In [1]:
import pandas as pd

In [10]:

In [10]:

df=pd.read\_csv(r"C:\Users\sadiy\Downloads\Datascience\pandas\dataframe\home\_data.csv")

In [11]:

df

Out[11]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view		grade	sqft_above	sqft_basement
0	7129300520	20141013T000000	221900	3	1.00	1180	5650	1.0	0	0		7	1180	0
1	6414100192	20141209T000000	538000	3	2.25	2570	7242	2.0	0	0		7	2170	400
2	5631500400	20150225T000000	180000	2	1.00	770	10000	1.0	0	0		6	770	0
3	2487200875	20141209T000000	604000	4	3.00	1960	5000	1.0	0	0		7	1050	910
4	1954400510	20150218T000000	510000	3	2.00	1680	8080	1.0	0	0		8	1680	0
21608	263000018	20140521T000000	360000	3	2.50	1530	1131	3.0	0	0		8	1530	0
21609	6600060120	20150223T000000	400000	4	2.50	2310	5813	2.0	0	0		8	2310	0
21610	1523300141	20140623T000000	402101	2	0.75	1020	1350	2.0	0	0		7	1020	0
21611	291310100	20150116T000000	400000	3	2.50	1600	2388	2.0	0	0		8	1600	0
21612	1523300157	20141015T000000	325000	2	0.75	1020	1076	2.0	0	0		7	1020	0
21612	rows x 21 co	lumne												
	21613 rows × 21 columns													
4														<b>)</b>

In [12]:

df.head()

Out[12]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view		grade	sqft_above	sqft_basement	yr_bui
0	7129300520	20141013T000000	221900	3	1.00	1180	5650	1.0	0	0		7	1180	0	195
1	6414100192	20141209T000000	538000	3	2.25	2570	7242	2.0	0	0		7	2170	400	195
2	5631500400	20150225T000000	180000	2	1.00	770	10000	1.0	0	0		6	770	0	193
3	2487200875	20141209T000000	604000	4	3.00	1960	5000	1.0	0	0		7	1050	910	196
4	1954400510	20150218T000000	510000	3	2.00	1680	8080	1.0	0	0		8	1680	0	198
i r	rows × 21 columns														
	4														•

```
In [14]:
                                                                                                                                      M
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21613 entries, 0 to 21612
Data columns (total 21 columns):
    Column
                   Non-Null Count Dtype
a
    id
                   21613 non-null int64
1
     date
                   21613 non-null
                                   object
     price
                   21613 non-null
                                   int64
     bedrooms
                   21613 non-null
                                   int64
     bathrooms
                   21613 non-null
                                   float64
     sqft_living
                   21613 non-null
                                   int64
     sqft_lot
                   21613 non-null
    floors
                   21613 non-null
                                   float64
8
     waterfront
                   21613 non-null
     view
                   21613 non-null
10 condition
                   21613 non-null
11
     grade
                   21613 non-null
                                   int64
    sqft_above
                   21613 non-null
     sqft_basement 21613 non-null
    yr_built
                   21613 non-null
15
    yr_renovated
                   21613 non-null
                                   int64
    zipcode
                   21613 non-null
                                   int64
17
     lat
                   21613 non-null
                                   float64
18 long
                   21613 non-null float64
    sqft_living15 21613 non-null
                                   int64
19
20 sqft_lot15
                   21613 non-null int64
dtypes: float64(4), int64(16), object(1)
memory usage: 3.5+ MB
In [16]:
                                                                                                                                      M
df.shape
Out[16]:
(21613, 21)
no of records
df.shape[0]
                                                                                                                                      M
In [18]:
df.bathrooms
Out[18]:
0
        1.00
1
        2.25
        1.00
3
         3.00
4
         2.00
21608
         2.50
21609
```

## changing the values in numeric

Name: bathrooms, Length: 21613, dtype: float64

0.75

2.50 0.75

21610 21611

21612

```
In [27]:
df[['bathrooms', 'floors']] = df[['bathrooms', 'floors']].apply(pd.to_numeric)
```

```
In [28]:
                                                                                                                                               M
df.dtypes
Out[28]:
id
                    int64
                   object
date
price
                    int64
                    int64
bedrooms
                  float64
bathrooms
sqft_living
                    int64
sqft_lot
                    int64
floors
                  float64
waterfront
                    int64
view
                    int64
condition
                    int64
grade
                    int64
sqft_above
                    int64
sqft_basement
                    int64
yr_built
                    int64
yr_renovated
                    int64
zipcode
                    int64
lat
                  float64
long
                  float64
sqft_living15
                    int64
sqft_lot15
                    int64
dtype: object
                                                                                                                                               M
In [29]:
df
Out[29]:
              id
                            date
                                   price
                                        bedrooms bathrooms sqft_living sqft_lot floors
                                                                                   waterfront view
                                                                                                  ... grade sqft_above sqft_basement
    0 7129300520 20141013T000000
                                                                                           0
                                                                                                0
                                                                                                                 1180
                                                                                                                                 0
                                 221900
                                               3
                                                       1.00
                                                                 1180
                                                                        5650
                                                                                1.0
    1 6414100192 20141209T000000
                                                       2.25
                                                                 2570
                                                                                           0
                                                                                                0
                                                                                                                 2170
                                                                                                                               400
                                               3
                                                                        7242
                                                                                2.0
    2 5631500400 20150225T000000
                                 180000
                                               2
                                                       1.00
                                                                 770
                                                                       10000
                                                                                1.0
                                                                                           0
                                                                                                0
                                                                                                         6
                                                                                                                  770
                                                                                                                                 0
                                                                                                         7
    3 2487200875 20141209T000000 604000
                                                                        5000
                                                                                           0
                                                                                                0 ...
                                                                                                                               910
                                                       3.00
                                                                 1960
                                                                                1.0
                                                                                                                 1050
       1954400510 20150218T000000 510000
                                               3
                                                       2.00
                                                                 1680
                                                                        8080
                                                                                1.0
                                                                                           0
                                                                                                0
                                                                                                         8
                                                                                                                 1680
                                                                                                                                 0
 21608
       263000018 20140521T000000 360000
                                               3
                                                       2.50
                                                                 1530
                                                                         1131
                                                                                3.0
                                                                                           0
                                                                                                0 ...
                                                                                                         8
                                                                                                                 1530
                                                                                                                                 0
       6600060120 20150223T000000
                                 400000
                                               4
                                                       2.50
                                                                 2310
                                                                        5813
                                                                                2.0
                                                                                           0
                                                                                                0
                                                                                                         8
                                                                                                                 2310
                                                                                                                                 0
                                                                                                         7
                                               2
                                                       0.75
                                                                                2.0
                                                                                           0
                                                                                                0 ...
                                                                                                                                 0
 21610 1523300141 20140623T000000 402101
                                                                 1020
                                                                        1350
                                                                                                                 1020
                                                                                                                                 0
       291310100 20150116T000000 400000
                                               3
                                                       2.50
                                                                 1600
                                                                        2388
                                                                                2.0
                                                                                           0
                                                                                                0 ...
                                                                                                         8
                                                                                                                 1600
 21612 1523300157 20141015T000000 325000
                                               2
                                                       0.75
                                                                 1020
                                                                        1076
                                                                                2.0
                                                                                           0
                                                                                                0
                                                                                                                 1020
                                                                                                                                 0
21613 rows × 21 columns
In [21]:
                                                                                                                                               M
df.columns
Out[21]:
In [40]:
                                                                                                                                               M
```

In [44]:

df["floors"]=df["floors"].astype(int)

In [39]: M df Out[39]: bedrooms bathrooms sqft\_living sqft\_lot floors waterfront view price ... grade sqft\_above sqft\_basement yr\_built yr\_r 7129300520 20141013T000000 1.00 1.0 6414100192 20141209T000000 2.25 2.0 5631500400 20150225T000000 1.00 1.0 0 ... 3 2487200875 20141209T000000 604000 3.00 1.0 0 ... 0 ... 1954400510 20150218T000000 510000 2.00 1.0 263000018 20140521T000000 360000 2.50 3.0 6600060120 20150223T000000 400000 2.50 2.0 0 ... 21610 1523300141 20140623T000000 402101 0.75 M In [42]: df["bathrooms"]=df["bathrooms"].astype(int) In [43]: df Out[43]: date price bedrooms bathrooms sqft\_living sqft\_lot floors waterfront view ... grade sqft\_above sqft\_basement 7129300520 20141013T000000 221900 6414100192 20141209T000000 2.0 5631500400 20150225T000000 1.0 3 2487200875 20141209T000000 604000 1.0 4 1954400510 20150218T000000 510000 0 ... 1.0 263000018 20140521T000000 360000 3.0 0 ... 21609 6600060120 20150223T000000 400000 2.0 0 ... 1523300141 20140623T000000 291310100 20150116T000000 400000 0 ... 1523300157 20141015T000000 325000 0 ... 2.0 21613 rows × 21 columns

localhost:8888/notebooks/Downloads/Datascience/pandas/dataframe/Sadiya Desao assignment.ipynb#

M

```
In [45]:
                                                                                                                                                M
df
Out[45]:
                                   price bedrooms bathrooms sqft_living sqft_lot floors waterfront view ... grade sqft_above sqft_basement
    0 7129300520 20141013T000000
                                 221900
                                               3
                                                          1
                                                                 1180
                                                                         5650
                                                                                           0
                                                                                                                  1180
                                                                                                                                  0
    1 6414100192 20141209T000000 538000
                                                          2
                                                                                                                  2170
                                                3
                                                                 2570
                                                                         7242
                                                                                  2
                                                                                           0
                                                                                                 0 ...
                                                                                                                                400
    2 5631500400 20150225T000000 180000
                                                          1
                                                                  770
                                                                        10000
                                                                                           0
                                                                                                 0 ...
                                                                                                          6
                                                                                                                  770
                                                                                                                                  0
                                                                                                          7
    3 2487200875 20141209T000000 604000
                                               4
                                                          3
                                                                 1960
                                                                         5000
                                                                                           0
                                                                                                 0 ...
                                                                                                                  1050
                                                                                                                                910
                                                          2
                                                                                                 0 ...
    4 1954400510 20150218T000000 510000
                                               3
                                                                 1680
                                                                         8080
                                                                                           0
                                                                                                          8
                                                                                                                  1680
                                                                                                                                  0
 21608
       263000018 20140521T000000 360000
                                               3
                                                          2
                                                                 1530
                                                                         1131
                                                                                  3
                                                                                           0
                                                                                                 0 ...
                                                                                                          8
                                                                                                                  1530
                                                                                                                                  0
 21609 6600060120 20150223T000000 400000
                                                          2
                                                                         5813
                                                                                           0
                                                                                                 0 ...
                                                                                                          8
                                                                                                                  2310
                                                                                                                                  0
                                               4
                                                                 2310
                                                                                  2
 21610 1523300141 20140623T000000 402101
                                               2
                                                          0
                                                                         1350
                                                                                           0
                                                                                                                  1020
                                                                                                                                  0
                                                          2
                                                                                                          8
 21611
       291310100 20150116T000000 400000
                                               3
                                                                 1600
                                                                         2388
                                                                                  2
                                                                                           0
                                                                                                 0 ...
                                                                                                                  1600
                                                                                                                                  0
 21612 1523300157 20141015T000000 325000
                                               2
                                                          0
                                                                 1020
                                                                         1076
                                                                                  2
                                                                                           0
                                                                                                 0 ...
                                                                                                                  1020
                                                                                                                                  0
21613 rows × 21 columns
4
In [47]:
                                                                                                                                                Ы
df["bedrooms"].value_counts()
Out[47]:
3
      9824
4
      6882
      2760
5
      1601
       199
0
        13
8
9
Name: bedrooms, dtype: int64
In [48]:
                                                                                                                                                M
df["floors"].unique()
Out[48]:
array([1, 2, 3])
In [49]:
                                                                                                                                                M
df["floors"].nunique()
Out[49]:
total no of bedrooms
In [50]:
                                                                                                                                                M
df["bedrooms"].unique()
Out[50]:
array([ 3, 2, 4, 5, 1, 6, 7, 0, 8, 9, 11, 10, 33])
In [52]:
                                                                                                                                                M
df["bedrooms"].nunique()
Out[52]:
13
```

```
In [56]:
                                                                                                                                                     M
df.isna().sum()
Out[56]:
                  0
id
                  0
date
                  a
price
bedrooms
                  0
bathrooms
                  0
sqft_living
                  0
sqft_lot
                  a
floors
                  0
waterfront
                  0
view
                  0
condition
grade
                  0
sqft_above
sqft_basement
                  0
yr_built
                  0
yr_renovated
                  0
zipcode
                  0
lat
                  0
long
sqft_living15
                  0
sqft_lot15
dtype: int64
                                                                                                                                                     M
In [57]:
df
Out[57]:
               id
                             date
                                    price
                                         bedrooms bathrooms sqft_living sqft_lot floors
                                                                                       waterfront view
                                                                                                      ... grade sqft_above sqft_basement
    0 7129300520 20141013T000000
                                                                                               0
                                                                                                                      1180
                                                                                                                                      0
                                  221900
                                                 3
                                                                    1180
                                                                           5650
                                                                                                    0
    1 6414100192 20141209T000000
                                                            2
                                                                   2570
                                                                                               0
                                                                                                    0
                                                                                                                      2170
                                                                                                                                     400
                                                 3
                                                                           7242
    2 5631500400 20150225T000000
                                  180000
                                                 2
                                                            1
                                                                    770
                                                                          10000
                                                                                               0
                                                                                                    0
                                                                                                              6
                                                                                                                      770
                                                                                                                                      0
                                                                                                              7
    3 2487200875 20141209T000000 604000
                                                 4
                                                            3
                                                                                               0
                                                                                                    0 ...
                                                                                                                                    910
                                                                   1960
                                                                           5000
                                                                                                                      1050
       1954400510 20150218T000000 510000
                                                 3
                                                            2
                                                                    1680
                                                                           8080
                                                                                               0
                                                                                                    0 ...
                                                                                                              8
                                                                                                                      1680
                                                                                                                                      0
21608
       263000018 20140521T000000 360000
                                                 3
                                                            2
                                                                   1530
                                                                            1131
                                                                                     3
                                                                                              0
                                                                                                    0 ...
                                                                                                              8
                                                                                                                      1530
                                                                                                                                      0
       6600060120 20150223T000000
                                                 4
                                                            2
                                                                   2310
                                                                           5813
                                                                                     2
                                                                                               0
                                                                                                    0
                                                                                                              8
                                                                                                                      2310
                                                                                                                                      0
                                                            0
                                                                                                              7
                                                 2
                                                                                     2
                                                                                              0
                                                                                                    0 ...
                                                                                                                                      0
 21610 1523300141 20140623T000000 402101
                                                                   1020
                                                                           1350
                                                                                                                      1020
                                                                                                                                      0
       291310100 20150116T000000 400000
                                                 3
                                                                    1600
                                                                           2388
                                                                                     2
                                                                                               0
                                                                                                    0 ...
                                                                                                              8
                                                                                                                      1600
 21612 1523300157 20141015T000000 325000
                                                 2
                                                            0
                                                                    1020
                                                                            1076
                                                                                     2
                                                                                               0
                                                                                                    0
                                                                                                                      1020
                                                                                                                                      0
21613 rows × 21 columns
In [59]:
                                                                                                                                                     M
s=df["date"]
                                                                                                                                                     M
In [60]:
s
Out[60]:
a
          20141013T000000
1
          20141209T000000
2
          20150225T000000
3
          20141209T000000
4
          20150218T000000
          20140521T000000
21608
21609
          20150223T000000
21610
          20140623T000000
21611
          20150116T000000
21612
          20141015T000000
Name: date, Length: 21613, dtype: object
In [67]:
                                                                                                                                                     M
s=s.str.replace("000000"," ")
```

```
In [68]:
                                                                                                                                                           M
s
Out[68]:
0
          20141013T
1
          20141209T
          20150225T
2
3
          20141209T
4
          20150218T
          20140521T
21608
21609
          20150223T
21610
          20140623T
21611
          20150116T
21612
          20141015T
Name: date, Length: 21613, dtype: object
In [69]:
                                                                                                                                                           M
df
Out[69]:
                                           bedrooms bathrooms sqft_living sqft_lot floors
                id
                                                                                           waterfront view
                                                                                                              grade sqft_above sqft_basement
                              date
                                      price
                                                                                                                                            0
    0 7129300520 20141013T000000
                                    221900
                                                    3
                                                                      1180
                                                                               5650
                                                                                                   0
                                                                                                        0
                                                                                                                           1180
     1 6414100192 20141209T000000
                                    538000
                                                    3
                                                              2
                                                                      2570
                                                                               7242
                                                                                                   0
                                                                                                        0
                                                                                                                          2170
                                                                                                                                          400
    2 5631500400 20150225T000000
                                                              1
                                                                                                        0
                                                                                                                            770
                                                                                                                                            0
                                    180000
                                                   2
                                                                       770
                                                                              10000
                                                                                                   0
                                                                                                                  6
       2487200875 20141209T000000
                                                              3
                                                                               5000
                                                                                                                                          910
                                                                      1960
                                                                                                                           1050
                                                              2
                                                                                                        0 ...
       1954400510 20150218T000000 510000
                                                   3
                                                                      1680
                                                                               8080
                                                                                                   0
                                                                                                                  8
                                                                                                                           1680
                                                                                                                                            0
 21608
        263000018 20140521T000000 360000
                                                              2
                                                                      1530
                                                                               1131
                                                                                        3
                                                                                                   0
                                                                                                        0 ...
                                                                                                                          1530
                                                                                                                                            0
                                                              2
 21609
       6600060120 20150223T000000
                                    400000
                                                                      2310
                                                                               5813
                                                                                        2
                                                                                                   0
                                                                                                        0
                                                                                                                  8
                                                                                                                          2310
                                                                                                                                            0
       1523300141 20140623T000000
                                    402101
                                                              0
                                                                      1020
                                                                               1350
                                                                                        2
                                                                                                  0
                                                                                                                           1020
                                                                                                                                            0
 21610
                                                                                                        0 ...
        291310100 20150116T000000 400000
                                                    3
                                                              2
                                                                      1600
                                                                               2388
                                                                                        2
                                                                                                   0
                                                                                                                  8
                                                                                                                           1600
                                                                                                                                            0
                                                              0
                                                                                                        0 ...
 21612
       1523300157 20141015T000000 325000
                                                                      1020
                                                                               1076
                                                                                        2
                                                                                                   0
                                                                                                                           1020
                                                                                                                                            0
21613 rows × 21 columns
                                                                                                                                                           M
In [70]:
df['date']=s
In [71]:
                                                                                                                                                           M
df
Out[71]:
                                                bathrooms sqft_living sqft_lot floors
                                                                                     waterfront
                                                                                                        grade
                                                                                                               sqft_above sqft_basement
                                                                                                                                         yr_bui
    0 7129300520 20141013T
                              221900
                                             3
                                                         1
                                                                 1180
                                                                         5650
                                                                                             0
                                                                                                  0
                                                                                                                     1180
                                                                                                                                      0
                                                                                                                                           195
                                                        2
                                                                                             0
                                                                                                  0
                                                                                                                    2170
                                                                                                                                           195
     1 6414100192 20141209T 538000
                                             3
                                                                2570
                                                                        7242
                                                                                  2
                                                                                                                                    400
       5631500400 20150225T
                                                                 770
                                                                                             0
                                                                                                                                           193
                                                                        10000
                                                                                                            6
                                                                                                                      770
                                                                                                                                      0
     3
       2487200875 20141209T 604000
                                             4
                                                        3
                                                                1960
                                                                        5000
                                                                                             0
                                                                                                  0
                                                                                                                     1050
                                                                                                                                    910
                                                                                                                                           196
       1954400510 20150218T 510000
                                                                1680
                                                                                             0
                                                                                                  0
                                                                                                                                      0
                                                                                                                                           198
                                                                        8080
                                                                                                            8
                                                                                                                     1680
 21608
        263000018 20140521T 360000
                                             3
                                                        2
                                                                1530
                                                                         1131
                                                                                  3
                                                                                             0
                                                                                                  Ω
                                                                                                            8
                                                                                                                     1530
                                                                                                                                      0
                                                                                                                                           200
 21609
       6600060120 20150223T 400000
                                                                2310
                                                                        5813
                                                                                                                     2310
                                                                                                                                      0
                                                                                                                                           201
 21610
       1523300141 20140623T 402101
                                                        0
                                                                1020
                                                                         1350
                                                                                  2
                                                                                             0
                                                                                                  0
                                                                                                                     1020
                                                                                                                                      0
                                                                                                                                           200
        291310100 20150116T 400000
                                                        2
                                                                                            0
                                             3
                                                                1600
                                                                        2388
                                                                                  2
                                                                                                  0 ...
                                                                                                            8
                                                                                                                     1600
                                                                                                                                      0
                                                                                                                                           200
 21611
 21612 1523300157 20141015T 325000
                                                        0
                                                                1020
                                                                         1076
                                                                                                  0
                                                                                                                     1020
                                                                                                                                           200
21613 rows × 21 columns
4
In [74]:
                                                                                                                                                           M
df.drop(['waterfront','view'],axis=1,inplace=True)
```

```
In [75]:
                                                                                                                                                    M
df
Out[75]:
                              price bedrooms bathrooms sqft_living sqft_lot floors condition grade sqft_above sqft_basement yr_built yr_renov
    0 7129300520 20141013T 221900
                                           3
                                                      1
                                                              1180
                                                                     5650
                                                                                        3
                                                                                              7
                                                                                                      1180
                                                                                                                            1955
    1 6414100192 20141209T 538000
                                           3
                                                      2
                                                                                              7
                                                                                                      2170
                                                                                                                            1951
                                                             2570
                                                                     7242
                                                                              2
                                                                                        3
                                                                                                                     400
                                                                                                       770
    2 5631500400 20150225T 180000
                                                              770
                                                                    10000
                                                                                        3
                                                                                              6
                                                                                                                       0
                                                                                                                            1933
    3 2487200875 20141209T 604000
                                           4
                                                      3
                                                             1960
                                                                     5000
                                                                                        5
                                                                                              7
                                                                                                      1050
                                                                                                                     910
                                                                                                                            1965
    4 1954400510 20150218T 510000
                                           3
                                                      2
                                                             1680
                                                                     8080
                                                                                        3
                                                                                              8
                                                                                                      1680
                                                                                                                       0
                                                                                                                            1987
                                                                              1
 21608
       263000018 20140521T 360000
                                           3
                                                      2
                                                              1530
                                                                     1131
                                                                              3
                                                                                        3
                                                                                              8
                                                                                                      1530
                                                                                                                       0
                                                                                                                            2009
 21609 6600060120 20150223T 400000
                                           4
                                                      2
                                                             2310
                                                                                              8
                                                                                                      2310
                                                                                                                       0
                                                                                                                            2014
                                                                     5813
                                                                              2
                                                                                        3
 21610 1523300141 20140623T 402101
                                           2
                                                      0
                                                             1020
                                                                     1350
                                                                                              7
                                                                                                      1020
                                                                                                                       0
                                                                                                                            2009
                                                      2
 21611
       291310100 20150116T 400000
                                           3
                                                             1600
                                                                     2388
                                                                              2
                                                                                        3
                                                                                              8
                                                                                                      1600
                                                                                                                       0
                                                                                                                            2004
 21612 1523300157 20141015T 325000
                                                      0
                                                             1020
                                                                     1076
                                                                              2
                                                                                        3
                                                                                                      1020
                                                                                                                       0
                                                                                                                            2008
21613 rows × 19 columns
In [76]:
                                                                                                                                                    M
s
Out[76]:
          20141013T
0
1
          20141209T
2
          20150225T
3
          20141209T
4
          20150218T
21608
          20140521T
21609
          20150223T
21610
          20140623T
          20150116T
21611
21612
         20141015T
Name: date, Length: 21613, dtype: object
In [77]:
                                                                                                                                                    M
s=s.str.replace("T"," ")
In [78]:
                                                                                                                                                    M
s
Out[78]:
0
          20141013
          20141209
1
2
          20150225
          20141209
3
          20150218
4
          20140521
21608
21609
          20150223
21610
          20140623
          20150116
21611
21612
         20141015
Name: date, Length: 21613, dtype: object
In [79]:
                                                                                                                                                    M
df["date"]=s
```

```
In [80]:
                                                                                                                                                     M
s
Out[80]:
0
          20141013
          20141209
1
2
          20150225
          20141209
3
          20150218
4
          20140521
21608
          20150223
21609
21610
          20140623
21611
          20150116
21612
          20141015
Name: date, Length: 21613, dtype: object
In [81]:
                                                                                                                                                     M
df
Out[81]:
               id
                       date
                             price bedrooms bathrooms sqft_living sqft_lot floors condition grade sqft_above sqft_basement yr_built yr_renova
    0 7129300520 20141013 221900
                                           3
                                                             1180
                                                                     5650
                                                                                                      1180
                                                                                                                      0
                                                                                                                            1955
    1 6414100192 20141209 538000
                                           3
                                                     2
                                                             2570
                                                                     7242
                                                                              2
                                                                                       3
                                                                                              7
                                                                                                      2170
                                                                                                                     400
                                                                                                                            1951
                                                                                                                                        1
                                                                                                                      0
    2 5631500400 20150225 180000
                                           2
                                                     1
                                                                                       3
                                                                                              6
                                                                                                       770
                                                                                                                            1933
                                                              770
                                                                    10000
                                                                              1
    3 2487200875 20141209 604000
                                                     3
                                                             1960
                                                                                                      1050
                                                                                                                     910
                                                                                                                            1965
                                                                     5000
                                                     2
                                                                                       3
    4 1954400510 20150218 510000
                                           3
                                                             1680
                                                                     8080
                                                                                              8
                                                                                                      1680
                                                                                                                      0
                                                                                                                            1987
 21608
       263000018 20140521 360000
                                           3
                                                     2
                                                             1530
                                                                     1131
                                                                              3
                                                                                       3
                                                                                              8
                                                                                                      1530
                                                                                                                      0
                                                                                                                            2009
                                                     2
                                                                              2
                                                                                       3
                                                                                                                      0
 21609
       6600060120 20150223 400000
                                                             2310
                                                                     5813
                                                                                              8
                                                                                                      2310
                                                                                                                            2014
       1523300141 20140623 402101
                                                     0
                                                             1020
                                                                     1350
                                                                              2
                                                                                       3
                                                                                                      1020
                                                                                                                      0
                                                                                                                            2009
 21610
        291310100 20150116 400000
                                           3
                                                     2
                                                             1600
                                                                     2388
                                                                              2
                                                                                       3
                                                                                              8
                                                                                                      1600
                                                                                                                      0
                                                                                                                            2004
 21612 1523300157 20141015 325000
                                           2
                                                     0
                                                                     1076
                                                                              2
                                                                                       3
                                                                                                      1020
                                                                                                                      0
                                                             1020
                                                                                                                            2008
21613 rows × 19 columns
                                                                                                                                                     M
In [82]:
Out[82]:
0
          20141013
1
          20141209
2
          20150225
3
          20141209
4
          20150218
21608
          20140521
21609
          20150223
21610
          20140623
          20150116
          20141015
21612
Name: date, Length: 21613, dtype: object
In [83]:
                                                                                                                                                     M
s.str.split()
Out[83]:
0
          [20141013]
1
          [20141209]
2
          [20150225]
3
          [20141209]
          [20150218]
21608
          [20140521]
21609
          [20150223]
21610
          [20140623]
21611
          [20150116]
          [20141015]
21612
Name: date, Length: 21613, dtype: object
```

```
In [87]:
                                                                                                                                           M
m="1235"
print(list(m.split))
                                           Traceback (most recent call last)
\verb|-AppDataLocalTemp\&ipykernel_12536&4284156713.py in <module>|
      1 m="1235"
----> 2 print(list(m.split))
TypeError: 'builtin_function_or_method' object is not iterable
In [88]:
                                                                                                                                           M
Out[88]:
'1235'
In [90]:
str=s.str.split()
In [91]:
                                                                                                                                           M
str
Out[91]:
0
         [20141013]
1
         [20141209]
2
         [20150225]
3
         [20141209]
4
         [20150218]
21608
         [20140521]
21609
         [20150223]
21610
         [20140623]
21611
         [20150116]
21612
         [20141015]
Name: date, Length: 21613, dtype: object
In [92]:
                                                                                                                                           M
str[0]
Out[92]:
['20141013']
In [95]:
                                                                                                                                           Ы
str[0]=str[0][-1]
In [96]:
                                                                                                                                           H
str[0]
Out[96]:
'20141013'
In [99]:
                                                                                                                                           H
  File "C:\Users\sadiy\AppData\Local\Temp\ipykernel_12536\1640769444.py", line 1
    for i in str[0]
SyntaxError: invalid syntax
```

In [100]:

df

Out[100]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	condition	grade	sqft_above	sqft_basement	yr_built	yr_renova
0	7129300520	20141013	221900	3	1	1180	5650	1	3	7	1180	0	1955	
1	6414100192	20141209	538000	3	2	2570	7242	2	3	7	2170	400	1951	1
2	5631500400	20150225	180000	2	1	770	10000	1	3	6	770	0	1933	
3	2487200875	20141209	604000	4	3	1960	5000	1	5	7	1050	910	1965	
4	1954400510	20150218	510000	3	2	1680	8080	1	3	8	1680	0	1987	
				•••							***			
21608	263000018	20140521	360000	3	2	1530	1131	3	3	8	1530	0	2009	
21609	6600060120	20150223	400000	4	2	2310	5813	2	3	8	2310	0	2014	
21610	1523300141	20140623	402101	2	0	1020	1350	2	3	7	1020	0	2009	
21611	291310100	20150116	400000	3	2	1600	2388	2	3	8	1600	0	2004	
21612	1523300157	20141015	325000	2	0	1020	1076	2	3	7	1020	0	2008	
21613	21613 rows × 19 columns													
4														•

## sort according to price

In [102]:

df.sort\_values(["price"],inplace=True)

In [103]:

df

Out[103]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	condition	grade	sqft_above	sqft_basement	yr_built	yr_reno\
1149	3421079032	20150217	75000	1	0	670	43377	1	3	3	670	0	1966	
15293	40000362	20140506	78000	2	1	780	16344	1	1	5	780	0	1942	
465	8658300340	20140523	80000	1	0	430	5050	1	2	4	430	0	1912	
16198	3028200080	20150324	81000	2	1	730	9975	1	1	5	730	0	1943	
8274	3883800011	20141105	82000	3	1	860	10426	1	3	6	860	0	1954	
1448	8907500070	20150413	5350000	5	5	8000	23985	2	3	12	6720	1280	2009	
4411	2470100110	20140804	5570000	5	5	9200	35069	2	3	13	6200	3000	2001	
9254	9208900037	20140919	6885000	6	7	9890	31374	2	3	13	8860	1030	2001	
3914	9808700762	20140611	7062500	5	4	10040	37325	2	3	11	7680	2360	1940	
7252	6762700020	20141013	7700000	6	8	12050	27600	2	4	13	8570	3480	1910	

21613 rows × 19 columns

In [104]:

#last records
df.tail()

Out[104]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	condition	grade	sqft_above	sqft_basement	yr_built	yr_renova
1448	8907500070	20150413	5350000	5	5	8000	23985	2	3	12	6720	1280	2009	
4411	2470100110	20140804	5570000	5	5	9200	35069	2	3	13	6200	3000	2001	
9254	9208900037	20140919	6885000	6	7	9890	31374	2	3	13	8860	1030	2001	
3914	9808700762	20140611	7062500	5	4	10040	37325	2	3	11	7680	2360	1940	2
7252	6762700020	20141013	7700000	6	8	12050	27600	2	4	13	8570	3480	1910	1
4														<b>•</b>

M

```
In [106]:
                                                                                                                                                                M
#descriptive info
df.describe()
Out[106]:
                  id
                             price
                                      bedrooms
                                                   bathrooms
                                                                 sqft_living
                                                                                 sqft_lot
                                                                                                 floors
                                                                                                           condition
                                                                                                                                     sqft_above sqf
                                                                                                                            grade
 count 2.161300e+04
                     2.161300e+04
                                   21613.000000
                                                21613.000000
                                                              21613.000000 2.161300e+04
                                                                                         21613.000000
                                                                                                       21613.000000
                                                                                                                     21613.000000
                                                                                                                                   21613.000000
                                                                                                                                                  2
 mean 4.580302e+09 5.400881e+05
                                       3 370842
                                                     1.749734
                                                               2079.899736
                                                                            1.510697e+04
                                                                                              1 446213
                                                                                                            3 409430
                                                                                                                         7.656873
                                                                                                                                    1788 390691
       2.876566e+09 3.671272e+05
                                       0.930062
                                                     0.734873
                                                                918.440897
                                                                            4.142051e+04
                                                                                              0.551894
                                                                                                            0.650743
                                                                                                                         1.175459
                                                                                                                                     828.090978
   std
   min
       1.000102e+06 7.500000e+04
                                       0.000000
                                                     0.000000
                                                                290.000000 5.200000e+02
                                                                                              1.000000
                                                                                                            1.000000
                                                                                                                         1.000000
                                                                                                                                     290.000000
       2.123049e+09 3.219500e+05
                                       3.000000
                                                     1.000000
                                                                1427.000000 5.040000e+03
                                                                                              1.000000
                                                                                                            3.000000
                                                                                                                         7.000000
                                                                                                                                    1190.000000
  25%
        3.904930e+09
                                       3.000000
                                                     2.000000
                                                                1910.000000
                                                                            7.618000e+03
                                                                                              1.000000
                                                                                                            3.000000
                                                                                                                         7.000000
                                                                                                                                    1560.000000
  50%
                     4.500000e+05
  75%
       7.308900e+09 6.450000e+05
                                       4.000000
                                                     2.000000
                                                               2550.000000 1.068800e+04
                                                                                              2.000000
                                                                                                            4.000000
                                                                                                                         8.000000
                                                                                                                                    2210.000000
       9.900000e+09 7.700000e+06
                                      33.000000
                                                     8.000000
                                                              13540.000000 1.651359e+06
                                                                                              3.000000
                                                                                                            5.000000
                                                                                                                         13.000000
                                                                                                                                    9410.000000
  max
In [112]:
                                                                                                                                                                M
#show the higest price of room
df["price"].max()
Out[112]:
7700000
In [116]:
                                                                                                                                                                M
#coverting grade into int
df['grade']=df["grade"].astype(int)
In [117]:
                                                                                                                                                                M
df
Out[117]:
                id
                        date
                                 price
                                       bedrooms
                                                  bathrooms
                                                             sqft_living
                                                                        sqft_lot floors
                                                                                        condition
                                                                                                  grade
                                                                                                         sqft_above
                                                                                                                    sqft_basement
                                                                                                                                   yr_built
                                                                                                                                           yr_renov
                                                                                                      3
        3421079032
                    20150217
                                75000
                                               1
                                                                          43377
                                                                                                                                 0
  1149
                                                          0
                                                                   670
                                                                                               3
                                                                                                                670
                                                                                                                                      1966
 15293
          40000362 20140506
                                78000
                                               2
                                                                   780
                                                                          16344
                                                                                                      5
                                                                                                                780
                                                                                                                                 0
                                                                                                                                      1942
   465
        8658300340
                   20140523
                                80000
                                               1
                                                          0
                                                                   430
                                                                           5050
                                                                                               2
                                                                                                      4
                                                                                                                430
                                                                                                                                 0
                                                                                                                                      1912
        3028200080 20150324
                                               2
                                81000
                                                                           9975
                                                                                                      5
                                                                                                                                 0
                                                                                                                                      1943
 16198
                                                                   730
                                                                                     1
                                                                                               1
                                                                                                                730
        3883800011
                                               3
                                82000
                                                                   860
                                                                          10426
                                                                                                                860
                                                                                                                                 0
                                                                                                                                      1954
        8907500070 20150413 5350000
                                               5
                                                          5
                                                                  8000
                                                                          23985
                                                                                     2
                                                                                               3
                                                                                                     12
                                                                                                               6720
                                                                                                                              1280
                                                                                                                                      2009
  1448
        2470100110 20140804
                              5570000
                                               5
                                                          5
                                                                  9200
                                                                          35069
                                                                                     2
                                                                                               3
                                                                                                     13
                                                                                                               6200
                                                                                                                              3000
                                                                                                                                      2001
  9254
        9208900037 20140919
                              6885000
                                               6
                                                          7
                                                                  9890
                                                                          31374
                                                                                     2
                                                                                               3
                                                                                                     13
                                                                                                               8860
                                                                                                                              1030
                                                                                                                                      2001
        9808700762 20140611
                                               5
                                                          4
                                                                  10040
                                                                          37325
                                                                                     2
                                                                                               3
                                                                                                     11
                                                                                                               7680
                                                                                                                              2360
                                                                                                                                      1940
  3914
                              7062500
  7252 6762700020 20141013 7700000
                                                                  12050
                                                                          27600
                                                                                     2
                                                                                                     13
                                                                                                               8570
                                                                                                                              3480
                                                                                                                                      1910
21613 rows × 19 columns
4
In [118]:
                                                                                                                                                                M
import matplotlib.pyplot as plt
                                                                                                                                                                M
In [121]:
x=df['yr_built']
```

```
In [122]:
                                                                                                                                                  M
х
Out[122]:
1149
         1966
15293
         1942
465
         1912
16198
         1943
         1954
8274
          2009
1448
         2001
4411
9254
         2001
3914
         1940
         1910
7252
Name: yr_built, Length: 21613, dtype: int64
In [124]:
                                                                                                                                                  M
#data of year of construnction:
x.plot(kind='hist')
Out[124]:
<AxesSubplot:ylabel='Frequency'>
    3500
    3000
    2500
 Frequency 1500
    1000
     500
        0
                                 1940
                                            1960
                                                       1980
           1900
                      1920
                                                                  2000
                                                                             2020
                                                                                                                                                  M
In [125]:
df.head()
Out[125]:
```

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovat
1149	3421079032	20150217	75000	1	0	670	43377	1	3	3	670	0	1966	
15293	40000362	20140506	78000	2	1	780	16344	1	1	5	780	0	1942	
465	8658300340	20140523	80000	1	0	430	5050	1	2	4	430	0	1912	
16198	3028200080	20150324	81000	2	1	730	9975	1	1	5	730	0	1943	
8274	3883800011	20141105	82000	3	1	860	10426	1	3	6	860	0	1954	
4														<b>•</b>

```
In [128]:
                                                                                                                                        M
g1=list(df.groupby(["price","sqft_lot"]))
```

```
In [132]:
                                                                                                                                                  M
g1
Out[132]:
[((75000, 43377),
                 id
                           date price bedrooms bathrooms sqft_living
  1149 3421079032 20150217
                                  75000
         {\it sqft\_lot floors condition grade sqft\_above sqft\_basement yr\_built \ \setminus \\
  1149
           43377
        yr_renovated zipcode
                                   lat
                                              long sqft_living15 sqft_lot15
  1149
                          98022 47.2638 -121.906
                                                                          42882 ),
 ((78000, 16344),
                           date price bedrooms bathrooms sqft_living \
  15293 40000362 20140506
                                 78000
         \verb|sqft_lot| floors condition grade sqft_above sqft_basement \\ \\ \\ \\ \\
  15293
         yr_built yr_renovated zipcode
                                                 lat
                                                         long sqft_living15 \
                                                                                                                                                  M
In [135]:
pie1=df["bedrooms"]
In [136]:
                                                                                                                                                  M
pie1
Out[136]:
1149
         1
15293
         2
465
         1
16198
8274
1448
4411
9254
         6
3914
         5
7252
Name: bedrooms, Length: 21613, dtype: int32
In [142]:
                                                                                                                                                  Ы
df
Out[142]:
               id
                      date
                              price bedrooms bathrooms sqft_living sqft_lot floors condition grade sqft_above sqft_basement yr_built yr_renov
 1149 3421079032 20150217
                             75000
                                                     0
                                                             670
                                                                   43377
                                                                                             3
                                                                                                      670
                                                                                                                     0
                                                                                                                          1966
 15293
         40000362 20140506
                             78000
                                           2
                                                     1
                                                             780
                                                                   16344
                                                                                             5
                                                                                                      780
                                                                                                                     0
                                                                                                                          1942
  465 8658300340 20140523
                             80000
                                                     0
                                                                                       2
                                                                                             4
                                                                                                                     0
                                                                                                                          1912
                                                             430
                                                                    5050
                                                                                                      430
 16198 3028200080 20150324
                             81000
                                                             730
                                                                    9975
                                                                             1
                                                                                                      730
                                                                                                                     0
                                                                                                                          1943
 8274 3883800011 20141105
                             82000
                                           3
                                                             860
                                                                   10426
                                                                                      3
                                                                                             6
                                                                                                      860
                                                                                                                     0
                                                                                                                          1954
 1448
       8907500070 20150413 5350000
                                                     5
                                                            8000
                                                                   23985
                                                                                      3
                                                                                            12
                                                                                                     6720
                                                                                                                  1280
                                                                                                                          2009
 4411 2470100110 20140804 5570000
                                           5
                                                     5
                                                            9200
                                                                   35069
                                                                             2
                                                                                      3
                                                                                            13
                                                                                                     6200
                                                                                                                  3000
                                                                                                                          2001
                                                                                            13
                                                                                                     8860
                                                                                                                  1030
 9254 9208900037 20140919 6885000
                                           6
                                                            9890
                                                                   31374
                                                                                                                          2001
 3914 9808700762 20140611 7062500
                                           5
                                                     4
                                                            10040
                                                                   37325
                                                                             2
                                                                                            11
                                                                                                     7680
                                                                                                                  2360
                                                                                                                          1940
 7252 6762700020 20141013 7700000
                                           6
                                                            12050
                                                                   27600
                                                                             2
                                                                                            13
                                                                                                     8570
                                                                                                                  3480
                                                                                                                          1910
                                                     8
21613 rows × 19 columns
4
In [ ]:
                                                                                                                                                  M
                                                                                                                                                  M
In [ ]:
```