eng22cs0130

October 8, 2024

```
[173]: import pandas as pd
       import matplotlib.pyplot as plt
       import seaborn as sns
       import numpy as np
[174]: data = pd.read_csv(r"C:\Users\DSU-CSE513-16\Downloads\Housing (2).csv")
[175]:
       data
                               bedrooms
[175]:
                                          bathrooms
                                                      stories mainroad guestroom basement
                price
                        area
       0
             13300000
                        7420
                                       4
                                                   2
                                                             3
                                                                     yes
                                                                                 no
                                                                                           no
                                       4
                                                   4
                                                             4
       1
             12250000
                        8960
                                                                     yes
                                                                                 no
                                                                                           no
       2
                                       3
                                                   2
                                                             2
             12250000
                        9960
                                                                     yes
                                                                                 no
                                                                                          yes
                                                   2
                                                             2
       3
             12215000
                        7500
                                       4
                                                                     yes
                                                                                 no
                                                                                          yes
                                                             2
       4
             11410000
                        7420
                                       4
                                                   1
                                                                     yes
                                                                                yes
                                                                                          yes
       561
              1820000
                        3000
                                       2
                                                   1
                                                             1
                                                                     yes
                                                                                 no
                                                                                          yes
       562
              1767150
                        2400
                                       3
                                                   1
                                                             1
                                                                      no
                                                                                 no
                                                                                           no
       563
              1750000
                        3620
                                       2
                                                   1
                                                             1
                                                                     yes
                                                                                 nο
                                                                                           no
                                       3
       564
              1750000
                                                   1
                                                             1
                        2910
                                                                      no
                                                                                 no
                                                                                           no
       565
                                       3
                                                             2
              1750000
                        3850
                                                   1
                                                                     yes
                                                                                 no
                                                                                           no
                                                parking prefarea furnishingstatus
            hotwaterheating airconditioning
       0
                                           yes
                                                                            furnished
                                                               yes
                                                        3
       1
                                                                            furnished
                          no
                                           yes
                                                                no
       2
                                                        2
                                                                      semi-furnished
                                                               yes
                          no
                                            no
       3
                                                        3
                                                                            furnished
                          no
                                           yes
                                                               yes
       4
                                                        2
                                                                            furnished
                          no
                                           yes
                                                                no
                                                        2
                                                                         unfurnished
       561
                          no
                                            no
                                                                no
       562
                                                        0
                                                                      semi-furnished
                          no
                                            no
                                                                no
       563
                          no
                                            no
                                                        0
                                                                no
                                                                         unfurnished
       564
                                                        0
                                                                            furnished
                          no
                                            no
                                                                no
       565
                                                       0
                                                                         unfurnished
                          nο
                                            nο
                                                                nο
```

[566 rows x 13 columns]

```
print(data.head(10))
        Sample Data:
              price
                       area
                              bedrooms
                                          bathrooms
                                                      stories mainroad guestroom basement
       0
          13300000
                       7420
                                                             3
                                                                      yes
                                                                                  no
                                                                                             no
                                                   4
       1
          12250000
                       8960
                                      4
                                                             4
                                                                     yes
                                                                                  nο
                                                                                            no
          12250000
                       9960
                                      3
                                                   2
                                                             2
       2
                                                                     yes
                                                                                           yes
                                                                                  nο
       3
          12215000
                       7500
                                      4
                                                   2
                                                             2
                                                                     yes
                                                                                  no
                                                                                            yes
       4
                       7420
                                      4
                                                   1
                                                             2
          11410000
                                                                     yes
                                                                                 yes
                                                                                            yes
                                      3
                                                   3
       5
                                                             1
          10850000
                       7500
                                                                     yes
                                                                                  no
                                                                                            yes
                                                   3
       6
          10150000
                       8580
                                      4
                                                             4
                                                                     yes
                                                                                  no
                                                                                            no
                                      5
                                                   3
                                                             2
       7
          10150000
                      16200
                                                                     yes
                                                                                  no
                                                                                            no
                                                             2
       8
           9870000
                       8100
                                      4
                                                   1
                                                                     yes
                                                                                 yes
                                                                                           yes
       9
           9800000
                       5750
                                      3
                                                   2
                                                             4
                                                                      yes
                                                                                 yes
                                                                                            no
         hotwaterheating airconditioning
                                               parking prefarea furnishingstatus
       0
                                                      2
                        no
                                          yes
                                                               yes
                                                                           furnished
       1
                                                      3
                                                                           furnished
                        no
                                          yes
                                                               no
                                                      2
       2
                                                               ves
                                                                      semi-furnished
                                          no
       3
                                                      3
                                                                           furnished
                        no
                                          yes
                                                               yes
       4
                                                      2
                                                                           furnished
                        nο
                                          yes
                                                               no
       5
                                                      2
                        nο
                                          yes
                                                               yes
                                                                      semi-furnished
       6
                                                      2
                                                                      semi-furnished
                                                               yes
                        no
                                          yes
       7
                                                      0
                                                                         unfurnished
                        no
                                           no
                                                               no
       8
                                                      2
                                         yes
                                                                           furnished
                        no
                                                               yes
       9
                                                      1
                        no
                                          yes
                                                               yes
                                                                         unfurnished
[177]: print(data.tail(10))
                                                      stories mainroad guestroom basement
                              bedrooms
                                          bathrooms
               price
                       area
                                      2
            2100000
                       3360
                                                   1
       556
                                                             1
                                                                     yes
                                                                                  no
                                                                                             no
                       3420
                                      5
                                                   1
                                                             2
       557
             1960000
                                                                      no
                                                                                  no
                                                                                             no
                                      3
                                                   1
                                                             2
       558
             1890000
                       1700
                                                                     yes
                                                                                  no
                                                                                             no
       559
             1890000
                       3649
                                      2
                                                   1
                                                             1
                                                                     yes
                                                                                  nο
                                                                                             nο
       560
             1855000
                       2990
                                      2
                                                   1
                                                             1
                                                                                  nο
                                                                                            no
                                                                      no
       561
             1820000
                       3000
                                      2
                                                   1
                                                             1
                                                                     yes
                                                                                  no
                                                                                            yes
                                      3
       562
            1767150
                       2400
                                                   1
                                                             1
                                                                                  no
                                                                                            no
                                                                      no
                                      2
       563
             1750000
                       3620
                                                   1
                                                             1
                                                                     yes
                                                                                  no
                                                                                             no
                                      3
                                                   1
       564
             1750000
                       2910
                                                             1
                                                                      no
                                                                                  no
                                                                                             no
       565
             1750000
                       3850
                                      3
                                                   1
                                                             2
                                                                     yes
                                                                                             no
                              airconditioning
                                                  parking prefarea furnishingstatus
           hotwaterheating
       556
                                             no
                                                         1
                                                                           unfurnished
                          no
                                                                  no
                                                         0
       557
                                                                           unfurnished
                                                                  no
                          no
                                             nο
                                                         0
       558
                                                                           unfurnished
                          no
                                             no
                                                                  no
                                                         0
       559
                                                                           unfurnished
                           no
                                                                  no
```

[176]: print("\n Sample Data:")

```
560
                                                                unfurnished
                                               1
                  no
                                    no
                                                       no
561
                                               2
                                                                unfurnished
                  no
                                    no
                                                       no
562
                                               0
                                                             semi-furnished
                  no
                                    no
                                                       no
563
                                               0
                                                                unfurnished
                  no
                                    no
                                                       no
                                               0
                                                                  furnished
564
                  no
                                    no
                                                       no
565
                                               0
                                                                unfurnished
                  no
                                    no
                                                       no
```

[178]: print("basic Information:")
print(data.info())

basic Information:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 566 entries, 0 to 565
Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	price	566 non-null	int64
1	area	566 non-null	int64
2	bedrooms	566 non-null	int64
3	bathrooms	566 non-null	int64
4	stories	566 non-null	int64
5	mainroad	566 non-null	object
6	guestroom	566 non-null	object
7	basement	566 non-null	object
8	hotwaterheating	566 non-null	object
9	airconditioning	566 non-null	object
10	parking	566 non-null	int64
11	prefarea	566 non-null	object
12	furnishingstatus	566 non-null	object

dtypes: int64(6), object(7)
memory usage: 57.6+ KB

None

[179]: data.dtypes

[179]: price int64 area int64 int64 bedrooms bathrooms int64 stories int64 object mainroad guestroom object object basement hotwaterheating object object airconditioning parking int64object prefarea furnishingstatus object

```
dtype: object
[180]: data.columns
[180]: Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
               'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
               'parking', 'prefarea', 'furnishingstatus'],
             dtype='object')
[181]:
       data.shape
[181]: (566, 13)
[182]: print("\n Summary Statistics:")
       print(data.describe())
       Summary Statistics:
                     price
                                     area
                                              bedrooms
                                                          bathrooms
                                                                         stories
      count
              5.660000e+02
                               566.000000
                                            566.000000
                                                         566.000000
                                                                     566.000000
      mean
              4.666197e+06
                              5076.773852
                                              2.950530
                                                           1.275618
                                                                        1.786219
      std
              1.906052e+06
                              2168.049072
                                              0.746217
                                                           0.496008
                                                                        0.861294
              1.750000e+06
                              1650.000000
      min
                                              1.000000
                                                           1.000000
                                                                        1.000000
      25%
              3.360000e+06
                              3514.000000
                                              2.000000
                                                           1.000000
                                                                        1.000000
      50%
              4.270000e+06
                              4500.000000
                                              3.000000
                                                           1.000000
                                                                       2.000000
      75%
              5.639375e+06
                              6357.500000
                                              3.000000
                                                           2.000000
                                                                        2.000000
              1.330000e+07
                             16200.000000
                                              6.000000
                                                           4.000000
                                                                        4.000000
      max
                 parking
              566.000000
      count
      mean
                0.674912
                0.856194
      std
      min
                0.000000
      25%
                0.000000
      50%
                0.000000
      75%
                1.000000
      max
                3.000000
[183]: data.isnull()
[183]:
                           bedrooms
                                                                      guestroom
            price
                     area
                                      bathrooms
                                                  stories
                                                           mainroad
       0
                              False
                                          False
                                                    False
            False
                    False
                                                               False
                                                                          False
       1
            False
                    False
                              False
                                          False
                                                    False
                                                               False
                                                                          False
       2
            False
                    False
                              False
                                          False
                                                    False
                                                               False
                                                                          False
       3
            False
                    False
                              False
                                          False
                                                    False
                                                               False
                                                                          False
       4
            False
                    False
                              False
                                          False
                                                    False
                                                               False
                                                                          False
                                                                •••
```

False

False

False

False

561

False

False

False

```
False
                                                   False
       562 False False
                              False
                                                              False
                                                                          False
       563 False False
                              False
                                          False
                                                   False
                                                              False
                                                                          False
       564 False False
                                                   False
                              False
                                          False
                                                              False
                                                                          False
       565 False False
                              False
                                          False
                                                   False
                                                              False
                                                                          False
                                         airconditioning parking prefarea \
            basement hotwaterheating
       0
               False
                                 False
                                                   False
                                                             False
                                                                        False
       1
               False
                                 False
                                                   False
                                                             False
                                                                        False
       2
               False
                                 False
                                                   False
                                                                        False
                                                             False
       3
               False
                                 False
                                                   False
                                                             False
                                                                        False
                                 False
                                                                        False
       4
               False
                                                   False
                                                             False
                 •••
       561
               False
                                 False
                                                   False
                                                             False
                                                                        False
       562
               False
                                 False
                                                   False
                                                                        False
                                                             False
       563
               False
                                 False
                                                   False
                                                             False
                                                                        False
       564
               False
                                 False
                                                   False
                                                             False
                                                                        False
       565
               False
                                 False
                                                   False
                                                             False
                                                                        False
            furnishingstatus
                        False
       0
       1
                        False
       2
                        False
       3
                        False
       4
                        False
       . .
       561
                        False
       562
                        False
       563
                        False
       564
                        False
       565
                        False
       [566 rows x 13 columns]
[184]: data.isnull().sum()
[184]: price
                            0
       area
                            0
       bedrooms
                            0
       bathrooms
                            0
       stories
                            0
       mainroad
                            0
       guestroom
                            0
```

0

0

0

0

0

basement

parking

prefarea

hotwaterheating

airconditioning

```
dtype: int64
[185]: data.dropna(inplace=True)
[186]: data.count()
[186]: price
                            566
                            566
       area
                            566
       bedrooms
       bathrooms
                            566
       stories
                            566
       mainroad
                            566
       guestroom
                            566
       basement
                            566
       hotwaterheating
                            566
                            566
       airconditioning
                            566
       parking
                            566
       prefarea
       furnishingstatus
                            566
       dtype: int64
[187]: duplicate_rows_df=data[data.duplicated()]
       print("number of duplicate row:",duplicate_rows_df.shape)
      number of duplicate row: (21, 13)
[188]: data.count()
[188]: price
                            566
                            566
       area
       bedrooms
                            566
       bathrooms
                            566
       stories
                            566
       mainroad
                            566
                            566
       guestroom
       basement
                            566
                            566
       hotwaterheating
       airconditioning
                            566
       parking
                            566
       prefarea
                            566
       furnishingstatus
                            566
       dtype: int64
[189]: data=data.drop_duplicates()
[190]: data.count()
```

0

furnishingstatus

```
[190]: price
                           545
                           545
       area
       bedrooms
                           545
       bathrooms
                           545
       stories
                           545
      mainroad
                           545
       guestroom
                           545
      basement
                           545
      hotwaterheating
                           545
       airconditioning
                           545
                           545
       parking
       prefarea
                           545
       furnishingstatus
                           545
       dtype: int64
[191]: features=data.columns
       features
[191]: Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
              'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
              'parking', 'prefarea', 'furnishingstatus'],
             dtype='object')
[192]: zero_val_cols=(data[features]==0).sum()
       zero_val_cols
[192]: price
                              0
       area
                              0
       bedrooms
                              0
       bathrooms
                              0
       stories
                              0
      mainroad
                              0
                              0
       guestroom
                              0
       basement
       hotwaterheating
                              0
       airconditioning
                              0
       parking
                            299
       prefarea
                              0
       furnishingstatus
                              0
       dtype: int64
[193]: data.isnull().sum()/len(data)*100
                            0.0
[193]: price
      area
                           0.0
       bedrooms
                           0.0
       bathrooms
                           0.0
       stories
                           0.0
```

```
basement
                            0.0
       hotwaterheating
                            0.0
                            0.0
       airconditioning
                            0.0
       parking
                            0.0
       prefarea
       furnishingstatus
                            0.0
       dtype: float64
[194]: | #data[['price', 'area']]=data[['price', 'area']].replace(0,np.NaN)
       # before removing Os we have to convert Os to NaN
       data.loc[:,['price','area']]=data[['price','area']].replace(0,np.NaN)
[195]: #filling null values with median of that column
       #data.area.fillna(data.area.median(),inplace=True)
       data.loc['area']=data['area'].fillna(data.area.median())
      C:\Users\DSU-CSE513-16\AppData\Local\Temp\ipykernel_6136\492706799.py:3:
      SettingWithCopyWarning:
      A value is trying to be set on a copy of a slice from a DataFrame
      See the caveats in the documentation: https://pandas.pydata.org/pandas-
      docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
        data.loc['area']=data['area'].fillna(data.area.median())
[196]: data
[196]:
                  price
                            area
                                  bedrooms
                                            bathrooms
                                                        stories mainroad guestroom \
       0
             13300000.0
                         7420.0
                                       4.0
                                                   2.0
                                                            3.0
                                                                      yes
                                                                                  no
       1
                                       4.0
                                                   4.0
                                                            4.0
             12250000.0
                          8960.0
                                                                      yes
                                                                                  no
       2
                                                   2.0
                                                            2.0
             12250000.0
                          9960.0
                                       3.0
                                                                      yes
                                                                                 no
       3
                                       4.0
                                                   2.0
             12215000.0
                          7500.0
                                                            2.0
                                                                      yes
                                                                                 no
       4
             11410000.0
                          7420.0
                                       4.0
                                                   1.0
                                                            2.0
                                                                      yes
                                                                                 yes
                          2400.0
                                                   1.0
                                                            1.0
       541
              1767150.0
                                       3.0
                                                                       no
                                                                                 no
       542
              1750000.0
                          3620.0
                                       2.0
                                                   1.0
                                                            1.0
                                                                      yes
                                                                                 no
       543
              1750000.0
                          2910.0
                                       3.0
                                                   1.0
                                                            1.0
                                                                       no
                                                                                 no
       544
              1750000.0
                          3850.0
                                       3.0
                                                   1.0
                                                            2.0
                                                                      yes
                                                                                 no
                    NaN
                             NaN
                                       NaN
       area
                                                   NaN
                                                            NaN
                                                                      NaN
                                                                                NaN
            basement hotwaterheating airconditioning
                                                        parking prefarea
       0
                                                            2.0
                  no
                                   no
                                                   yes
                                                                      yes
       1
                  no
                                   nο
                                                   yes
                                                            3.0
                                                                       no
       2
                                                            2.0
                 yes
                                   no
                                                    no
                                                                      yes
       3
                                                            3.0
                 yes
                                   no
                                                   yes
                                                                      yes
       4
                                                            2.0
                 yes
                                                   yes
                                                                       no
                                   no
```

mainroad

guestroom

0.0

E / 1					0 0			
541 542	no		no	no				
542 543	no		no	no				
544	no		no	no				
area	no NaN		no NaN	no NaN				
area	. Ivan		Nan	Ivaiv	IValv	Ivaiv		
	furnishings	tatus						
0	_	ished						
1	furn	ished						
2	semi-furn	ished						
3	furn	ished						
4	furn	ished						
•••		···						
541	semi-furn	ished						
542	unfurn	ished						
543	furn	ished						
544	unfurn	ished						
area		NaN						
[546	rows x 13 c	olumns]						
	# <i>b</i>							
ono	#one-hot en		dummi og (da	.+o columna	- [lmoinno	odll profix-	[lmoinmos	۱۱۱ه
	t("one-hot e			ica, corumns		oad'],prefix=	L Mainio	au] /
_			ata.)					
_	t(one_hot_en		ata.)					
prin		coded)	ava.)					
prin	t(one_hot_en not encoded d price	coded) lata: area	bedrooms	bathrooms	stories	guestroom ba	sement	\
prin	t(one_hot_en not encoded d price 13300000.0	coded) lata: area 7420.0	bedrooms	2.0	stories	guestroom ba	sement no	\
prin one-h	not encoded d price 13300000.0 12250000.0	data: area 7420.0 8960.0	bedrooms 4.0 4.0	2.0 4.0	3.0 4.0	_	no no	\
prin one-h 0 1 2	rt(one_hot_en not encoded d price 13300000.0 12250000.0	data: area 7420.0 8960.0 9960.0	bedrooms 4.0 4.0 3.0	2.0 4.0 2.0	3.0 4.0 2.0	no	no	\
prin one-h 0 1 2 3	not encoded d price 13300000.0 12250000.0 12250000.0 12215000.0	coded) lata: area 7420.0 8960.0 9960.0 7500.0	bedrooms 4.0 4.0 3.0 4.0	2.0 4.0 2.0 2.0	3.0 4.0 2.0 2.0	no no	no no yes yes	\
prin one-h 0 1 2	rt(one_hot_en not encoded d price 13300000.0 12250000.0	data: area 7420.0 8960.0 9960.0	bedrooms 4.0 4.0 3.0	2.0 4.0 2.0	3.0 4.0 2.0	no no no	no no yes	\
prin one-h 0 1 2 3 4	not encoded d price 13300000.0 12250000.0 12250000.0 12215000.0 11410000.0	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0	bedrooms 4.0 4.0 3.0 4.0 4.0	2.0 4.0 2.0 2.0 1.0	3.0 4.0 2.0 2.0 2.0	no no no no	no no yes yes	\
prin one-h 0 1 2 3 4 541	t(one_hot_en price 13300000.0 12250000.0 12250000.0 12215000.0 11410000.0 1767150.0	data: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0	2.0 4.0 2.0 2.0 1.0 	3.0 4.0 2.0 2.0 2.0 	no no no no	no no yes yes	\
prin one-h 0 1 2 3 4 541 542	t(one_hot_en price 13300000.0 12250000.0 12215000.0 11410000.0 1767150.0 1750000.0	data: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 2.0	2.0 4.0 2.0 2.0 1.0 1.0	3.0 4.0 2.0 2.0 2.0 1.0	no no no no yes no	no no yes yes yes	\
prin one-h 0 1 2 3 4 541 542 543	t(one_hot_en price 13300000.0 12250000.0 12250000.0 12215000.0 11410000.0 1767150.0 1750000.0	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0	bedrooms 4.0 4.0 3.0 4.0 4.0 2.0 3.0	2.0 4.0 2.0 2.0 1.0 1.0 1.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0	no no no no yes no no	no no yes yes yes no no no	\
prin one-h 0 1 2 3 4 541 542 543 544	t(one_hot_en price 13300000.0 12250000.0 12250000.0 12215000.0 11410000.0 1767150.0 1750000.0 1750000.0	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 3.0 3.0 3.0	2.0 4.0 2.0 2.0 1.0 1.0 1.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0	no no no no yes no no no	no no yes yes yes no no no	\
prin one-h 0 1 2 3 4 541 542 543	t(one_hot_en price 13300000.0 12250000.0 12250000.0 12215000.0 11410000.0 1767150.0 1750000.0	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0	bedrooms 4.0 4.0 3.0 4.0 4.0 2.0 3.0	2.0 4.0 2.0 2.0 1.0 1.0 1.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0	no no no no yes no no	no no yes yes yes no no no	\
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 3.0 2.0 3.0 3.0 NaN	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN	no no no no yes no no no NaN	no no yes yes yes no no no NaN	\
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 3.0 NaN	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN	no no no no yes no no no no no no furnishingst	no no yes yes yes no no no no NaN	\
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN cing airc	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 2.0 3.0 3.0 NaN conditioning	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN g parking s 2.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN prefarea yes	no no no no yes no no no no no for no furnishingst furni	no no yes yes yes no no no no shaN	
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN cing airc	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 2.0 3.0 NaN conditioning	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN g parking s 2.0 3.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN prefarea yes no	no no no no yes no no no no no furnishingst furni furni	no no yes yes yes no no no NaN atus \ shed shed	
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN sing airc	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 2.0 3.0 3.0 NaN conditioning yes	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN g parking s 2.0 3.0 2.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN prefarea yes no yes	no no no no yes no no no no no no furnishingst furni furni semi-furni	no no yes yes yes no no no NaN atus \ shed shed	
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN cing airc	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 2.0 3.0 3.0 NaN conditioning yes yes	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN g parking s 2.0 s 3.0 c 2.0 s 3.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN prefarea yes no yes yes	no no no no no yes no no no no no no to no to trinishingst furni furni semi-furni furni	no no yes yes yes no no no no shed shed shed	
prin one-h 0 1 2 3 4 541 542 543 544 area 0 1 2 3 4	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN sing airc	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 2.0 3.0 3.0 NaN conditioning yes	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN g parking s 2.0 s 3.0 2.0 s 3.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN prefarea yes no yes	no no no no yes no no no no no no furnishingst furni furni semi-furni	no no yes yes yes no no no no shed shed shed	
prin one-h 0 1 2 3 4 541 542 543 544 area	t(one_hot_en not encoded d	coded) lata: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN cing airc	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 2.0 3.0 3.0 NaN conditioning yes yes	2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 1.0 3.0 2.0 3.0 2.0 3.0 2.0	3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN prefarea yes no yes yes	no no no no no yes no no no no no no to no to trinishingst furni furni semi-furni furni	no no yes yes yes no no no no NaN atus \ shed shed shed shed shed	

[197]:

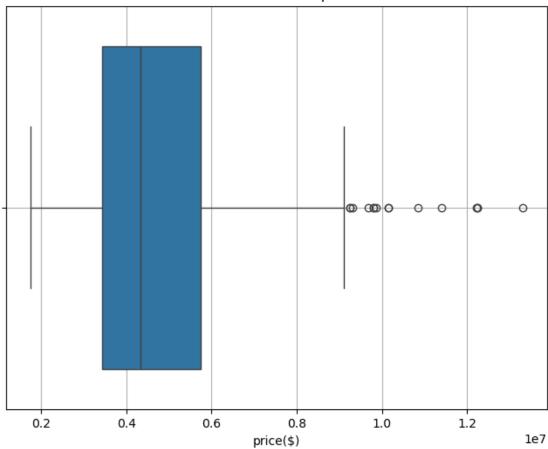
542		no	r	0.0	no	unf	urnished	
543		no	n	0.0	no	fı	urnished	
544		no	n	0.0	no	unf	urnished	
area		NaN	Na	ıN NaN	NaN		NaN	
	mainroad_no		•					
0	False		True					
1	False		True					
2	False		True					
3	False		True					
4	False		True					
•••	•••	•••						
541	True		False					
542	False		True					
543	True		False					
544	False		True					
area	False		False					
data	a['Guestroom_	LabelEnc	oded']=lab	el_encoder.:	fit_trans	form(data	a['guestroo	m'])
<pre>label_encoder = Lab data['Guestroom_Lab print("\n Label Enc print(data)</pre>		LabelEnc	oded']=lab	el_encoder.	fit_trans	form(data	a['guestroo	m'])
data prin	a['Guestroom_ nt("\n Label	LabelEnc Encoded	oded']=lab	el_encoder.	fit_trans	sform(data	a['guestroo	m'])
data prin	a['Guestroom_int("\n Label Int(data)	LabelEnc Encoded	oded']=lab	el_encoder.			a['guestroo	m'])
data prin	a['Guestroom_] nt("\n Label nt(data) el Encoded Da	LabelEnc Encoded : ta: area	oded']=lab Data:")					
data prin prin	a['Guestroom_int("\n Label int(data) el Encoded Da price 13300000.0	LabelEnc Encoded : ta: area	oded']=lab Data:") bedrooms	bathrooms	stories	mainroad	guestroom	
data prin prin	a['Guestroom_int("\n Label int(data) el Encoded Da price 13300000.0	ta: area 7420.0	oded']=lab Data:") bedrooms 4.0	bathrooms 2.0	stories	mainroad yes	guestroom no	
data prin prin Labo	a['Guestroom_lat("\n Label lat(data) el Encoded Da	ta: area 7420.0 8960.0	oded']=lab Data:") bedrooms 4.0 4.0	bathrooms 2.0 4.0	stories 3.0 4.0	mainroad yes yes	guestroom no no	
data prin prin Labe	a['Guestroom_lat("\n Label lat(data) el Encoded Da	ta: area 7420.0 8960.0	oded']=lab Data:") bedrooms 4.0 4.0 3.0	bathrooms 2.0 4.0 2.0	stories 3.0 4.0 2.0	mainroad yes yes yes	guestroom no no	
data prin prin Labe	a['Guestroom_int("\n Label int(data) el Encoded Da	ta: area 7420.0 8960.0 9960.0 7500.0 7420.0	oded']=lab Data:") bedrooms 4.0 4.0 3.0 4.0	bathrooms 2.0 4.0 2.0 2.0 1.0	stories 3.0 4.0 2.0 2.0	mainroad yes yes yes yes	guestroom no no no	
data prin prin Labe	a['Guestroom_int("\n Label int(data) el Encoded Da	ta: area 7420.0 8960.0 9960.0 7500.0 7420.0	oded']=lab Data:") bedrooms	bathrooms 2.0 4.0 2.0 2.0 1.0	stories 3.0 4.0 2.0 2.0	mainroad yes yes yes yes	guestroom no no no	
data prin prin Labe	el Encoded Da price 13300000.0 12250000.0 12215000.0 11410000.0	ta: area 7420.0 8960.0 9960.0 7500.0 7420.0	oded']=lab Data:") bedrooms	bathrooms 2.0 4.0 2.0 2.0 1.0	stories 3.0 4.0 2.0 2.0 2.0	mainroad yes yes yes yes	guestroom no no no no yes	
data prin prin 1	el Encoded Da price 1330000.0 12250000.0 12410000.0 1750000.0 1750000.0	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0	bedrooms 4.0 4.0 4.0 4.0 2.0 3.0	bathrooms 2.0 4.0 2.0 2.0 1.0 1.0 1.0	stories 3.0 4.0 2.0 2.0 2.0 1.0	mainroad yes yes yes yes yes	guestroom no no no no yes	
data prin prin 1	a['Guestroom_lat("\n Label Int(data) el Encoded Daprice 13300000.0 12250000.0 12215000.0 11410000.0 1767150.0 1750000.0	ta: area 7420.0 8960.0 9960.0 7500.0 7420.0 2400.0 3620.0	bedrooms 4.0 4.0 4.0 4.0 3.0 4.0 2.0	bathrooms 2.0 4.0 2.0 2.0 1.0	stories 3.0 4.0 2.0 2.0 2.0 	mainroad yes yes yes yes yes yes	guestroom no no no no yes no	
data prin prin 1	a['Guestroom_int("\n Label int(data) el Encoded Da price 13300000.0 12250000.0 12250000.0 12215000.0 11410000.0 1767150.0 1750000.0 1750000.0	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0	bedrooms 4.0 4.0 4.0 4.0 2.0 3.0	bathrooms 2.0 4.0 2.0 2.0 1.0 1.0 1.0	stories 3.0 4.0 2.0 2.0 2.0 1.0	mainroad yes yes yes yes yes yes no	guestroom no no no no yes no no	
data prin prin Labe 0 1 2 3 4 541 542 543 544	a['Guestroom_] at("\n Label] at("\n Label] at(data) el Encoded Da	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 4.0 4.0 3.0 4.0 3.0 A.0 A.0 NaN	bathrooms 2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN	stories 3.0 4.0 2.0 2.0 2.0 1.0 1.0 2.0 NaN	mainroad yes yes yes yes yes no yes no yes NaN	guestroom no no no no yes no no no no NaN	
data prin prin 1	a['Guestroom_int("\n Label int(data) el Encoded Da price 13300000.0 12250000.0 12250000.0 12215000.0 1750000.0 1750000.0 NaN basement hot	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 3.0 4.0 3.0 4.0 3.0 4.0	bathrooms 2.0 4.0 2.0 2.0 1.0 1.0 1.0 1.0 NaN	stories 3.0 4.0 2.0 2.0 1.0 1.0 1.0 2.0 NaN	mainroad yes yes yes yes yes no yes NaN	guestroom no no no no yes no no no no NaN	
data prin prin 1	a['Guestroom_lat("\n Label Int(data) el Encoded Da price 13300000.0 12250000.0 12250000.0 12215000.0 1750000.0 1750000.0 1750000.0 NaN basement hot	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 4.0 bedrooms 4.0 4.0 3.0 4.0 4.0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	bathrooms	stories 3.0 4.0 2.0 2.0 2.0 1.0 1.0 1.0 parking 2.0	mainroad yes yes yes yes yes no yes no yes NaN prefarea	guestroom no no no no yes no no no no NaN	
data prin prin 1	a['Guestroom_lat("\n Label Int(data) el Encoded Da price 13300000.0 12250000.0 12215000.0 1750000.0 1750000.0 NaN basement hot no no	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 5.0 3.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0	bathrooms	stories 3.0 4.0 2.0 2.0 2.0 1.0 1.0 1.0 2.0 NaN parking 2.0 3.0	mainroad yes yes yes yes yes no yes NaN prefarea yes no	guestroom no no no no yes no no no no NaN	
data prin prin 1	a['Guestroom_lat("\n Label Int(data) el Encoded Da price 13300000.0 12250000.0 12250000.0 12215000.0 1750000.0 1750000.0 1750000.0 NaN basement hot	ta: area 7420.0 8960.0 7500.0 7420.0 2400.0 3620.0 2910.0 3850.0 NaN	bedrooms 4.0 4.0 3.0 4.0 4.0 3.0 4.0 bedrooms 4.0 4.0 3.0 4.0 4.0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	bathrooms	stories 3.0 4.0 2.0 2.0 2.0 1.0 1.0 1.0 parking 2.0	mainroad yes yes yes yes yes no yes no yes NaN prefarea	guestroom no no no no yes no no no no NaN	

```
541
                                                           0.0
                                  no
                                                                     no
                 no
                                                   no
      542
                                                           0.0
                                  no
                                                                     no
                 no
                                                   no
      543
                                                           0.0
                 no
                                  no
                                                                     no
                                                   no
      544
                                                           0.0
                 no
                                  no
                                                   no
                                                                     no
                NaN
                                 NaN
                                                           NaN
                                                                    NaN
      area
                                                  NaN
                              Guestroom_LabelEncoded
           furnishingstatus
      0
                  furnished
      1
                  furnished
                                                    0
      2
             semi-furnished
                                                    0
      3
                  furnished
                                                    0
      4
                  furnished
                                                    1
                                                    0
      541
             semi-furnished
      542
                unfurnished
                                                    0
      543
                  furnished
                                                    0
                unfurnished
                                                    0
      544
                         NaN
                                                    2
      area
      [546 rows x 14 columns]
      C:\Users\DSU-CSE513-16\AppData\Local\Temp\ipykernel_6136\2421631340.py:2:
      SettingWithCopyWarning:
      A value is trying to be set on a copy of a slice from a DataFrame.
      Try using .loc[row_indexer,col_indexer] = value instead
      See the caveats in the documentation: https://pandas.pydata.org/pandas-
      docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
        data['Guestroom_LabelEncoded']=label_encoder.fit_transform(data['guestroom'])
[200]: #visualize the distribution of house prices
       plt.figure(figsize=(8,6))
       sns.histplot(data['price'],bins=20,kde=True)
       plt.title("Distribution of House Prices")
       plt.xlabel("Price ($)")
       plt.ylabel("Frequency")
       plt.grid(True)
       plt.show()
```



```
[201]: #visualize outliers in house price
plt.figure(figsize=(8,6))
sns.boxplot(data=data,x='price')
plt.title("outliers in house price")
plt.xlabel("price($)")
plt.grid(True)
plt.show()
```

outliers in house price



```
[202]: data = data[data['price']<9000000]

[203]: def find_boundaries(variable):
        q1 = data[variable].quantile(0.25)
        q3 = data[variable].quantile(0.75)
        iqr = q3 - q1
        lower_boundary = q1 - 1.5 * iqr
        upper_boundary = q3 + 1.5 * iqr
        return lower_boundary, upper_boundary

lower_price, upper_price = find_boundaries('price')
        data.price = np.where(data.price > upper_price, upper_price, data.price)
        data.price = np.where(data.price < lower_price, lower_price, data.price)</pre>
```

C:\Users\DSU-CSE513-16\AppData\Local\Temp\ipykernel_6136\2133715101.py:10: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame. Try using .loc[row_indexer,col_indexer] = value instead

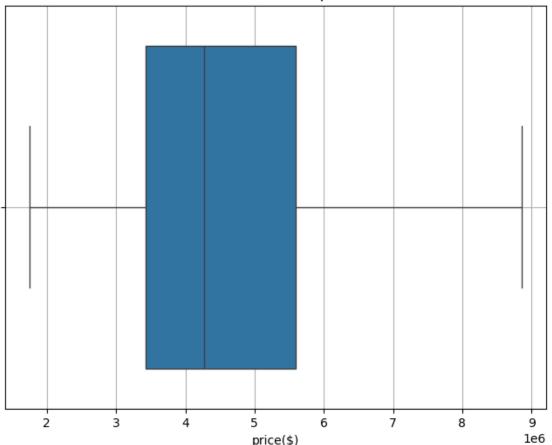
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy data.price = np.where(data.price > upper_price, upper_price, data.price)
C:\Users\DSU-CSE513-16\AppData\Local\Temp\ipykernel_6136\2133715101.py:11:
SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame. Try using .loc[row_indexer,col_indexer] = value instead

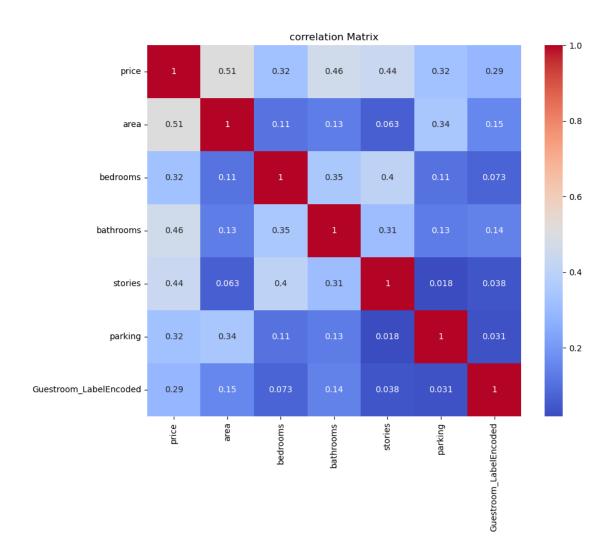
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy data.price = np.where(data.price < lower_price, lower_price, data.price)

```
[204]: plt.figure(figsize=(8,6))
    sns.boxplot(data=data,x='price')
    plt.title("outliers in house price")
    plt.xlabel("price($)")
    plt.grid(True)
    plt.show()
```

outliers in house price



```
[205]: #calculate correlation matrix
      numeric_data =data.select_dtypes(include=[np.number])
      correlation_matrix=numeric_data.corr()
      print(correlation_matrix)
                                           area bedrooms bathrooms
                                                                       stories \
                                price
      price
                              1.000000 0.511642 0.323128
                                                            0.460610 0.436452
                              0.511642 1.000000 0.109293
                                                            0.132166 0.063436
      area
                                                            0.349523 0.404938
      bedrooms
                             0.323128 0.109293 1.000000
      bathrooms
                             0.460610 0.132166 0.349523
                                                            1.000000 0.308414
                             0.436452 0.063436 0.404938
      stories
                                                            0.308414 1.000000
      parking
                             0.323307 0.343992 0.105479
                                                            0.128327
                                                                      0.018348
      Guestroom_LabelEncoded 0.286845 0.153728 0.072505
                                                            0.141416 0.037742
                              parking Guestroom_LabelEncoded
                              0.323307
                                                     0.286845
      price
      area
                             0.343992
                                                     0.153728
      bedrooms
                             0.105479
                                                     0.072505
      bathrooms
                             0.128327
                                                     0.141416
      stories
                             0.018348
                                                     0.037742
                                                     0.030774
      parking
                              1.000000
                                                     1.000000
      Guestroom_LabelEncoded 0.030774
[207]: #visualize correlation matrix
      plt.figure(figsize=(10,8))
      sns.heatmap(correlation_matrix,annot=True,cmap="coolwarm")
      plt.title("correlation Matrix")
      plt.show()
```



```
[209]: plt.figure(figsize=(10,6))
    sns.scatterplot(data=data,x='area',y='price')
    plt.title("House Price vs. Area")
    plt.xlabel("Area")
    plt.ylabel("Price($)")
    plt.grid(True)
    plt.show()
```



[]: