Panda's assignment

December 14, 2017

```
In [1]: ### Name:Thinley Jurmi ID:u6018302
```

```
In [ ]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt #importing pandas
        df=pd.read_csv("/Users/thinleyjurmi/Downloads/untitled folder/CArealestate.csv")
        #csv used
```

```
In [2]: df.head(20)#first 20 data of the csv file
```

```
Out[2]:                          street             city    zip state  beds  baths  \
        0                  3526 HIGH ST       SACRAMENTO  95838    CA     2      1
        1                   51 OMAHA CT       SACRAMENTO  95823    CA     3      1
        2                2796 BRANCH ST       SACRAMENTO  95815    CA     2      1
        3              2805 JANETTE WAY       SACRAMENTO  95815    CA     2      1
        4               6001 MCMAHON DR       SACRAMENTO  95824    CA     2      1
        5            5828 PEPPERMILL CT       SACRAMENTO  95841    CA     3      1
        6           6048 OGDEN NASH WAY       SACRAMENTO  95842    CA     3      2
        7                 2561 19TH AVE       SACRAMENTO  95820    CA     3      1
        8   11150 TRINITY RIVER DR Unit 114  RANCHO CORDOVA  95670    CA     2      2
        9                  7325 10TH ST        RIO LINDA  95673    CA     3      2
        10             645 MORRISON AVE       SACRAMENTO  95838    CA     3      2
        11                4085 FAWN CIR       SACRAMENTO  95823    CA     3      2
        12              2930 LA ROSA RD       SACRAMENTO  95815    CA     1      1
        13                2113 KIRK WAY       SACRAMENTO  95822    CA     3      1
        14           4533 LOCH HAVEN WAY       SACRAMENTO  95842    CA     2      2
        15               7340 HAMDEN PL       SACRAMENTO  95842    CA     2      2
        16                  6715 6TH ST        RIO LINDA  95673    CA     2      1
        17        6236 LONGFORD DR Unit 1   CITRUS HEIGHTS  95621    CA     2      1
        18              250 PERALTA AVE       SACRAMENTO  95833    CA     2      1
        19              113 LEEWILL AVE        RIO LINDA  95673    CA     3      2

            sq__ft         type                     sale_date   price   latitude  \
        0      836  Residential  Wed May 21 00:00:00 EDT 2008   59222  38.631913
        1     1167  Residential  Wed May 21 00:00:00 EDT 2008   68212  38.478902
        2      796  Residential  Wed May 21 00:00:00 EDT 2008   68880  38.618305
        3      852  Residential  Wed May 21 00:00:00 EDT 2008   69307  38.616835
        4      797  Residential  Wed May 21 00:00:00 EDT 2008   81900  38.519470
```

```
5     1122        Condo  Wed May 21 00:00:00 EDT 2008   89921  38.662595
6     1104  Residential  Wed May 21 00:00:00 EDT 2008   90895  38.681659
7     1177  Residential  Wed May 21 00:00:00 EDT 2008   91002  38.535092
8      941        Condo  Wed May 21 00:00:00 EDT 2008   94905  38.621188
9     1146  Residential  Wed May 21 00:00:00 EDT 2008   98937  38.700909
10     909  Residential  Wed May 21 00:00:00 EDT 2008  100309  38.637663
11    1289  Residential  Wed May 21 00:00:00 EDT 2008  106250  38.470746
12     871  Residential  Wed May 21 00:00:00 EDT 2008  106852  38.618698
13    1020  Residential  Wed May 21 00:00:00 EDT 2008  107502  38.482215
14    1022  Residential  Wed May 21 00:00:00 EDT 2008  108750  38.672914
15    1134        Condo  Wed May 21 00:00:00 EDT 2008  110700  38.700051
16     844  Residential  Wed May 21 00:00:00 EDT 2008  113263  38.689591
17     795        Condo  Wed May 21 00:00:00 EDT 2008  116250  38.679776
18     588  Residential  Wed May 21 00:00:00 EDT 2008  120000  38.612099
19    1356  Residential  Wed May 21 00:00:00 EDT 2008  121630  38.689999

     longitude
0  -121.434879
1  -121.431028
2  -121.443839
3  -121.439146
4  -121.435768
5  -121.327813
6  -121.351705
7  -121.481367
8  -121.270555
9  -121.442979
10 -121.451520
11 -121.458918
12 -121.435833
13 -121.492603
14 -121.359340
15 -121.351278
16 -121.452239
17 -121.314089
18 -121.469095
19 -121.463220
```

In [3]:
```
df.describe()
```

Out[3]:
```
                zip        beds       baths        sq__ft          price  \
count    499.000000  499.000000  499.000000    499.000000     499.000000
mean   95758.643287    3.062124    1.871743   1410.086172  232728.452906
std       84.660893    1.145161    0.838089    775.033290  121169.372390
min    95603.000000    0.000000    0.000000      0.000000   40000.000000
25%    95670.000000    3.000000    1.000000   1019.000000  145500.000000
50%    95815.000000    3.000000    2.000000   1326.000000  215100.000000
75%    95828.000000    4.000000    2.000000   1776.000000  285000.000000
```

```
max     95864.000000    8.000000    5.000000  3992.000000  839000.000000

         latitude   longitude
count  499.000000  499.000000
mean    38.580031 -121.369985
std      0.129570    0.132910
min     38.241514 -121.551704
25%     38.475161 -121.450145
50%     38.602416 -121.401482
75%     38.676239 -121.315538
max     39.008159 -120.603872
```

In [6]:
```
df.groupby(["city", "street"]).size()#street grouped to a city accordingly
```

Out[6]:
```
city            street
ANTELOPE        3228 BAGGAN CT             1
                3305 RIO ROCA CT           1
                3318 DAVIDSON DR           1
                3828 BLACKFOOT WAY         1
                4437 MITCHUM CT            1
                4508 OLD DAIRY DR          1
                4844 CLYDEBANK WAY         1
                5308 MARBURY WAY           1
                5312 MARBURY WAY           1
                5708 RIDGEPOINT DR         1
                5712 MELBURY CIR           1
                7837 ABBINGTON WAY         1
                7863 CRESTLEIGH CT         1
                7895 CABER WAY             1
                8020 WALERGA RD            1
                8108 FILIFERA WAY          1
                8721 SPRUCE RIDGE WAY      1
AUBURN          1740 HIGH ST               1
                220 OLD AIRPORT RD         1
                2231 COUNTRY VILLA CT      1
CAMERON PARK    2778 KAWEAH CT             1
                3361 BOW MAR CT            1
                3429 FERNBROOK CT          1
CARMICHAEL      2109 HAMLET PL             1
                4010 ALEX LN               1
                4622 MEYER WAY             1
                5332 SANDSTONE ST          1
                5709 RIVER OAK WAY         1
                5847 DEL CAMPO LN          1
                5907 ELLERSLEE DR          1
                                          ..
SACRAMENTO      8219 GWINHURST CIR         1
                8244 SUNBIRD WAY           1
```

```
                 8299 HALBRITE WAY          1
                 8317 SUNNY CREEK WAY       1
                 8344 FIELDPOPPY CIR        1
                 8345 STAR THISTLE WAY      1
                 8354 SUNRISE WOODS WAY     1
                 8363 LANGTREE WAY          1
                 8421 SUNBLAZE WAY          1
                 8495 DARTFORD DR           1
                 8515 DARTFORD DR           1
                 8593 DERLIN WAY            1
                 8612 WILLOW GROVE WAY      1
                 8728 CRYSTAL RIVER WAY     1
                 8788 LA MARGARITA WAY      1
                 8929 SUTTERS GOLD DR       1
                 9 PASTURE CT               1
                 9012 KIEFER BLVD           1
                 9013 CASALS ST             1
                 9058 MONTOYA ST            1
                 9080 BEDROCK CT            1
                 909 SINGINGWOOD RD         1
                 9186 KINBRACE CT           1
                 923 FULTON AVE             1
                 9257 CALDERA WAY           1
                 9273 PREMIER WAY           1
                 9570 HARVEST ROSE WAY      1
                 9861 CULP WAY              1
WILTON           11215 SHARRMONT CT         1
                 9741 SADDLEBRED CT         1
Length: 496, dtype: int64
```

In [7]:
```python
pd.crosstab(df.city,df.street)
```

Out[7]:
```
street           1 KENNELFORD CIR  10 SEA FOAM CT  100 REBECCA WAY  \
city
ANTELOPE                        0               0                0
AUBURN                          0               0                0
CAMERON PARK                    0               0                0
CARMICHAEL                      0               0                0
CITRUS HEIGHTS                  0               0                0
EL DORADO                       0               0                0
EL DORADO HILLS                 0               0                0
ELK GROVE                       0               0                0
ELVERTA                         0               0                0
FAIR OAKS                       0               0                0
FOLSOM                          0               0                1
GALT                            0               0                0
GOLD RIVER                      0               0                0
GREENWOOD                       0               0                0
```

LINCOLN	0	0	0
LOOMIS	0	0	0
MATHER	0	0	0
MEADOW VISTA	0	0	0
NORTH HIGHLANDS	0	0	0
ORANGEVALE	0	0	0
PLACERVILLE	0	0	0
POLLOCK PINES	0	0	0
RANCHO CORDOVA	0	0	0
RANCHO MURIETA	0	0	0
RIO LINDA	0	0	0
ROCKLIN	0	0	0
ROSEVILLE	0	0	0
SACRAMENTO	1	1	0
WILTON	0	0	0

street	100 TOURMALINE CIR	10001 WOODCREEK OAKS BLVD Unit 1415 \
city		
ANTELOPE	0	0
AUBURN	0	0
CAMERON PARK	0	0
CARMICHAEL	0	0
CITRUS HEIGHTS	0	0
EL DORADO	0	0
EL DORADO HILLS	0	0
ELK GROVE	0	0
ELVERTA	0	0
FAIR OAKS	0	0
FOLSOM	0	0
GALT	0	0
GOLD RIVER	0	0
GREENWOOD	0	0
LINCOLN	0	0
LOOMIS	0	0
MATHER	0	0
MEADOW VISTA	0	0
NORTH HIGHLANDS	0	0
ORANGEVALE	0	0
PLACERVILLE	0	0
POLLOCK PINES	0	0
RANCHO CORDOVA	0	0
RANCHO MURIETA	0	0
RIO LINDA	0	0
ROCKLIN	0	0
ROSEVILLE	0	1
SACRAMENTO	1	0
WILTON	0	0

street	10062 LINCOLN VILLAGE DR	10104 ANNIE ST	10158 CRAWFORD WAY \
city			
ANTELOPE	0	0	0
AUBURN	0	0	0
CAMERON PARK	0	0	0
CARMICHAEL	0	0	0
CITRUS HEIGHTS	0	0	0
EL DORADO	0	0	0
EL DORADO HILLS	0	0	0
ELK GROVE	0	1	0
ELVERTA	0	0	0
FAIR OAKS	0	0	0
FOLSOM	0	0	0
GALT	0	0	0
GOLD RIVER	0	0	0
GREENWOOD	0	0	0
LINCOLN	0	0	0
LOOMIS	0	0	0
MATHER	0	0	0
MEADOW VISTA	0	0	0
NORTH HIGHLANDS	0	0	0
ORANGEVALE	0	0	0
PLACERVILLE	0	0	0
POLLOCK PINES	0	0	0
RANCHO CORDOVA	0	0	0
RANCHO MURIETA	0	0	0
RIO LINDA	0	0	0
ROCKLIN	0	0	0
ROSEVILLE	0	0	0
SACRAMENTO	1	0	1
WILTON	0	0	0

street	1016 CONGRESS AVE	10245 LOS PALOS DR \
city		
ANTELOPE	0	0
AUBURN	0	0
CAMERON PARK	0	0
CARMICHAEL	0	0
CITRUS HEIGHTS	0	0
EL DORADO	0	0
EL DORADO HILLS	0	0
ELK GROVE	0	0
ELVERTA	0	0
FAIR OAKS	0	0
FOLSOM	0	0
GALT	0	0
GOLD RIVER	0	0
GREENWOOD	0	0

```
LINCOLN                           0                   0
LOOMIS                            0                   0
MATHER                            0                   0
MEADOW VISTA                      0                   0
NORTH HIGHLANDS                   0                   0
ORANGEVALE                        0                   0
PLACERVILLE                       0                   0
POLLOCK PINES                     0                   0
RANCHO CORDOVA                    0                   1
RANCHO MURIETA                    0                   0
RIO LINDA                         0                   0
ROCKLIN                           0                   0
ROSEVILLE                         0                   0
SACRAMENTO                        1                   0
WILTON                            0                   0

street                     ...           9629 CEDAR OAK WAY  \
city                       ...
ANTELOPE                   ...                            0
AUBURN                     ...                            0
CAMERON PARK               ...                            0
CARMICHAEL                 ...                            0
CITRUS HEIGHTS             ...                            0
EL DORADO                  ...                            0
EL DORADO HILLS            ...                            0
ELK GROVE                  ...                            1
ELVERTA                    ...                            0
FAIR OAKS                  ...                            0
FOLSOM                     ...                            0
GALT                       ...                            0
GOLD RIVER                 ...                            0
GREENWOOD                  ...                            0
LINCOLN                    ...                            0
LOOMIS                     ...                            0
MATHER                     ...                            0
MEADOW VISTA               ...                            0
NORTH HIGHLANDS            ...                            0
ORANGEVALE                 ...                            0
PLACERVILLE                ...                            0
POLLOCK PINES              ...                            0
RANCHO CORDOVA             ...                            0
RANCHO MURIETA             ...                            0
RIO LINDA                  ...                            0
ROCKLIN                    ...                            0
ROSEVILLE                  ...                            0
SACRAMENTO                 ...                            0
WILTON                     ...                            0
```

street	9688 NATURE TRAIL WAY	9720 LITTLE HARBOR WAY	\
city			
ANTELOPE	0	0	
AUBURN	0	0	
CAMERON PARK	0	0	
CARMICHAEL	0	0	
CITRUS HEIGHTS	0	0	
EL DORADO	0	0	
EL DORADO HILLS	0	0	
ELK GROVE	1	1	
ELVERTA	0	0	
FAIR OAKS	0	0	
FOLSOM	0	0	
GALT	0	0	
GOLD RIVER	0	0	
GREENWOOD	0	0	
LINCOLN	0	0	
LOOMIS	0	0	
MATHER	0	0	
MEADOW VISTA	0	0	
NORTH HIGHLANDS	0	0	
ORANGEVALE	0	0	
PLACERVILLE	0	0	
POLLOCK PINES	0	0	
RANCHO CORDOVA	0	0	
RANCHO MURIETA	0	0	
RIO LINDA	0	0	
ROCKLIN	0	0	
ROSEVILLE	0	0	
SACRAMENTO	0	0	
WILTON	0	0	

street	9723 TERRAPIN CT	9741 SADDLEBRED CT	9760 LAZULITE CT	\
city				
ANTELOPE	0	0	0	
AUBURN	0	0	0	
CAMERON PARK	0	0	0	
CARMICHAEL	0	0	0	
CITRUS HEIGHTS	0	0	0	
EL DORADO	0	0	0	
EL DORADO HILLS	0	0	0	
ELK GROVE	1	0	1	
ELVERTA	0	0	0	
FAIR OAKS	0	0	0	
FOLSOM	0	0	0	
GALT	0	0	0	
GOLD RIVER	0	0	0	
GREENWOOD	0	0	0	

```
LINCOLN                            0                   0                  0
LOOMIS                             0                   0                  0
MATHER                             0                   0                  0
MEADOW VISTA                       0                   0                  0
NORTH HIGHLANDS                    0                   0                  0
ORANGEVALE                         0                   0                  0
PLACERVILLE                        0                   0                  0
POLLOCK PINES                      0                   0                  0
RANCHO CORDOVA                     0                   0                  0
RANCHO MURIETA                     0                   0                  0
RIO LINDA                          0                   0                  0
ROCKLIN                            0                   0                  0
ROSEVILLE                          0                   0                  0
SACRAMENTO                         0                   0                  0
WILTON                             0                   1                  0

street           9837 CORTE DORADO CT  9861 CULP WAY  993 MANTON CT  \
city
ANTELOPE                            0              0              0
AUBURN                              0              0              0
CAMERON PARK                        0              0              0
CARMICHAEL                          0              0              0
CITRUS HEIGHTS                      0              0              0
EL DORADO                           0              0              0
EL DORADO HILLS                     0              0              0
ELK GROVE                           1              0              0
ELVERTA                             0              0              0
FAIR OAKS                           0              0              0
FOLSOM                              0              0              0
GALT                                0              0              1
GOLD RIVER                          0              0              0
GREENWOOD                           0              0              0
LINCOLN                             0              0              0
LOOMIS                              0              0              0
MATHER                              0              0              0
MEADOW VISTA                        0              0              0
NORTH HIGHLANDS                     0              0              0
ORANGEVALE                          0              0              0
PLACERVILLE                         0              0              0
POLLOCK PINES                       0              0              0
RANCHO CORDOVA                      0              0              0
RANCHO MURIETA                      0              0              0
RIO LINDA                           0              0              0
ROCKLIN                             0              0              0
ROSEVILLE                           0              0              0
SACRAMENTO                          0              1              0
WILTON                              0              0              0
```

```
street          9970 STATE HIGHWAY 193
city
ANTELOPE                             0
AUBURN                               0
CAMERON PARK                         0
CARMICHAEL                           0
CITRUS HEIGHTS                       0
EL DORADO                            0
EL DORADO HILLS                      0
ELK GROVE                            0
ELVERTA                              0
FAIR OAKS                            0
FOLSOM                               0
GALT                                 0
GOLD RIVER                           0
GREENWOOD                            0
LINCOLN                              0
LOOMIS                               0
MATHER                               0
MEADOW VISTA                         0
NORTH HIGHLANDS                      0
ORANGEVALE                           0
PLACERVILLE                          1
POLLOCK PINES                        0
RANCHO CORDOVA                       0
RANCHO MURIETA                       0
RIO LINDA                            0
ROCKLIN                              0
ROSEVILLE                            0
SACRAMENTO                           0
WILTON                               0

[29 rows x 496 columns]
```

In [4]:
```
df["city"].value_counts()
```

Out[4]:
```
SACRAMENTO         248
ELK GROVE           66
CITRUS HEIGHTS      19
ANTELOPE            17
ROSEVILLE           17
RANCHO CORDOVA      15
GALT                14
NORTH HIGHLANDS     13
FOLSOM              10
CARMICHAEL          10
RIO LINDA           10
EL DORADO HILLS      8
```

```
LINCOLN              7
PLACERVILLE          5
ROCKLIN              5
ORANGEVALE           5
FAIR OAKS            5
ELVERTA              4
GOLD RIVER           3
CAMERON PARK         3
AUBURN               3
RANCHO MURIETA       3
POLLOCK PINES        2
WILTON               2
GREENWOOD            1
LOOMIS               1
MATHER               1
MEADOW VISTA         1
EL DORADO            1
Name: city, dtype: int64
```

In [66]:
```
df["type"].value_counts()
```

Out[66]:
```
Residential     453
Condo            37
Multi-Family      9
Name: type, dtype: int64
```

In [8]:
```
df.loc[(df["city"]=="SACRAMENTO") &(df["type"]=="Condo")& (df['price']<=150000),['city
```

Out[8]:

	city	type	price
5	SACRAMENTO	Condo	89921
15	SACRAMENTO	Condo	110700
39	SACRAMENTO	Condo	150000
335	SACRAMENTO	Condo	40000
336	SACRAMENTO	Condo	48000
343	SACRAMENTO	Condo	77000
344	SACRAMENTO	Condo	82732
349	SACRAMENTO	Condo	90000
354	SACRAMENTO	Condo	100000
368	SACRAMENTO	Condo	120000
377	SACRAMENTO	Condo	126960

In [9]:
```
#Creating a new function:
def num_missing(x):
      return sum(x.isnull())

#Applying per column:
print("Missing values per column:")
print (df.apply(num_missing, axis=0)) #axis=0 defines that function is to be applied o
```

```
        print("\nMissing values per row:")
        print(df.apply(num_missing, axis=1).head()) #axis=1 defines that function is to be app
```

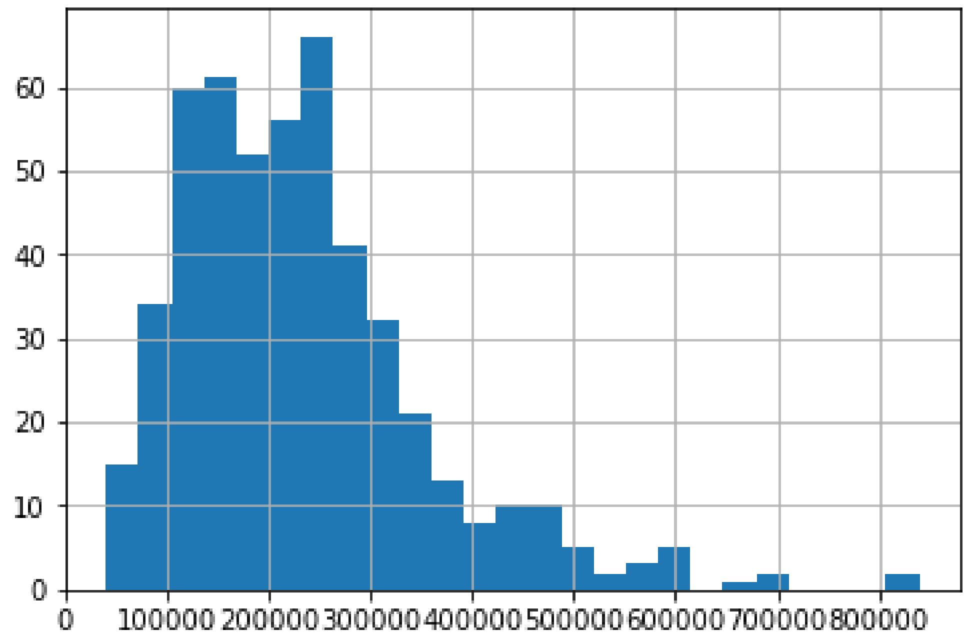
```
Missing values per column:
street       0
city         0
zip          0
state        0
beds         0
baths        0
sq__ft       0
type         0
sale_date    0
price        0
latitude     0
longitude    0
dtype: int64

Missing values per row:
0    0
1    0
2    0
3    0
4    0
dtype: int64
```

```
In [10]:  df.apply(lambda x: sum(x.isnull()),axis=0) #using lambda
```

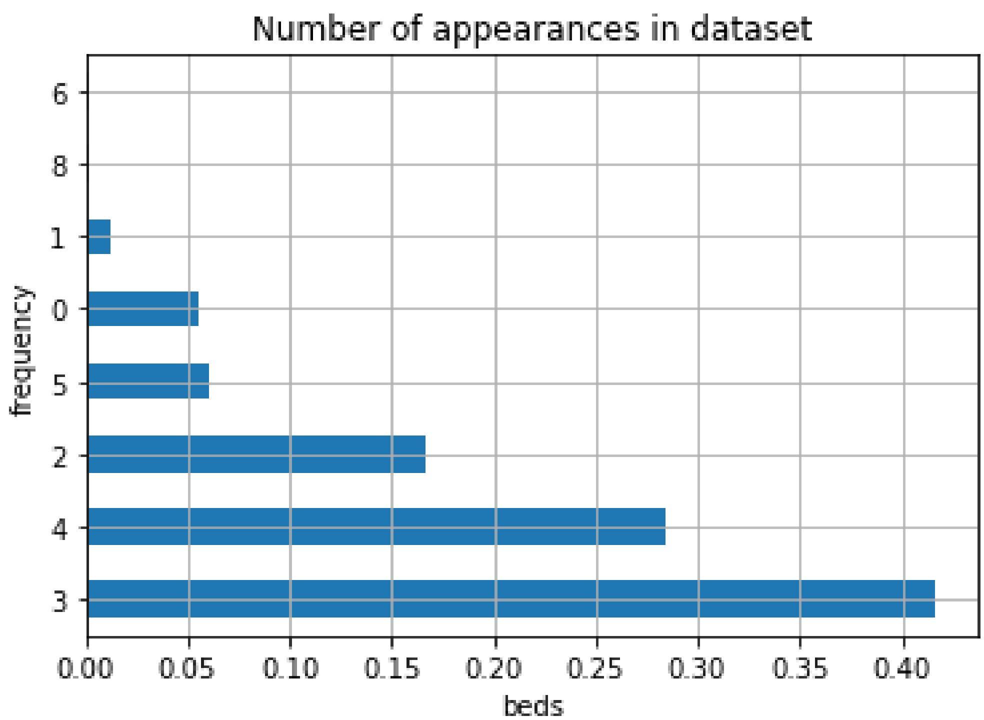
```
Out[10]: street       0
         city         0
         zip          0
         state        0
         beds         0
         baths        0
         sq__ft       0
         type         0
         sale_date    0
         price        0
         latitude     0
         longitude    0
         dtype: int64
```

```
In [11]: %matplotlib inline
         import matplotlib.pyplot as plt
         df["price"].hist(bins=25)#histogram of the first 25 data of the price
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x118c5ba20>
```

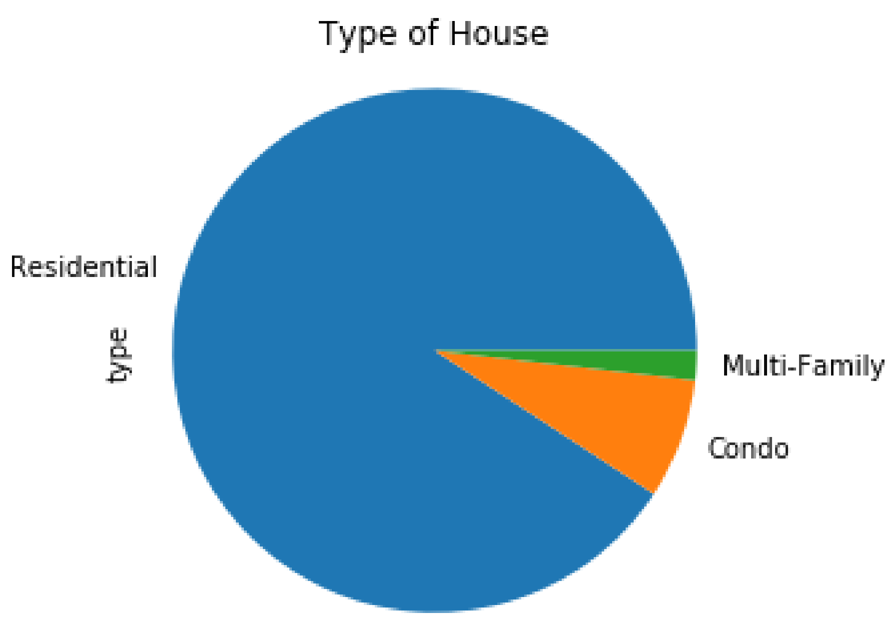



```
In [42]: df.beds.value_counts(25).plot(kind='barh',grid=True)
         plt.title('Number of appearances in dataset')
         plt.xlabel('beds')
         plt.ylabel('frequency') #histogram of city.Incorrect graph.

Out[42]: <matplotlib.text.Text at 0x11dc67a90>
```



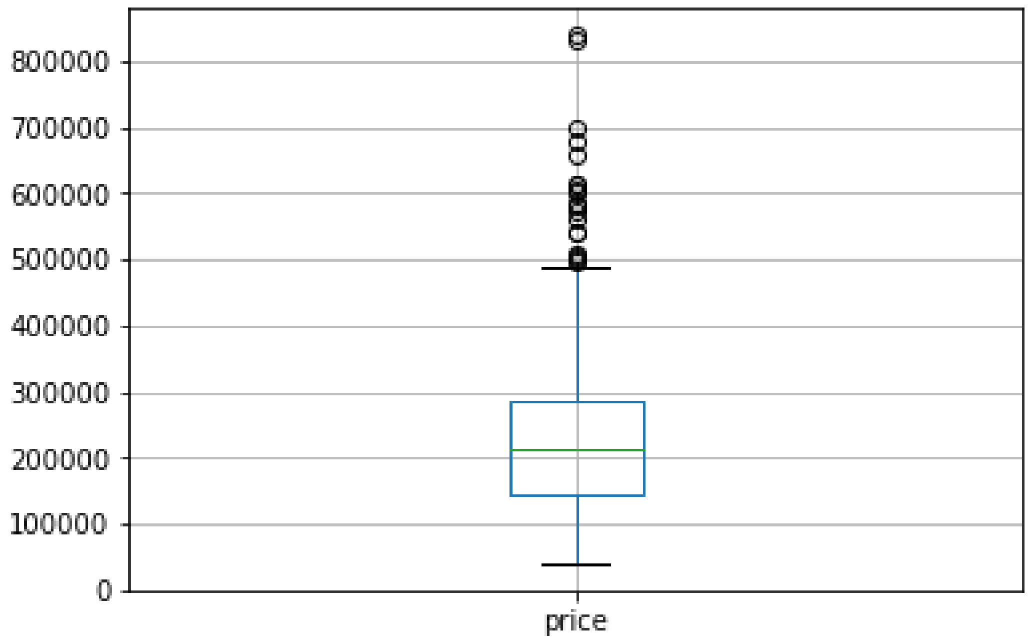

```
In [16]: df.type.value_counts().plot(kind='pie')
         plt.axis('scaled')
         plt.title('Type of House')#pie chart version of the type of house
```

```
Out[16]: <matplotlib.text.Text at 0x11c5984e0>
```



```
In [18]: df.boxplot(column='price')
```

```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x11c699b38>
```



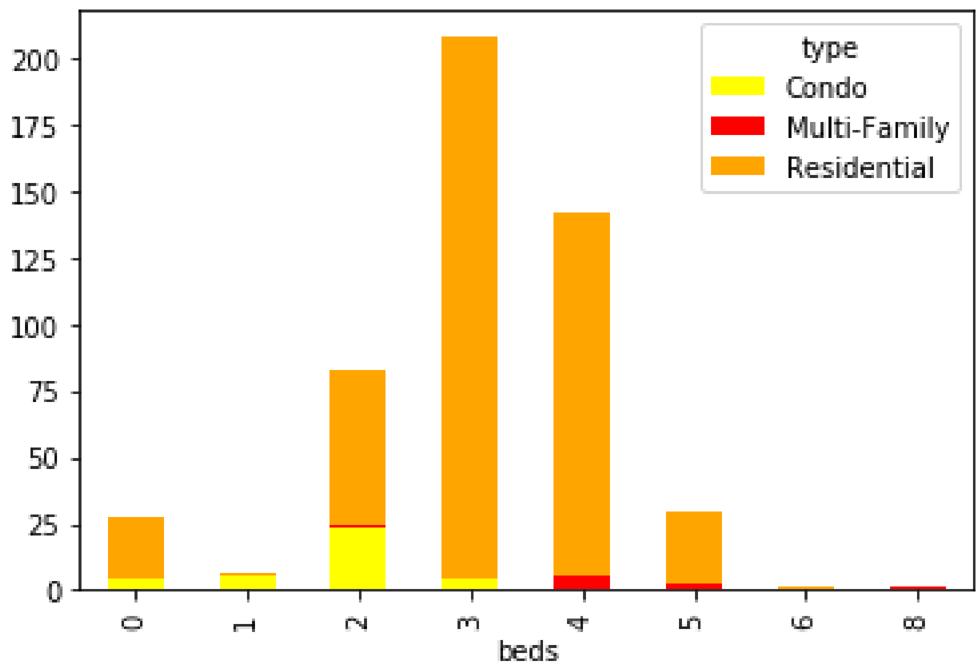

```
In [33]: #pivot table
         temp1 = df['beds'].value_counts(ascending=True)
         print('Frequency Table for Bedrooms:' )
         print(temp1)
         temp2 = df.pivot_table(values='type',index=['beds'],aggfunc=lambda x: x.map({'Residen
         print(temp2) #not sure about temp2 did the same as intructed on the website.
```

```
Frequency Table for Bedrooms:
6      1
8      1
1      6
0     28
5     30
2     83
4    142
3    208
Name: beds, dtype: int64
          type
beds
```

```
0       0.857143
1       0.166667
2       0.707317
3       0.980769
4       1.000000
5       1.000000
6       1.000000
8            NaN
```

```
In [80]: temp3=pd.crosstab(df['beds'],df['type'])
         temp3.plot(kind='bar',stacked=True, color=['Yellow','red','orange'],grid=False)#stack
```

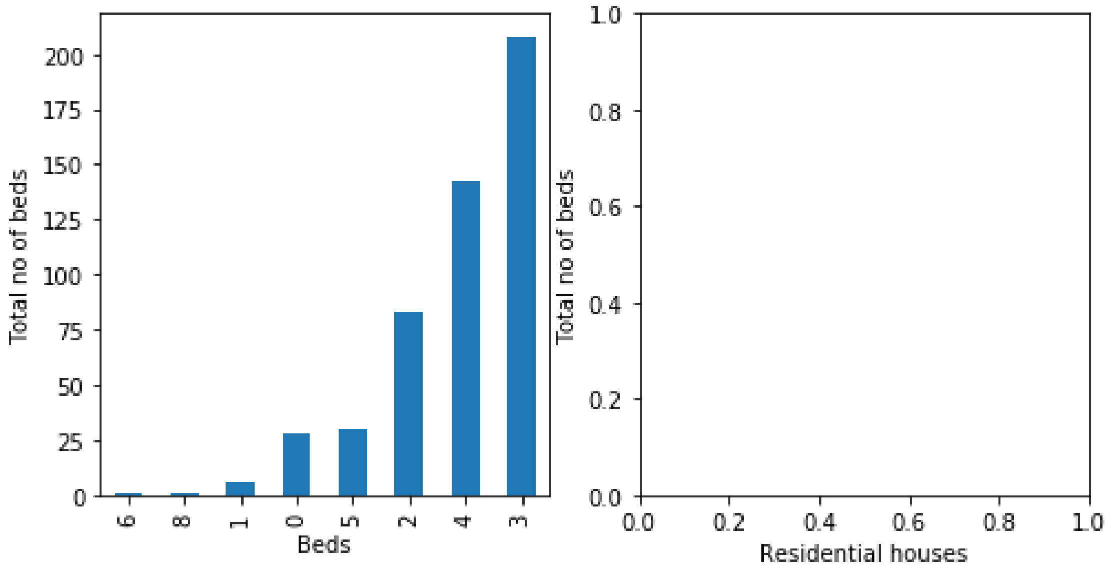
```
Out[80]: <matplotlib.axes._subplots.AxesSubplot at 0x11dff0b00>
```

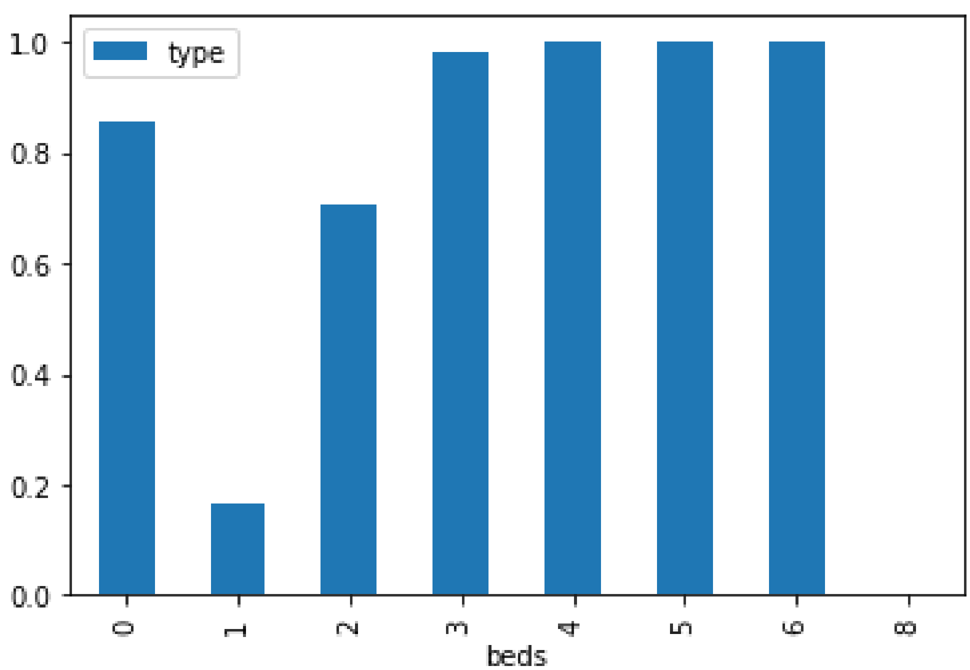



```
In [47]: import matplotlib.pyplot as plt
         fig = plt.figure(figsize=(8,4))
         ax1 = fig.add_subplot(121)
         ax1.set_xlabel('Beds')
         ax1.set_ylabel('Total no of beds')
         ax1.set_title("")
         temp1.plot(kind='bar')

         ax2 = fig.add_subplot(122)
         temp2.plot(kind = 'bar')
```

```
ax2.set_xlabel('Residential houses')
ax2.set_ylabel('Total no of beds')
ax2.set_title("")
```

Out[47]: <matplotlib.text.Text at 0x11ea254e0>





```
In [52]: pd.crosstab(df["beds"],df["type"],margins=False) #crosstab
```

```
Out[52]: type  Condo  Multi-Family  Residential
         beds
         0         4             0           24
         1         5             0            1
         2        24             1           58
         3         4             0          204
         4         0             5          137
         5         0             2           28
         6         0             0            1
         8         0             1            0
```

```
In [53]: pd.crosstab(df["beds"],df["type"],margins=True) #crosstab
```

```
Out[53]: type  Condo  Multi-Family  Residential  All
         beds
         0         4             0           24   28
         1         5             0            1    6
         2        24             1           58   83
         3         4             0          204  208
         4         0             5          137  142
         5         0             2           28   30
         6         0             0            1    1
         8         0             1            0    1
         All      37             9          453  499
```

```
In [58]: prop_rates = pd.DataFrame([1000, 5000, 12000], index=['Residential','COndo','Multi-Fa
         prop_rates
```

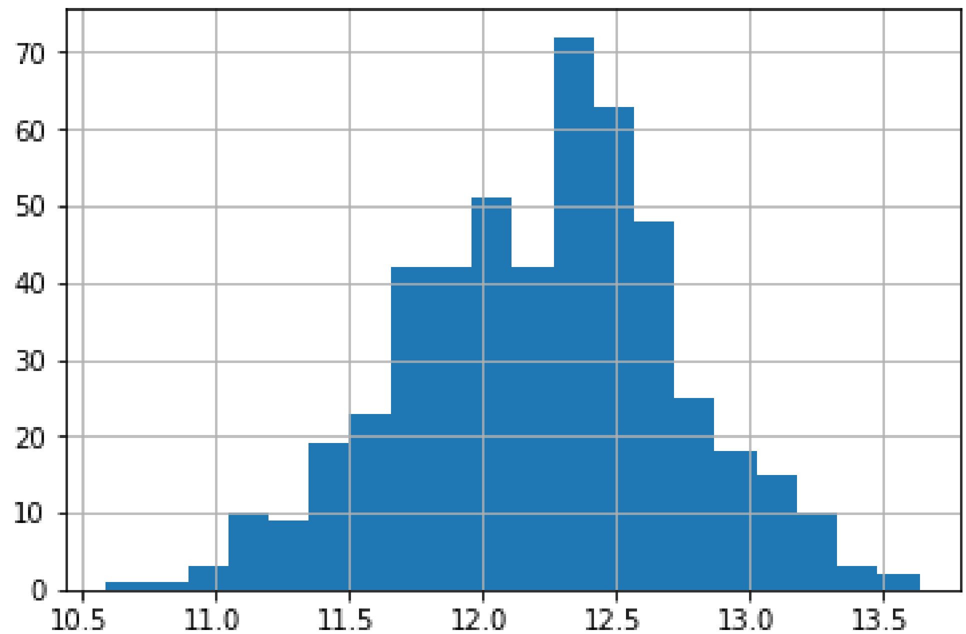
```
Out[58]:               rates
         Residential    1000
         COndo          5000
         Multi-Family  12000
```

```
In [62]: data_sorted = df.sort_values(['beds','baths'], ascending=False)
         data_sorted[['beds','baths']].head(10)
         #sorted the datas
```

```
Out[62]:      beds  baths
         108     8      4
         157     6      5
         332     5      5
         120     5      4
         151     5      4
         258     5      4
         303     5      4
         307     5      4
         451     5      4
         78      5      3
```

In [68]:
```
df['price_log'] = np.log(df['price'])
df['price_log'].hist(bins=20)#used log for extreme numbers
```

Out[68]: <matplotlib.axes._subplots.AxesSubplot at 0x11e5fff60>



In [78]:
```
df.dtypes
```

Out[78]:
```
street          object
city            object
zip              int64
state           object
beds             int64
baths            int64
sq__ft           int64
type            object
sale_date       object
price            int64
latitude       float64
longitude      float64
price_log      float64
dtype: object
```

In [89]:
```
import numpy as np

import matplotlib as plt
```

```python
from sklearn import datasets

from sklearn import metrics

from sklearn.linear_model import LogisticRegression #importing scikit
```

In [111]:
```python
from sklearn.preprocessing import LabelEncoder
var_mod = ['street','state','city','type','sale_date']
le = LabelEncoder()
for i in var_mod:
    df[i] = le.fit_transform(df[i])
df.dtypes
```

Out[111]:
```
street                 int64
city                   int64
zip                    int64
state                  int64
beds                   int64
baths                  int64
sq__ft                 int64
type                   int64
sale_date              int64
price                  int64
latitude             float64
longitude            float64
price_log            float64
Totalbed&baths         int64
TotalIncome            int64
TotalIncome_log      float64
dtype: object
```

In [123]:
```python
#i tried decision tree, logistic regression but it didn't work
```

In [ ]:

In [ ]: