DATA OF HOUSING IN LONDON WAS ANALYSED AND QUESTIONS ANSWERED TO BETTER UNDERSTAND THE REAL ESTATE MARKET IN LONDON

QUESTION 1: CONVERT DATA TYPE OF DATE COLUMN TO DATE TIME FORMAT?

In [29]:
```
import pandas as pd

data= pd.read_csv("C:\\Users\\Joseph\\Desktop\\Data Files\\CSV Files\\housing_london_variables.csv")
data.head()
```

Out[29]:

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

Check the data types contained in the dataset

In [30]:
```
data.dtypes
```

Out[30]:
```
date              object
area              object
average_price      int64
code              object
houses_sold      float64
no_of_crimes     float64
borough_flag       int64
dtype: object
```

In [31]:
```
data["date"]= pd.to_datetime(data["date"])
```

In [32]:
```
data.dtypes
```

Out[32]:
```
date             datetime64[ns]
area                     object
average_price             int64
code                     object
houses_sold             float64
no_of_crimes            float64
borough_flag              int64
dtype: object
```

QUESTION 2: ADD A NEW COLUMN "YEAR" TO THE DATA FRAME WHICH CONTAINS YEAR ONLY

In [33]:
```
data["year"]= data["date"].dt.year
data.head()
```

Out[33]:

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

QUESTION 3: ADD NEW COLUMN "MONTH" TO THE SECOND COLUMN OF THE DATA FRAME WHICH CONTAIN MONTH ONLY?

In [35]:
```
data["month"]= data["date"].dt.month

data.head()
```

Out[35]:

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

In [46]:
```
data= data.iloc[:, [0,8,1,2,3,4,5,6,7]]
data.head(5)
```

Out[46]:

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

QUESTION 4: REMOVE THE COLUMN CODE AND HOUSES_SOLD FROM THE DATA FRAME

In [ ]:
```
data.drop(["code", "houses_sold"], axis=1, inplace= True)
```

In [60]:
```
data.head()
```

Out[60]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
0	1995-01-01	1	city of london	91449	NaN	1	1995
1	1995-02-01	2	city of london	82203	NaN	1	1995
2	1995-03-01	3	city of london	79121	NaN	1	1995
3	1995-04-01	4	city of london	77101	NaN	1	1995
4	1995-05-01	5	city of london	84409	NaN	1	1995

code and house_hold column has been dropped from the data frame

QUESTION 5: SHOW ALL THE RECORD WHERE NUMBER OF CRIME IS ZERO, HOW MANY OF SUCH RECORD ARE THERE?

In [61]:
```
data.head(5)
```

Out[61]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
0	1995-01-01	1	city of london	91449	NaN	1	1995
1	1995-02-01	2	city of london	82203	NaN	1	1995
2	1995-03-01	3	city of london	79121	NaN	1	1995
3	1995-04-01	4	city of london	77101	NaN	1	1995
4	1995-05-01	5	city of london	84409	NaN	1	1995

In [63]:
```
crim_data= data[data["no_of_crimes"]== 0]
crim_data.head()
```

Out[63]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
72	2001-01-01	1	city of london	284262	0.0	1	2001
73	2001-02-01	2	city of london	198137	0.0	1	2001
74	2001-03-01	3	city of london	189033	0.0	1	2001
75	2001-04-01	4	city of london	205494	0.0	1	2001
76	2001-05-01	5	city of london	223459	0.0	1	2001

In [64]:
```
len(crim_data)
```

Out[64]:
```
104
```

There are 104 records where number of crime is 0

QUESTION 6: WHAT IS THE MINIMUM AND MAXIMUM "AVERAGE_PRICE" IN ENGLAND PER YEAR?

In [74]:
```
eng_data= data[data["area"]== "england"]
eng_data.head()
```

Out[74]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
13248	1995-01-01	1	england	53203	NaN	0	1995
13249	1995-02-01	2	england	53096	NaN	0	1995
13250	1995-03-01	3	england	53201	NaN	0	1995
13251	1995-04-01	4	england	53591	NaN	0	1995
13252	1995-05-01	5	england	53678	NaN	0	1995

In [77]:
```
eng_data.groupby("year")["average_price"].max()
```

Out[77]:
```
year
1995     53901
1996     55755
1997     61564
1998     65743
1999     75071
2000     84191
2001     95992
2002    119982
2003    138985
2004    160330
2005    167244
2006    182031
2007    194764
2008    191750
2009    174136
2010    180807
2011    177335
2012    180129
2013    188544
2014    203639
2015    219582
2016    231922
2017    242628
2018    248620
2019    250410
2020    247355
Name: average_price, dtype: int64
```

The maximum average price of houses in england is shown above

In [79]:
```
eng_data.groupby("year")["average_price"].min()
```

Out[79]:
```
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
2017    231593
2018    240428
2019    243281
2020    247355
Name: average_price, dtype: int64
```

The minimum average price of houses in england is shown above

QUESTION 7: WHAT IS THE MAXIMUM AND MINIMUM NUMBER OF CRIMES RECORDED PER AREA?

In [68]:
```
data.head()
```

Out[68]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
0	1995-01-01	1	city of london	91449	NaN	1	1995
1	1995-02-01	2	city of london	82203	NaN	1	1995
2	1995-03-01	3	city of london	79121	NaN	1	1995
3	1995-04-01	4	city of london	77101	NaN	1	1995
4	1995-05-01	5	city of london	84409	NaN	1	1995

In [81]:
```
data.groupby("area")["no_of_crimes"].max()
```

Out[81]:
```
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
```

The maximum number of crime per area is shown above

In [84]:
```
data.groupby("area")["no_of_crimes"].min()
```

Out[84]:
```
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
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
```

The minimum number of crime per area is shown above

QUESTION 8: SHOW THE TOTAL COUNTS OF RECORD FOR EACH AREA WHERE THE AVERAGE PRICE IS LESS THAN 100000?

In [85]:
```
data.head()
```

Out[85]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
0	1995-01-01	1	city of london	91449	NaN	1	1995
1	1995-02-01	2	city of london	82203	NaN	1	1995
2	1995-03-01	3	city of london	79121	NaN	1	1995
3	1995-04-01	4	city of london	77101	NaN	1	1995
4	1995-05-01	5	city of london	84409	NaN	1	1995

In [87]:
```
area_data= data[data["average_price"] < 100000]
area_data
```

Out[87]:

	date	month	area	average_price	no_of_crimes	borough_flag	year
0	1995-01-01	1	city of london	91449	NaN	1	1995
1	1995-02-01	2	city of london	82203	NaN	1	1995
2	1995-03-01	3	city of london	79121	NaN	1	1995
3	1995-04-01	4	city of london	77101	NaN	1	1995
4	1995-05-01	5	city of london	84409	NaN	1	1995
...	...	...	...	...	...	...	...
13330	2001-11-01	11	england	95083	NaN	0	2001
13331	2001-12-01	12	england	95992	NaN	0	2001
13332	2002-01-01	1	england	96215	NaN	0	2002
13333	2002-02-01	2	england	96676	NaN	0	2002
13334	2002-03-01	3	england	98962	NaN	0	2002

2209 rows × 7 columns

In [88]:
```
area_data["area"].value_counts()
```

Out[88]:
```
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: area, dtype: int64
```

In [ ]: