1               NaN  
2               NaN  
3               NaN  
4               NaN  
...             ...  
13544           NaN  
13545           NaN  
13546           NaN  
13547           NaN  
13548           NaN  

[13549 rows x 6 columns]
```

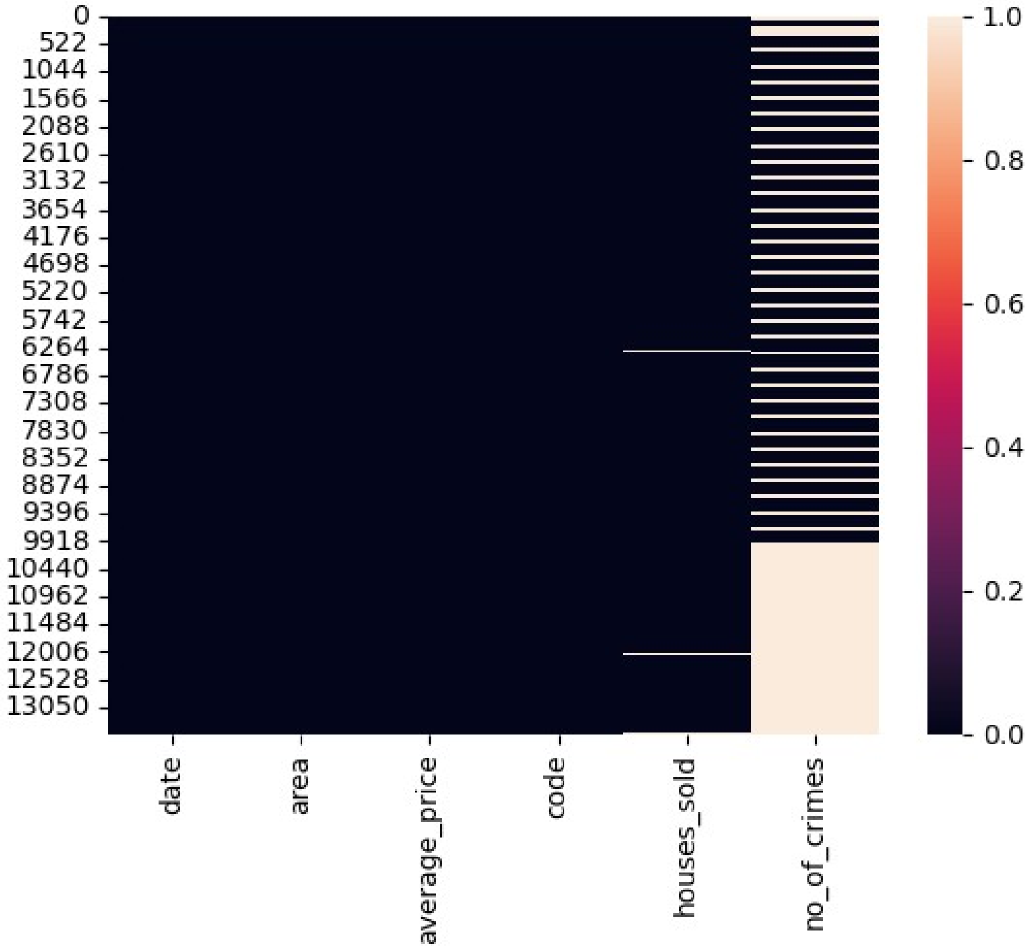
```
lhd.count()
```

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

```
lhd.isnull().sum()
```

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

```
sns.heatmap(lhd.isnull())
plt.show()
```



Convert the datatype of date column to date-time format?

```
lhd.head(2)

       date            area  average_price        code  houses_sold  \
0  1/1/1995  city of london          91449   E09000001         17.0
1  2/1/1995  city of london          82203   E09000001          7.0

   no_of_crimes
0           NaN
1           NaN

lhd.dtypes

date                 object
area                 object
average_price         int64
```

```
code                  object
houses_sold          float64
no_of_crimes         float64
dtype: object

# Command to chnage the data type ofthe coloumn!

lhd.date = pd.to_datetime(lhd.date)
```

Add a new coloumn "Year" in the dataframe, which contain yearas only?

```
lhd.head(2)

        date              area  average_price       code  houses_sold  \
0 1995-01-01  city of london          91449  E09000001         17.0
1 1995-02-01  city of london          82203  E09000001          7.0

   no_of_crimes
0           NaN
1           NaN

# Extract year\month coloumn form the date coloumn!

lhd["year"] = lhd.date.dt.year
# lhd["month"] = lhd.date.dt.month
```

Add a new coloumn "month" as 2nd coloumn in the dataframe, which contain month only?

```
# Command to insert new coloumn as specific postion

lhd.insert(1 , "month" , lhd.date.dt.month)

lhd.head(2)

        date  month              area  average_price       code
houses_sold  \
0 1995-01-01      1  city of london          91449  E09000001
17.0
1 1995-02-01      2  city of london          82203  E09000001
7.0

   no_of_crimes  year
```

```
0           NaN  1995
1           NaN  1995
```

Remove the coloumn "Year" and "Month" from the dataframe?

```
lhd.drop(["month" , "year"], axis = 1, inplace = True)

lhd.head(2)

        date             area  average_price       code  houses_sold  \
0 1995-01-01  city of london          91449  E09000001         17.0   
1 1995-02-01  city of london          82203  E09000001          7.0   

   no_of_crimes  
0           NaN  
1           NaN  
```

Show all record where number of crime is 0, and how many such record are their?

```
lhd.head(2)

        date             area  average_price       code  houses_sold  \
0 1995-01-01  city of london          91449  E09000001         17.0   
1 1995-02-01  city of london          82203  E09000001          7.0   

   no_of_crimes  
0           NaN  
1           NaN  

lhd[lhd.no_of_crimes == 0]
#len(lhd[lhd.no_of_crimes == 0])

            date             area  average_price       code  houses_sold
\
72    2001-01-01   city of london         284262   E09000001         24.0

73    2001-02-01   city of london         198137   E09000001         37.0

74    2001-03-01   city of london         189033   E09000001         44.0

75    2001-04-01   city of london         205494   E09000001         38.0

76    2001-05-01   city of london         223459   E09000001         30.0
```

```
..          ...             ...            ...         ...           ...
178 2009-11-01  city of london          397909  E09000001          11.0
179 2009-12-01  city of london          411955  E09000001          16.0
180 2010-01-01  city of london          464436  E09000001          20.0
181 2010-02-01  city of london          490525  E09000001           9.0
182 2010-03-01  city of london          498241  E09000001          15.0

     no_of_crimes
72            0.0
73            0.0
74            0.0
75            0.0
76            0.0
..            ...
178           0.0
179           0.0
180           0.0
181           0.0
182           0.0

[104 rows x 6 columns]
```

What is the Maximum and Minimum "Average_Price" per year in England?

```
lhd["year"] = lhd.date.dt.year

lhd.head(2)

        date            area  average_price        code  houses_sold  \
0 1995-01-01  city of london          91449  E09000001         17.0
1 1995-02-01  city of london          82203  E09000001          7.0

   no_of_crimes  year
0           NaN  1995
1           NaN  1995

engdf = lhd[lhd.area == "england"]

engdf
```

```
             date     area  average_price       code  houses_sold  \
13248 1995-01-01  england          53203  E92000001      47639.0
13249 1995-02-01  england          53096  E92000001      47880.0
13250 1995-03-01  england          53201  E92000001      67025.0
13251 1995-04-01  england          53591  E92000001      56925.0
13252 1995-05-01  england          53678  E92000001      64192.0
...          ...      ...            ...        ...          ...
13544 2019-09-01  england         249942  E92000001      64605.0
13545 2019-10-01  england         249376  E92000001      68677.0
13546 2019-11-01  england         248515  E92000001      67814.0
13547 2019-12-01  england         250410  E92000001          NaN
13548 2020-01-01  england         247355  E92000001          NaN

       no_of_crimes  year
13248           NaN  1995
13249           NaN  1995
13250           NaN  1995
13251           NaN  1995
13252           NaN  1995
...             ...   ...
13544           NaN  2019
13545           NaN  2019
13546           NaN  2019
13547           NaN  2019
13548           NaN  2020

[301 rows x 7 columns]
```

```
engdf.groupby("year").average_price.max().head(5)
```

```
year
1995    53901
1996    55755
1997    61564
1998    65743
1999    75071
Name: average_price, dtype: int64
```

```
engdf.groupby("year").average_price.min().head(5)
```

```
year
1995    52788
1996    52333
1997    55789
1998    61659
1999    65522
Name: average_price, dtype: int64
```

```
engdf.groupby("year").average_price.mean().round(2).head(5)
```

```
year
1995    53322.42
1996    54151.50
1997    59160.67
1998    64301.67
1999    70070.75
Name: average_price, dtype: float64
```

What is the Maximum and Minimum number of cases recorded per area?

```
lhd.head(2)
```

```
        date            area  average_price       code  houses_sold  \
0 1995-01-01  city of london          91449  E09000001         17.0   
1 1995-02-01  city of london          82203  E09000001          7.0   

   no_of_crimes  year  
0           NaN  1995  
1           NaN  1995  
```

```
lhd.groupby("area").no_of_crimes.max().head(5)
```

```
area
barking and dagenham    2049.0
barnet                  2893.0
bexley                  1914.0
brent                   2937.0
bromley                 2637.0
Name: no_of_crimes, dtype: float64
```

```
lhd.groupby("area").no_of_crimes.min().sort_values(ascending =
True).head(5)
```

```
area
city of london            0.0
kingston upon thames    692.0
richmond upon thames    700.0
sutton                  787.0
merton                  819.0
Name: no_of_crimes, dtype: float64
```

Show the total count of record of each area, where average price is less then 100000.

```
lhd.head()
```

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

   no_of_crimes  year  
0           NaN  1995  
1           NaN  1995  
2           NaN  1995  
3           NaN  1995  
4           NaN  1995  
```

```
lhd[lhd.average_price < 100000].area.value_counts()
```

```
area
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
hackney                    53
redbridge                  52
southwark                  48
tower hamlets              47
outer london               46
hillingdon                 44
lambeth                    41
hounslow                   41
```

```
brent                    40
london                   39
merton                   35
haringey                 33
bromley                  33
inner london             31
ealing                   31
kingston upon thames     30
harrow                   30
wandsworth               26
barnet                   25
islington                19
city of london           11
Name: count, dtype: int64
```