```
In [1]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
In [2]: data = pd.read_csv(r'C:/Users/aksha_0iafvb0/OneDrive/Desktop/housing.csv')
In [3]: data
Out[3]:
                longitude latitude housing_median_age total_rooms total_bedrooms population households median_income median_house_value ocean_proximity
                  -122.23
                           37.88
                                               41.0
                                                          880.0
                                                                        129.0
                                                                                   322.0
                                                                                              126.0
                                                                                                           8.3252
                                                                                                                             452600.0
                                                                                                                                          NEAR BAY
             1
                  -122.22
                           37.86
                                               21.0
                                                         7099.0
                                                                        1106.0
                                                                                  2401.0
                                                                                             1138.0
                                                                                                           8.3014
                                                                                                                             358500.0
                                                                                                                                          NEAR BAY
                  -122.24
                                               52.0
                                                         1467.0
                                                                         190.0
                                                                                   496.0
                                                                                              177.0
                                                                                                            7.2574
                                                                                                                             352100.0
                                                                                                                                          NEAR BAY
             2
                           37.85
                  -122.25
             3
                           37.85
                                               52.0
                                                         1274.0
                                                                        235.0
                                                                                  558.0
                                                                                              219.0
                                                                                                           5.6431
                                                                                                                             341300.0
                                                                                                                                          NEAR BAY
                  -122.25
                           37.85
                                               52.0
                                                         1627.0
                                                                        280.0
                                                                                   565.0
                                                                                              259.0
                                                                                                           3.8462
                                                                                                                             342200.0
                                                                                                                                          NEAR BAY
          20635
                  -121.09
                           39.48
                                               25.0
                                                         1665.0
                                                                        374.0
                                                                                  845.0
                                                                                              330.0
                                                                                                            1.5603
                                                                                                                             78100.0
                                                                                                                                             INLAND
          20636
                  -121.21
                           39.49
                                               18.0
                                                          697.0
                                                                        150.0
                                                                                  356.0
                                                                                              114.0
                                                                                                            2.5568
                                                                                                                             77100.0
                                                                                                                                             INLAND
          20637
                  -121.22
                           39.43
                                               17.0
                                                         2254.0
                                                                        485.0
                                                                                  1007.0
                                                                                              433.0
                                                                                                            1.7000
                                                                                                                             92300.0
                                                                                                                                             INLAND
                  -121.32
                                                                        409.0
                                                                                  741.0
                                                                                                                              84700.0
          20638
                                                18.0
                                                         1860.0
                                                                                              349.0
                                                                                                            1.8672
                                                                                                                                             INLAND
                                                                                                                              89400.0
          20639
                  -121 24
                           39 37
                                               16.0
                                                         2785.0
                                                                        616.0
                                                                                  1387 0
                                                                                              530.0
                                                                                                            2 3886
                                                                                                                                             INI AND
         20640 rows × 10 columns
In [4]: data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 20640 entries, 0 to 20639
         Data columns (total 10 columns):
                                    Non-Null Count Dtype
              Column
          0
              longitude
                                    20640 non-null
                                                     float64
          1
              latitude
                                    20640 non-null
                                                     float64
              housing_median_age
                                    20640 non-null
                                                     float64
              total rooms
                                    20640 non-null
                                                     float64
              total bedrooms
                                    20433 non-null
                                                     float64
                                    20640 non-null
              population
                                                     float64
                                    20640 non-null
          6
              households
                                                     float64
                                    20640 non-null
              median income
                                                     float64
              median_house_value
                                    20640 non-null
                                                     float64
              ocean_proximity
                                    20640 non-null
                                                     object
         dtypes: float64(9), object(1)
         memory usage: 1.6+ MB
In [5]: data.dropna(inplace=True)
In [6]: data.info()
         <class 'pandas.core.frame.DataFrame'>
         Int64Index: 20433 entries, 0 to 20639
         Data columns (total 10 columns):
          #
              Column
                                    Non-Null Count Dtype
          0
              longitude
                                    20433 non-null
                                                      float64
              latitude
                                    20433 non-null
                                                     float64
              housing_median_age
                                    20433 non-null
                                                     float64
               total_rooms
                                    20433 non-null
                                                     float64
              total bedrooms
                                    20433 non-null
                                                     float64
                                    20433 non-null
          5
              population
                                                     float64
              households
                                    20433 non-null
          6
                                                     float64
                                    20433 non-null
              median income
                                                     float64
              median_house_value
                                    20433 non-null
                                                     float64
              ocean_proximity
                                    20433 non-null
                                                     object
         dtypes: float64(9), object(1)
         memory usage: 1.7+ MB
In [7]: from sklearn.model_selection import train_test_split
         # seperating the dependent variable column and independent column
         x = data.drop(['median_house_value'],axis=1) #Independent variables
         y = data['median_house_value'] #dependent variables
```

```
In [8]: x
Out[8]:
                   longitude latitude
                                     housing_median_age total_rooms
                                                                       total_bedrooms population households median_income ocean_proximity
                     -122.23
                                                     41.0
                                                                                                                                    NEAR BAY
                0
                               37.88
                                                                880.0
                                                                                129.0
                                                                                           322.0
                                                                                                        126.0
                                                                                                                       8.3252
                     -122.22
                               37.86
                                                     21.0
                                                                7099.0
                                                                                1106.0
                                                                                           2401.0
                                                                                                       1138.0
                                                                                                                       8.3014
                                                                                                                                    NEAR BAY
                2
                     -122.24
                               37.85
                                                     52.0
                                                                1467.0
                                                                                190.0
                                                                                           496.0
                                                                                                        177.0
                                                                                                                       7.2574
                                                                                                                                    NEAR BAY
                3
                     -122.25
                               37.85
                                                     52.0
                                                                1274.0
                                                                                235.0
                                                                                           558.0
                                                                                                        219.0
                                                                                                                       5.6431
                                                                                                                                    NEAR BAY
                     -122.25
                               37.85
                                                     52.0
                                                                1627.0
                                                                                280.0
                                                                                           565.0
                                                                                                        259.0
                                                                                                                       3.8462
                                                                                                                                    NEAR BAY
            20635
                     -121.09
                                                     25.0
                                                                1665.0
                                                                                374.0
                                                                                           845.0
                                                                                                        330.0
                                                                                                                       1.5603
                                                                                                                                      INLAND
                               39.48
            20636
                     -121 21
                               39 49
                                                     18.0
                                                                697.0
                                                                                150.0
                                                                                           356.0
                                                                                                        114.0
                                                                                                                       2 5568
                                                                                                                                      INI AND
                                                     17.0
                                                                                485.0
                                                                                           1007.0
                                                                                                                                      INLAND
            20637
                     -121.22
                               39.43
                                                               2254.0
                                                                                                        433.0
                                                                                                                       1.7000
            20638
                     -121.32
                               39.43
                                                     18.0
                                                                1860.0
                                                                                409.0
                                                                                           741.0
                                                                                                        349.0
                                                                                                                       1.8672
                                                                                                                                      INLAND
            20639
                     -121.24
                               39.37
                                                     16.0
                                                               2785.0
                                                                                616.0
                                                                                           1387.0
                                                                                                        530.0
                                                                                                                       2.3886
                                                                                                                                      INLAND
           20433 rows × 9 columns
In [9]: y
Out[9]: 0
                     452600.0
                     358500.0
                      352100.0
           2
                      341300.0
           4
                      342200.0
                       78100.0
           20635
           20636
                       77100.0
           20637
                       92300.0
           20638
                       84700.0
           20639
                       89400.0
          Name: median_house_value, Length: 20433, dtype: float64
In [10]: x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2)
In [11]: train_data = x_train.join(y_train)
In [12]: train_data
Out[12]:
                   longitude latitude housing_median_age total_rooms total_bedrooms population households median_income ocean_proximity median_house_value
            15372
                     -117.24
                               33.38
                                                     16.0
                                                               2792.0
                                                                                525.0
                                                                                           1696.0
                                                                                                        516.0
                                                                                                                       3.6680
                                                                                                                                   <1H OCEAN
                                                                                                                                                          171200.0
                                                     16.0
                                                                                345.0
                                                                                                                                                          365800.0
             1654
                     -122.01
                               37.92
                                                               2638.0
                                                                                           1055.0
                                                                                                        334.0
                                                                                                                       8.1163
                                                                                                                                    NEAR BAY
            20506
                     -118.73
                               34.29
                                                      8.0
                                                               4983.0
                                                                                754.0
                                                                                           2510.0
                                                                                                        725.0
                                                                                                                       6.9454
                                                                                                                                   <1H OCEAN
                                                                                                                                                          276500.0
                     -121.93
                                                                                852.0
                                                                                           1582.0
                                                                                                                                 NEAR OCEAN
                                                                                                                                                          313900.0
             9831
                               36.63
                                                     28.0
                                                               3983.0
                                                                                                        778.0
                                                                                                                       3.5147
             6318
                     -117.97
                               34.01
                                                     33.0
                                                               2006.0
                                                                                381.0
                                                                                           1410.0
                                                                                                        346.0
                                                                                                                       3.7083
                                                                                                                                   <1H OCEAN
                                                                                                                                                          165500.0
            14388
                     -117.23
                               32.74
                                                     16.0
                                                                735.0
                                                                                139.0
                                                                                           299.0
                                                                                                        134.0
                                                                                                                       4.6354
                                                                                                                                 NEAR OCEAN
                                                                                                                                                          179200.0
                     -118.06
                                                     28.0
                                                                1778.0
                                                                                605.0
                                                                                           2184.0
                                                                                                        574.0
                                                                                                                       1.9189
                                                                                                                                   <1H OCEAN
                                                                                                                                                          165900.0
             6530
             3902
                     -118.51
                               34 20
                                                     34 0
                                                               2871 0
                                                                                581.0
                                                                                           1350.0
                                                                                                        535.0
                                                                                                                       3 7049
                                                                                                                                   <1H OCEAN
                                                                                                                                                          227500 0
            10111
                     -117.94
                                                     25.0
                                                                3250.0
                                                                                546.0
                                                                                           1452.0
                                                                                                        501.0
                                                                                                                       5.1084
                                                                                                                                   <1H OCEAN
                                                                                                                                                          303800.0
                                                     37.0
                     -118.25
                                                                2119.0
                                                                                442.0
                                                                                           1372.0
                                                                                                        406.0
                                                                                                                       1.9605
                                                                                                                                   <1H OCEAN
                                                                                                                                                          106200.0
            7593
                               33.90
           16346 rows × 10 columns
```

```
In [13]: train_data.hist(figsize=(15, 8))
<AxesSubplot:title={'center':'latitude'}>,
    <AxesSubplot:title={'center':'housing_median_age'}>],
[<AxesSubplot:title={'center':'total_rooms'}>,
    <AxesSubplot:title={'center':'total_bedrooms'}>,
    <AxesSubplot:title={'center':'population'}>],
[<AxesSubplot:title={'center':'households'}>,
    <AxesSubplot:title={'center':'median_income'}>,
    <AxesSubplot:title={'center':'median_house_value'}>]],
dtyne=obiect)
                         dtype=object)
                                              longitude
                                                                                                                     latitude
                                                                                                                                                                                  housing_median_age
                  5000
                                                                                                                                                             2500
                                                                                        6000
                  4000
                                                                                                                                                             2000
                  3000
                                                                                        4000
                                                                                                                                                             1500
                                                                                        2000
                  1000
                                                                                                                                                               500
                                                                                            0
                                                                                                                                                                  0
                           -124
                                     -122
                                               -120
                                                        -118
                                                                  -116
                                                                           -114
                                                                                                        34
                                                                                                                                                 42
                                                                                                                                                                               10
                                                                                                                                                                                                           40
                                                                                                                                                                                                                     50
                                                                                                                                                                                        population
                                             total_rooms
                                                                                                                total_bedrooms
                                                                                       12500
                                                                                                                                                            15000
                 12500
                                                                                       10000
                 10000
                                                                                                                                                            10000
                  7500
                                                                                        5000
                  5000
                                                                                                                                                             5000
                  2500
                                                                                                                                                                  0
                                     10000
                                                                                                              2000
                                                                                                                                            6000
                                                                                                                                                                                 10000
                                                                                                                                                                                               20000
                                                 20000
                                                              30000
                                                                           40000
                                                                                                                              4000
                                                                                                                                                                                                             30000
                                             households
                                                                                                                                                                                 median house value
                                                                                                               median income
                 12500
                                                                                        6000
                 10000
                                                                                        4000
                  7500
                                                                                                                                                             2000
                  5000
                                                                                        2000
                                                                                                                                                              1000
                  2500
                                 1000 2000 3000 4000 5000 6000
                                                                                              0.0
                                                                                                                                        12.5
                                                                                                       2.5
                                                                                                                               10.0
                                                                                                                                                                           100000 200000 300000 400000 500000
                                                                                                               5.0
```

In [14]: train_data.corr()

Out[14]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
longitude	1.000000	-0.925096	-0.111758	0.051010	0.076175	0.102704	0.061867	-0.017517	-0.046921
latitude	-0.925096	1.000000	0.014189	-0.042309	-0.074197	-0.111976	-0.077274	-0.076567	-0.143006
housing_median_age	-0.111758	0.014189	1.000000	-0.360009	-0.319265	-0.293732	-0.302194	-0.120917	0.105039
total_rooms	0.051010	-0.042309	-0.360009	1.000000	0.932893	0.854297	0.920263	0.195959	0.132696
total_bedrooms	0.076175	-0.074197	-0.319265	0.932893	1.000000	0.875031	0.979374	-0.006331	0.051541
population	0.102704	-0.111976	-0.293732	0.854297	0.875031	1.000000	0.904967	0.004790	-0.025053
households	0.061867	-0.077274	-0.302194	0.920263	0.979374	0.904967	1.000000	0.013222	0.065891
median_income	-0.017517	-0.076567	-0.120917	0.195959	-0.006331	0.004790	0.013222	1.000000	0.685612
median_house_value	-0.046921	-0.143006	0.105039	0.132696	0.051541	-0.025053	0.065891	0.685612	1.000000

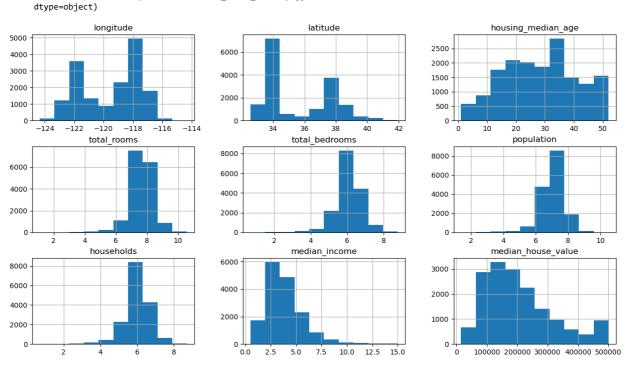
```
In [15]: plt.figure(figsize = (15,8))
sns.heatmap(train_data.corr(),annot=True, cmap="YlGnBu")
```

Out[15]: <AxesSubplot:>



```
In [16]: train_data['total_rooms'] = np.log(train_data['total_rooms'] + 1)
    train_data['total_bedrooms'] = np.log(train_data['total_bedrooms'] + 1)
    train_data['population'] = np.log(train_data['population'] + 1)
    train_data['households'] = np.log(train_data['households'] + 1)
```

```
In [17]: train_data.hist(figsize =(15,8))
```



```
In [18]: train_data.ocean_proximity.value_counts()
Out[18]:
            <1H OCEAN
                               7264
            INLAND
                               5171
            NEAR OCEAN
                               2110
            NEAR BAY
                              1797
            ISLAND
            Name: ocean_proximity, dtype: int64
In [19]: train_data = train_data.join(pd.get_dummies(train_data.ocean_proximity)).drop(['ocean_proximity'], axis = 1)
In [20]: train_data
Out[20]:
                                                                                                                                                                   <1H
OCEAN
                     longitude latitude
                                         housing_median_age total_rooms total_bedrooms population households median_income median_house_value
                                                                                                                                                                             INLAND
                       -117.24
             15372
                                  33.38
                                                                    7.934872
                                                                                      6.265301
                                                                                                   7.436617
                                                                                                                6.248043
                                                                                                                                    3.6680
                                                                                                                                                         171200.0
              1654
                       -122.01
                                  37.92
                                                           16.0
                                                                    7.878155
                                                                                                  6.962243
                                                                                                                5.814131
                                                                                                                                    8.1163
                                                                                                                                                         365800.0
                                                                                                                                                                                   C
                                                                                      5.846439
                                                                                                                                                                          0
             20506
                       -118.73
                                  34.29
                                                            8.0
                                                                    8.513988
                                                                                      6.626718
                                                                                                   7.828436
                                                                                                                6.587550
                                                                                                                                    6.9454
                                                                                                                                                         276500.0
                                                                                      6.748760
                                                                                                   7.367077
                                                                                                                6.658011
              9831
                       -121.93
                                  36.63
                                                           28.0
                                                                    8.290042
                                                                                                                                    3.5147
                                                                                                                                                         313900.0
                                                                                                                                                                          0
                                                                                                                                                                                   C
              6318
                       -117.97
                                  34.01
                                                           33.0
                                                                    7.604396
                                                                                      5.945421
                                                                                                   7.252054
                                                                                                                5.849325
                                                                                                                                    3.7083
                                                                                                                                                         165500.0
             14388
                       -117.23
                                  32.74
                                                           16.0
                                                                    6.601230
                                                                                      4.941642
                                                                                                   5.703782
                                                                                                                4.905275
                                                                                                                                    4.6354
                                                                                                                                                         179200.0
                                                                                                                                                                          0
              6530
                       -118.06
                                  34.06
                                                           28.0
                                                                    7.483807
                                                                                      6.406880
                                                                                                   7.689371
                                                                                                                6.354370
                                                                                                                                    1.9189
                                                                                                                                                         165900.0
              3902
                       -118.51
                                  34.20
                                                           34.0
                                                                    7.962764
                                                                                      6.366470
                                                                                                   7.208600
                                                                                                                6.284134
                                                                                                                                    3.7049
                                                                                                                                                         227500.0
             10111
                       -117.94
                                  33.94
                                                           25.0
                                                                    8.086718
                                                                                      6.304449
                                                                                                   7.281386
                                                                                                                6.218600
                                                                                                                                    5.1084
                                                                                                                                                         303800.0
                                                           37.0
                       -118.25
                                                                    7.659171
                                                                                      6.093570
                                                                                                  7.224753
                                                                                                                6.008813
                                                                                                                                    1.9605
              7593
                                  33.90
                                                                                                                                                         106200.0
            16346 rows × 14 columns
In [21]: plt.figure(figsize = (15,8))
            sns.heatmap(train_data.corr(),annot=True, cmap="YlGnBu")
Out[21]: <AxesSubplot:>
                                                                                                                                                                             1.00
                                              -0.93
                                                      -0.11
                                                                                                   -0.018
                        lonaitude
                                                                                                           -0.047
                                                                                                                             -0.055
                                                                                                                                                -0.47
                                    -0.93
                                                               -0.039
                                                                        -0.076
                                                                                  -0.14
                                                                                          -0.094
                                                                                                   -0.077
                                                                                                            -0.14
                                                                                                                     -0.45
                                                                                                                                      -0.017
                                                                                                                                                         -0.16
                          latitude
                                                                                                                                                                            0.75
             housing_median_age
                                     -0.11
                                                                -0.31
                                                                        -0.27
                                                                                  -0.24
                                                                                          -0.24
                                                                                                   -0.12
                                                                                                                              -0.24
                      total_rooms
                                                       -0.31
                                                                                                                                               -0.019
                                                                                                                                                                            0.50
                   total_bedrooms
                                             -0.076
                                                       -0.27
                                                                                  0.9
                                                                                                   -0.028
                                                                                                                              -0.046
                                                                                                                                                -0.02
                       population
                                              -0.14
                                                       -0.24
                                                                0.86
                                                                         0.9
                                                                                          0.93
                                                                                                            -0.022
                                                                                                                             -0.073
                                                                                                                                      -0.0098
                                                                                                                                               -0.063
                                                                                                                                                        -0.015
                                                                                                                                                                            0.25
                      households
                                             -0.094
                                                      -0.24
                                                                0.93
                                                                        0.97
                                                                                 0.93
                                                                                                                             -0.088
                                                                                                                                               -0.0095
                                                                                                                                                                            0.00
                                                                                                                                      -0.0092
                  median income
                                    -0.018
                                             -0.077
                                                       -0.12
                                                                        -0.028
                                                                                                            0.69
                                                                                                                              -0.24
                                                                                                   0.69
             median house value
                                    -0.047
                                              -0.14
                                                                                 -0.022
                                                                                                                              -0.49
                                                                                                                                                                             -0.25
                      <1H OCEAN
                                              -0.45
                                                                                                                              -0.61
                                                                                                                                      -0.014
                                                                                                                                                -0.31
                                                                                                                                                         -0.34
                                                                                                                                       -0.011
                          INLAND
                                    -0.055
                                                      -0.24
                                                               -0.014
                                                                        -0.046
                                                                                 -0.073
                                                                                          -0.088
                                                                                                            -0.49
                                                                                                                     -0.61
                                                                                                                                                -0.24
                                                                                                                                                         -0.26
                                                                                                                                                                            -0.50
                          ISLAND
                                             -0.017
                                                                                                  -0.0092
                                                                                                                     -0.014
                                                                                                                              -0.011
                        NEAR BAY
                                    -0.47
                                                               -0.019
                                                                        -0.02
                                                                                 -0.063
                                                                                         -0.0095
                                                                                                                     -0.31
                                                                                                                              -0.24
                                                                                                                                                         -0.14
                                                                                                                                                                             -0.75
                     NEAR OCEAN
                                              -0.16
                                                                                 -0.015
                                                                                                                     -0.34
                                                                                                                              -0.26
                                                                                                                                                -0.14
                                                                                                                      <1H OCEAN
                                                                                                                               INLAND
                                                                                                                                                         NEAR OCEAN
                                                                total_rooms
                                                                                                                                                 NEAR BAY
                                               latitude
                                                        housing_median_age
                                                                         total_bedrooms
                                                                                                             median_house_value
```

C

C

C

C

C

C

C

```
In [22]: plt.figure(figsize=(15,8))
            sns.scatterplot (x='latitude',y='longitude',data=train\_data,hue="median\_house\_value",palette="coolwarm") \\
Out[22]: <AxesSubplot:xlabel='latitude', ylabel='longitude'>
                -114
                                                                                                                                                       median house value
                                                                                                                                                               100000
                                                                                                                                                                200000
                                                                                                                                                                300000
                                                                                                                                                                400000
                                                                                                                                                                500000
                -116
                -118
             longitude
                -120
                -122
                -124
                                                                                                                                                                     42
                                                                                             latitude
In [23]: train_data['bedroom_ratio'] = train_data['total_bedrooms']/train_data['total_rooms']
            train_data['household_rooms'] = train_data['total_rooms'] /train_data['households']
In [24]: plt.figure(figsize=(15,8))
            sns.heatmap(train_data.corr(),annot=True,cmap="YlGnBu")
Out[24]: <AxesSubplot:>
                                                                                                                                                                      1.00
                       longitude
                                          -0.93
                                                  -0.11
                                                                                       -0.018
                                                                                              -0.047
                                                                                                             -0.055
                                                                                                                             -0.47
                                                                                                                                                   -0.071
                         latitude
                                                         -0.039
                                                                -0.076
                                                                        -0.14
                                                                                -0.094
                                                                                       -0.077
                                                                                               -0.14
                                                                                                      -0.45
                                                                                                                     -0.017
                                                                                                                                     -0.16
                                                                                                                                            -0.12
                                                                                                                                                                      0.75
                                                          -0.31
                                                                                        -0.12
             housing_median_age
                                          -0.039
                                                  -0.31
                                                                                                                            -0.019 0.00059
                                                                                                                                                    -0.37
                     total_rooms
                                                                                                                                                                      0.50
                                                         0.95
                  total bedrooms
                                          -0.076
                                                  -0.27
                                                                         0.9
                                                                                0.97
                                                                                       -0.028
                                                                                                                             -0.02
                                                                                                                                            0.68
                                                                                                                                                    -0.54
                                                         0.86
                                                                 0.9
                                                                                0.93
                                                                                              -0.022
                                                                                                             -0.073 -0.0098 -0.063 -0.015
                                                                                                                                            0.58
                      population
                                           -0.14
                                                  -0.24
                                                                                                                                                    -0.61
                                                                                                                                                                      0.25
                      households
                                          -0.094
                                                  -0.24
                                                         0.93
                                                                 0.97
                                                                        0.93
                                                                                                             -0.088
                                                                                                                     0.0067 -0.0095
                                                                                                                                            0.64
                                                                                                                                                    -0.65
                  median_income
                                   -0.018
                                          -0.077
                                                  -0.12
                                                                -0.028
                                                                                               0.69
                                                                                                              -0.24
                                                                                                                                            -0.51
                                                                                                                                                                      0.00
             median_house_value
                                   -0.047
                                           -0.14
                                                                                                              -0.49
                                                                                                                                             -0.2
                     <1H OCEAN
                                          -0.45
                                                                                                              -0.61
                                                                                                                             -0.31
                                                                                                                                     -0.34
                                                                                                                                                    -0.13
                                                                                                                                                                      -0.25
                         INLAND
                                                  -0.24
                                                                                       -0.24
                                                                                               -0.49
                                                                                                      -0.61
                                                                                                                             -0.24
                                                                                                                                     -0.26
                         ISLAND
                                          -0.017
                                                                                      -0.0092
                                                                                                      -0.014
                                                                                                             -0.011
                                   -0.47
                                                         -0.019
                                                                 -0.02
                                                                        -0.063 -0.0095
                                                                                                      -0.31
                                                                                                              -0.24
                                                                                                                                     -0.14
                                                                                                                                            -0.015
                                                                                                                                                   -0.013
                       NEAR BAY
                                                                                                                                                                      -0.50
                    NEAR OCEAN
                                           -0.16
                                                                        -0.015
                                                                                                      -0.34
                                                                                                              -0.26
                                                                                                                             -0.14
                                                                                                                                                   -0.041
                   bedroom_ratio
                                           -0.12
                                                  -0.033
                                                                 0.68
                                                                         0.58
                                                                                       -0.51
                                                                                               -0.2
                                                                                                              -0.1
                                                                                                                             -0.015
                                                                                                                                                    -0.74
                                                                                                                                                                       -0.75
                household_rooms
                                   -0.071
                                                 -0.039
                                                         -0.37
                                                                 -0.54
                                                                         -0.61
                                                                                -0.65
                                                                                                      -0.13
                                                                                                                            -0.013 -0.041
                                                          total_rooms
                                                                                                       <1H OCEAN
                                                                                                               INLAND
                                                                                                                              ΒAΥ
                                                                                                                                     NEAR OCEAN
                                    ongitude
                                                                                         median_incom
                                                                                                median house value
                                                                                                                                                     household room
In [25]: from sklearn.linear_model import LinearRegression
            x_train, y_train = train_data.drop(['median_house_value'],axis=1),train_data['median_house_value']
           reg = LinearRegression()
           reg.fit(x_train, y_train)
Out[25]: LinearRegression()
```

```
In [26]: test_data = x_test.join(y_test)
           test_data['total_rooms'] = np.log(test_data['total_rooms']+1)
          test_data['total_bedrooms'] = np.log(test_data['total_bedrooms']+1)
test_data['population'] = np.log(test_data['population']+1)
           test_data['households'] = np.log(test_data['households']+1)
           test_data = test_data.join(pd.get_dummies(test_data.ocean_proximity)).drop(['ocean_proximity'], axis=1)
           test_data['bedroom_ratio'] = test_data['total_bedrooms']/test_data['total_rooms']
           test_data['household_rooms'] = test_data['total_rooms']/ test_data['households']
In [27]: x test, y test = test data.drop(['median house value'],axis=1),test data['median house value']
In [28]: test_data
Out[28]:
                                                                                                                                                   <1H
OCEAN
                   longitude latitude housing_median_age total_rooms total_bedrooms population households median_income median_house_value
                                                                                                                                                           INLAND
             7246
                     -118.11
                               34.00
                                                     38.0
                                                             7.853216
                                                                             6.184149
                                                                                        7.358194
                                                                                                     6.131226
                                                                                                                       3.0208
                                                                                                                                         193700.0
                                                                                                                                                                 C
             9486
                     -123.18
                               39 26
                                                     25.0
                                                             8 028455
                                                                             6.347389
                                                                                        7 351800
                                                                                                     6 284134
                                                                                                                       3 7880
                                                                                                                                         134200.0
                                                                                                                                                                 ſ
            14050
                     -117.13
                               32.76
                                                     33.0
                                                             7.372746
                                                                             6.135565
                                                                                        6.678342
                                                                                                     6.054439
                                                                                                                       2.6333
                                                                                                                                         140000.0
                                                                                                                                                                 C
             4733
                     -118.36
                               34.05
                                                     50.0
                                                             8.165932
                                                                             6.700731
                                                                                        7.452982
                                                                                                     6.632002
                                                                                                                       3.0833
                                                                                                                                         338100.0
                                                                                                                                                                 C
             2394
                     -119.40
                               37.09
                                                     22.0
                                                             7.701652
                                                                             6.169611
                                                                                         6.651572
                                                                                                     5.666427
                                                                                                                       3.3269
                                                                                                                                         102700.0
                                                                                                                                                         0
            17298
                     -119.51
                                                     24.0
                                                             8.138273
                                                                             6.391917
                                                                                         7.475339
                                                                                                     6.400257
                                                                                                                       5.2039
                                                                                                                                         301300.0
                                                                                                                                                         0
                               34.40
                                                                                                                                                                 C
                                                                                                                       2.6776
                                                                                                                                                         0
            12035
                     -117.48
                               33.93
                                                     31.0
                                                             7.692570
                                                                             6.131226
                                                                                        7.355641
                                                                                                     6.111467
                                                                                                                                         122000.0
             6828
                     -118.11
                               34.08
                                                     30.0
                                                             7.762596
                                                                             6.159095
                                                                                         6.852243
                                                                                                     6.148468
                                                                                                                       3.3421
                                                                                                                                         201000.0
                                                                                                                                                                 C
            17553
                     -121.88
                               37.33
                                                     41.0
                                                             5.981414
                                                                             5.105945
                                                                                        6.309918
                                                                                                     5.220356
                                                                                                                       2.3750
                                                                                                                                         175000.0
                                                                                                                                                         1
                                                                                                                                                                 C
             1487
                     -122 03
                               37.96
                                                     20.0
                                                             7 877397
                                                                             6 539586
                                                                                        7 041412
                                                                                                     6 442540
                                                                                                                       2.1083
                                                                                                                                         162500.0
                                                                                                                                                         0
                                                                                                                                                                 ٢
           4087 rows × 16 columns
In [29]: train data
Out[29]:
                                                                                                                                                   <1H
OCEAN
                                                                                                                                                           INLAND
                   longitude latitude housing median age total rooms total bedrooms population households median income median house value
            15372
                     -117.24
                               33.38
                                                     16.0
                                                             7.934872
                                                                             6.265301
                                                                                         7.436617
                                                                                                     6.248043
                                                                                                                       3.6680
                                                                                                                                         171200.0
                     -122.01
                                                             7.878155
                                                                                        6.962243
             1654
                               37.92
                                                     16.0
                                                                             5.846439
                                                                                                     5.814131
                                                                                                                       8.1163
                                                                                                                                         365800.0
                                                                                                                                                         0
                                                                                                                                                                 C
            20506
                     -118.73
                                                             8.513988
                                                                             6.626718
                                                                                         7.828436
                                                                                                     6.587550
                                                                                                                       6.9454
                                                                                                                                         276500.0
                               34.29
                                                      8.0
                     -121.93
                                                                                        7.367077
                                                                                                     6.658011
             9831
                               36.63
                                                     28.0
                                                             8.290042
                                                                             6.748760
                                                                                                                       3.5147
                                                                                                                                         313900.0
                                                                                                                                                         0
                                                                                                                                                                 C
             6318
                     -117.97
                               34.01
                                                     33.0
                                                             7 604396
                                                                             5.945421
                                                                                        7.252054
                                                                                                     5.849325
                                                                                                                       3.7083
                                                                                                                                         165500.0
                                                                                                                                                                 ٢
            14388
                     -117.23
                               32.74
                                                     16.0
                                                             6.601230
                                                                             4.941642
                                                                                        5.703782
                                                                                                     4.905275
                                                                                                                       4.6354
                                                                                                                                         179200.0
                                                                                                                                                         0
                                                                                                                                                                 C
             6530
                     -118.06
                               34.06
                                                     28.0
                                                             7.483807
                                                                             6.406880
                                                                                         7.689371
                                                                                                     6.354370
                                                                                                                       1.9189
                                                                                                                                         165900.0
                                                                                                                                                                 C
             3902
                     -118.51
                               34.20
                                                     34.0
                                                             7.962764
                                                                             6.366470
                                                                                        7.208600
                                                                                                     6.284134
                                                                                                                       3.7049
                                                                                                                                         227500.0
                                                                                                                                                                 C
            10111
                     -117.94
                               33.94
                                                     25.0
                                                             8.086718
                                                                             6.304449
                                                                                        7.281386
                                                                                                     6.218600
                                                                                                                       5.1084
                                                                                                                                         303800.0
                                                                                                                                                                 C
                     -118.25
                                                     37.0
                                                             7.659171
                                                                             6.093570
                                                                                        7.224753
                                                                                                     6.008813
                                                                                                                                         106200.0
            7593
                               33.90
                                                                                                                       1.9605
           16346 rows × 16 columns
In [30]: reg.score(x_test, y_test)
Out[30]: 0.6815871611317289
In [31]: from sklearn.ensemble import RandomForestRegressor
           forest = RandomForestRegressor()
           forest.fit(x_train, y_train)
Out[31]: RandomForestRegressor()
In [32]: forest.score(x_test,y_test)# test preprocessing remaining
Out[32]: 0.8192411768651547
```

```
In [39]: #performing cross validation for above dataset
                                  {\bf from} \  \, {\bf sklearn.model\_selection} \  \, {\bf import} \  \, {\bf GridSearchCV}
                                 param_grid = {
                                                 "n_estimators":[30,50,100],
"max_features":[8,12,20],
                                                 "min_samples_split":[2,4,6,8]
                                  grid_search = GridSearchCV(forest, param_grid,cv=5,
                                                                                                                                scoring="neg_mean_squared_error",
return_train_score=True)
                                 grid_search.fit(x_train, y_train)
                                 {\tt C:\ProgramData\Anaconda3\lib\site-packages\klearn\model\_selection\vertex} : FitFailed {\tt Warning: Packages\klearn\model\_selection\vertex} : FitFailed {\tt Warning: Packages\klearn\model\_selection\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\needed\need
                                  60 fits failed out of a total of 180.
                                  The score on these train-test partitions for these parameters will be set to nan.
                                  If these failures are not expected, you can try to debug them by setting error_score='raise'.
                                  Below are more details about the failures:
                                  60 fits failed with the following error:
                                  Traceback (most recent call last):
                                        File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\model_selection\_validation.py", line 680, in _fit_and_score estimator.fit(X_train, y_train, **fit_params)
                                         File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\ensemble\_forest.py", line 450, in fit
                                                trees = Parallel(
                                         File "C:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line 1043, in __call__
                                                 if self.dispatch_one_batch(iterator):
                                         File "C:\ProgramData\Anaconda3\lib\site-packages\joblib\parallel.py", line 861, in dispatch_one_batch
                                                 self._dispatch(tasks)
                                         \label{linear_file_file} File \ "C:\Pr{ogramData}\ Anaconda3\ lib\ site-packages\ joblib\ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ in \ \_dispatch \ parallel.py", \ line \ 779, \ line
                                          job = self._backend.apply_async(batch, callback=cb)
                                                                                                                                                                                                                                                                        11 1 5 5 4 11 12 200 2
In [40]: best_forest = grid_search.best_estimator_
In [41]: best_forest.score(x_test, y_test)
Out[41]: 0.8176602269832993
  In [ ]:
```