Melbourne_house_predictions

October 28, 2022

[2]:
```
import pandas as pd
```

[3]:
```
home_data = pd.read_csv( r"C:\Users\sound\Downloads\melb_data.csv")
```

[4]:
```
home_data.describe()
```

[4]:

	Rooms	Price	Distance	Postcode	Bedroom2
count	13580.000000	1.358000e+04	13580.000000	13580.000000	13580.000000
mean	2.937997	1.075684e+06	10.137776	3105.301915	2.914728
std	0.955748	6.393107e+05	5.868725	90.676964	0.965921
min	1.000000	8.500000e+04	0.000000	3000.000000	0.000000
25%	2.000000	6.500000e+05	6.100000	3044.000000	2.000000
50%	3.000000	9.030000e+05	9.200000	3084.000000	3.000000
75%	3.000000	1.330000e+06	13.000000	3148.000000	3.000000
max	10.000000	9.000000e+06	48.100000	3977.000000	20.000000

	Bathroom	Car	Landsize	BuildingArea	YearBuilt
count	13580.000000	13518.000000	13580.000000	7130.000000	8205.000000
mean	1.534242	1.610075	558.416127	151.967650	1964.684217
std	0.691712	0.962634	3990.669241	541.014538	37.273762
min	0.000000	0.000000	0.000000	0.000000	1196.000000
25%	1.000000	1.000000	177.000000	93.000000	1940.000000
50%	1.000000	2.000000	440.000000	126.000000	1970.000000
75%	2.000000	2.000000	651.000000	174.000000	1999.000000
max	8.000000	10.000000	433014.000000	44515.000000	2018.000000

	Lattitude	Longtitude	Propertycount
count	13580.000000	13580.000000	13580.000000
mean	-37.809203	144.995216	7454.417378
std	0.079260	0.103916	4378.581772
min	-38.182550	144.431810	249.000000
25%	-37.856822	144.929600	4380.000000
50%	-37.802355	145.000100	6555.000000
75%	-37.756400	145.058305	10331.000000
max	-37.408530	145.526350	21650.000000

[6]:
```
home_data.head()
```

[6]:
```
        Suburb           Address  Rooms Type      Price Method SellerG  \
0  Abbotsford      85 Turner St      2    h  1480000.0      S  Biggin
1  Abbotsford   25 Bloomburg St      2    h  1035000.0      S  Biggin
2  Abbotsford      5 Charles St      3    h  1465000.0     SP  Biggin
3  Abbotsford  40 Federation La      3    h   850000.0     PI  Biggin
4  Abbotsford       55a Park St      4    h  1600000.0     VB  Nelson

         Date  Distance  Postcode  …  Bathroom  Car  Landsize  BuildingArea  \
0  3/12/2016       2.5    3067.0  …       1.0  1.0     202.0           NaN
1  4/02/2016       2.5    3067.0  …       1.0  0.0     156.0          79.0
2  4/03/2017       2.5    3067.0  …       2.0  0.0     134.0         150.0
3  4/03/2017       2.5    3067.0  …       2.0  1.0      94.0           NaN
4  4/06/2016       2.5    3067.0  …       1.0  2.0     120.0         142.0

   YearBuilt  CouncilArea Lattitude  Longtitude             Regionname  \
0        NaN        Yarra  -37.7996    144.9984  Northern Metropolitan
1     1900.0        Yarra  -37.8079    144.9934  Northern Metropolitan
2     1900.0        Yarra  -37.8093    144.9944  Northern Metropolitan
3        NaN        Yarra  -37.7969    144.9969  Northern Metropolitan
4     2014.0        Yarra  -37.8072    144.9941  Northern Metropolitan

  Propertycount
0        4019.0
1        4019.0
2        4019.0
3        4019.0
4        4019.0

[5 rows x 21 columns]
```

[7]:
```
home_data.tail()
```

[7]:
```
              Suburb        Address  Rooms Type      Price Method   SellerG  \
13575  Wheelers Hill   12 Strada Cr      4    h  1245000.0      S     Barry
13576   Williamstown  77 Merrett Dr      3    h  1031000.0     SP  Williams
13577   Williamstown    83 Power St      3    h  1170000.0      S     Raine
13578   Williamstown   96 Verdon St      4    h  2500000.0     PI   Sweeney
13579     Yarraville     6 Agnes St      4    h  1285000.0     SP   Village

             Date  Distance  Postcode  …  Bathroom  Car  Landsize  \
13575  26/08/2017      16.7    3150.0  …       2.0  2.0     652.0
13576  26/08/2017       6.8    3016.0  …       2.0  2.0     333.0
13577  26/08/2017       6.8    3016.0  …       2.0  4.0     436.0
13578  26/08/2017       6.8    3016.0  …       1.0  5.0     866.0
13579  26/08/2017       6.3    3013.0  …       1.0  1.0     362.0

       BuildingArea  YearBuilt  CouncilArea Lattitude  Longtitude  \
```

```
13575           NaN     1981.0          NaN -37.90562   145.16761
13576         133.0     1995.0          NaN -37.85927   144.87904
13577           NaN     1997.0          NaN -37.85274   144.88738
13578         157.0     1920.0          NaN -37.85908   144.89299
13579         112.0     1920.0          NaN -37.81188   144.88449

                      Regionname Propertycount
13575  South-Eastern Metropolitan        7392.0
13576        Western Metropolitan        6380.0
13577        Western Metropolitan        6380.0
13578        Western Metropolitan        6380.0
13579        Western Metropolitan        6543.0

[5 rows x 21 columns]
```

[8]:
```
home_data.mean()
```

[8]:
```
Rooms            2.937997e+00
Price            1.075684e+06
Distance         1.013778e+01
Postcode         3.105302e+03
Bedroom2         2.914728e+00
Bathroom         1.534242e+00
Car              1.610075e+00
Landsize         5.584161e+02
BuildingArea     1.519676e+02
YearBuilt        1.964684e+03
Lattitude       -3.780920e+01
Longtitude       1.449952e+02
Propertycount    7.454417e+03
dtype: float64
```

[9]:
```
home_data.std()
```

[9]:
```
Rooms               0.955748
Price          639310.724296
Distance            5.868725
Postcode           90.676964
Bedroom2            0.965921
Bathroom            0.691712
Car                 0.962634
Landsize         3990.669241
BuildingArea      541.014538
YearBuilt          37.273762
Lattitude           0.079260
Longtitude          0.103916
Propertycount    4378.581772
```

```
dtype: float64
```

[10]:
```
home_data.max()
```

[10]:
```
Suburb                 Yarraville
Address              9b Stewart St
Rooms                           10
Type                             u
Price                    9000000.0
Method                          VB
SellerG                     iTRAK
Date                     9/09/2017
Distance                      48.1
Postcode                    3977.0
Bedroom2                      20.0
Bathroom                       8.0
Car                           10.0
Landsize                  433014.0
BuildingArea               44515.0
YearBuilt                   2018.0
Lattitude                 -37.40853
Longtitude                145.52635
Regionname       Western Victoria
Propertycount              21650.0
dtype: object
```

[11]:
```
home_data.min()
```

[11]:
```
Suburb                     Abbotsford
Address                   1 Adelle Ct
Rooms                               1
Type                                h
Price                         85000.0
Method                             PI
SellerG                     @Realty
Date                       1/07/2017
Distance                         0.0
Postcode                      3000.0
Bedroom2                         0.0
Bathroom                         0.0
Car                              0.0
Landsize                         0.0
BuildingArea                     0.0
YearBuilt                     1196.0
Lattitude                   -38.18255
Longtitude                  144.43181
Regionname       Eastern Metropolitan
```

```
Propertycount                  249.0
dtype: object
```

[16]:
```
home_data_features=['Rooms','Lattitude','Longtitude','Landsize']
X=home_data[home_data_features]
```

[18]:
```
X.describe()
```

[18]:

	Rooms	Lattitude	Longtitude	Landsize
count	13580.000000	13580.000000	13580.000000	13580.000000
mean	2.937997	-37.809203	144.995216	558.416127
std	0.955748	0.079260	0.103916	3990.669241
min	1.000000	-38.182550	144.431810	0.000000
25%	2.000000	-37.856822	144.929600	177.000000
50%	3.000000	-37.802355	145.000100	440.000000
75%	3.000000	-37.756400	145.058305	651.000000
max	10.000000	-37.408530	145.526350	433014.000000

[19]:
```
X.head()
```

[19]:

	Rooms	Lattitude	Longtitude	Landsize
0	2	-37.7996	144.9984	202.0
1	2	-37.8079	144.9934	156.0
2	3	-37.8093	144.9944	134.0
3	3	-37.7969	144.9969	94.0
4	4	-37.8072	144.9941	120.0

[20]:
```
Y=home_data.Price
```

[22]:
```
Y.head()
```

[22]:
```
0    1480000.0
1    1035000.0
2    1465000.0
3     850000.0
4    1600000.0
Name: Price, dtype: float64
```

[27]:
```
from sklearn.tree import DecisionTreeRegressor

# Define model. Specify a number for random_state to ensure same results each␣
 ↪run
home_data_model = DecisionTreeRegressor(random_state=1)

# Fit model
home_data_model.fit(X, Y)
```

[27]:
```
DecisionTreeRegressor(random_state=1)
```

[26]:
```
print("Making predictions for the following 5 houses:")
print(X.head())
print("The predictions are")
print(home_data_model.predict(X.head()))
```

```
Making predictions for the following 5 houses:
   Rooms  Lattitude  Longtitude  Landsize
0      2   -37.7996    144.9984     202.0
1      2   -37.8079    144.9934     156.0
2      3   -37.8093    144.9944     134.0
3      3   -37.7969    144.9969      94.0
4      4   -37.8072    144.9941     120.0
The predictions are
[1480000. 1035000. 1465000.  850000. 1600000.]
```

[29]:
```
from sklearn.metrics import mean_absolute_error

predicted_home_prices = home_data_model.predict(X)
mean_absolute_error(Y, predicted_home_prices)
```

[29]:
```
1492.787236131566
```

[31]:
```
from sklearn.model_selection import train_test_split

# split data into training and validation data, for both features and target.

train_X, val_X, train_Y, val_Y = train_test_split(X, Y, random_state = 0)

# Define model
home_data_model = DecisionTreeRegressor()

# Fit model
home_data_model.fit(train_X, train_Y)

# get predicted prices on validation data
val_predictions = home_data_model.predict(val_X)
print(mean_absolute_error(val_Y, val_predictions))
```

```
248615.38546882672
```

[32]:
```
from sklearn.metrics import mean_absolute_error
from sklearn.tree import DecisionTreeRegressor

def get_mae(max_leaf_nodes, train_X, val_X, train_y, val_y):
    model = DecisionTreeRegressor(max_leaf_nodes=max_leaf_nodes, random_state=0)
    model.fit(train_X, train_y)
    preds_val = model.predict(val_X)
    mae = mean_absolute_error(val_y, preds_val)
```

```
    return(mae)
```

[33]:
```
# compare MAE with differing values of max_leaf_nodes
for max_leaf_nodes in [5, 50, 500, 5000]:
    my_mae = get_mae(max_leaf_nodes, train_X, val_X, train_Y, val_Y)
    print("Max leaf nodes: %d  \t\t Mean Absolute Error:  %d" %(max_leaf_nodes,␣
 ↪my_mae))
```

```
Max leaf nodes: 5               Mean Absolute Error:  354662
Max leaf nodes: 50              Mean Absolute Error:  258912
Max leaf nodes: 500             Mean Absolute Error:  227078
Max leaf nodes: 5000            Mean Absolute Error:  245018
```

[34]:
```
#Random forest
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error

forest_model = RandomForestRegressor(random_state=1)
forest_model.fit(train_X, train_Y)
melb_preds = forest_model.predict(val_X)
print(mean_absolute_error(val_Y, melb_preds))
```

```
189895.85708714963
```

[ ]:
```
```