

Linear-Regression

March 8, 2022

0.0.1 Linear Regression In Python

Importing Libraries

```
[35]: import pandas as pd
      from matplotlib import pyplot as plt
      import seaborn as sns
```

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

```
[6]:   id    price  area  bedrooms  bathrooms  stories  stories.1  guestroom  \
0    0  13300000  7420         4          2        3          1          0
1    1  12250000  8960         4          4        4          1          0
2    2  12250000  9960         3          2        2          1          0
3    3  12215000  7500         4          2        2          1          0
4    4  11410000  7420         4          1        2          1          1
```

```
      basement  hotwaterheating  airconditioning  parking  prefarea  \
0            0                0                1        2          1
1            0                0                1        3          0
2            1                0                0        2          1
3            1                0                1        3          1
4            1                0                1        2          0
```

```
      furnishingstatus
0                1.0
1                1.0
2                0.5
3                1.0
4                1.0
```

```
[5]: df = pd.read_csv('../..../datasets/housing.csv')
```

```
[11]: df.drop(df.columns[[0]], axis=1, inplace=True)
```

```
[12]: df.head()
```

```
[12]:      price  area  bedrooms  bathrooms  stories  stories.1  guestroom  \
0  13300000  7420          4           2          3          1          0
1  12250000  8960          4           4          4          1          0
2  12250000  9960          3           2          2          1          0
3  12215000  7500          4           2          2          1          0
4  11410000  7420          4           1          2          1          1

      basement  hotwaterheating  airconditioning  parking  prefarea  \
0           0              0              1          2          1
1           0              0              1          3          0
2           1              0              0          2          1
3           1              0              1          3          1
4           1              0              1          2          0

      furnishingstatus
0              1.0
1              1.0
2              0.5
3              1.0
4              1.0
```

```
[33]: #To check which area is less then 2000
```

```
[21]: df[df['area']<2000]
```

```
[21]:      price  area  bedrooms  bathrooms  stories  stories.1  guestroom  \
271  4340000  1905          5           1          2          0          0
413  3430000  1950          3           2          2          1          0
449  3150000  1650          3           1          2          0          0
527  2275000  1836          2           1          1          0          0
530  2240000  1950          3           1          1          0          0
537  1890000  1700          3           1          2          1          0

      basement  hotwaterheating  airconditioning  parking  prefarea  \
271          1              0              0          0          0
413          1              0              0          0          1
449          1              0              0          0          0
527          1              0              0          0          0
530          0              1              0          0          0
537          0              0              0          0          0

      furnishingstatus
271              0.5
413              0.0
449              0.0
527              0.5
530              0.0
```

537

0.0

```
[ ]: #To check which price is greater then 2500000
```

```
[29]: df['price']>2500000
```

```
[29]: 0      True
      1      True
      2      True
      3      True
      4      True
      ...
      540   False
      541   False
      542   False
      543   False
      544   False
      Name: price, Length: 545, dtype: bool
```

```
[32]: #To check which price is greater then 2500000 for all data frame
```

```
[30]: df[df['price']>2500000]
```

```
[30]:
```

	price	area	bedrooms	bathrooms	stories	stories.1	guestroom	\
0	13300000	7420	4	2	3	1	0	
1	12250000	8960	4	4	4	1	0	
2	12250000	9960	3	2	2	1	0	
3	12215000	7500	4	2	2	1	0	
4	11410000	7420	4	1	2	1	1	
..		
508	2590000	4400	2	1	1	1	0	
509	2590000	3600	2	2	2	1	0	
510	2520000	2880	3	1	1	0	0	
511	2520000	3180	3	1	1	0	0	
512	2520000	3000	2	1	2	1	0	

	basement	hotwaterheating	airconditioning	parking	prefarea	\
0	0	0	1	2	1	
1	0	0	1	3	0	
2	1	0	0	2	1	
3	1	0	1	3	1	
4	1	0	1	2	0	
..		
508	0	0	0	0	0	
509	1	0	0	1	0	
510	0	0	0	0	0	

```

511         0         0         0         0         0
512         0         0         0         0         0

```

```

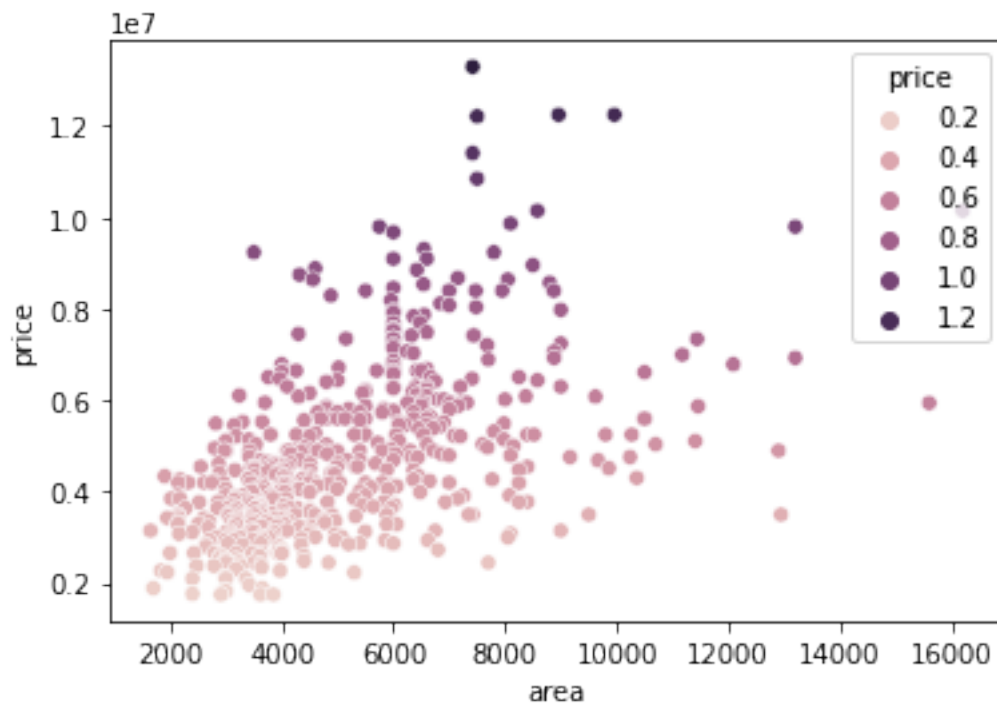
furnishingstatus
0         1.0
1         1.0
2         0.5
3         1.0
4         1.0
..        ...
508        0.0
509        1.0
510        0.0
511        0.0
512        1.0

```

```
[513 rows x 13 columns]
```

```
[67]: #visualze
```

```
[50]: sns.scatterplot(x='area',y='price',data=df,hue='price')
plt.show()
```



```
[51]: df.head()
```

```
[51]:
```

	price	area	bedrooms	bathrooms	stories	stories.1	guestroom	\
0	13300000	7420	4	2	3	1	0	
1	12250000	8960	4	4	4	1	0	
2	12250000	9960	3	2	2	1	0	
3	12215000	7500	4	2	2	1	0	
4	11410000	7420	4	1	2	1	1	

	basement	hotwaterheating	airconditioning	parking	prefarea	\
0	0	0	1	2	1	
1	0	0	1	3	0	
2	1	0	0	2	1	
3	1	0	1	3	1	
4	1	0	1	2	0	

	furnishingstatus
0	1.0
1	1.0
2	0.5
3	1.0
4	1.0

```
[74]: y = df[['price']]
```

```
[76]: x = df[['area']]
```

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

```
[80]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

```
[81]: x_train.head()
```

```
[81]:
```

	area
343	4080
86	6670
308	4046
349	4820
133	4800

```
[82]: x_test.head()
```

```
[82]:
```

	area
542	3620
129	11460
486	6000
510	2880
369	3600

```
[83]: y_train.head()
```

```
[83]:      area
343  4080
86   6670
308  4046
349  4820
133  4800
```

```
[84]: y_test.head()
```

```
[84]:      area
542  3620
129 11460
486  6000
510  2880
369  3600
```

```
[86]: from sklearn.linear_model import LinearRegression
```

```
[87]: ir = LinearRegression()
```

```
[88]: ir.fit(x_train,y_train)
```

```
[88]: LinearRegression()
```

```
[90]: y_pred=ir.predict(x_test)
```

```
[91]: y_test.head()
```

```
[91]:      area
542  3620
129 11460
486  6000
510  2880
369  3600
```

```
[92]: y_pred[0:5]
```

```
[92]: array([[ 3620.],
          [11460.],
          [ 6000.],
          [ 2880.],
          [ 3600.]])
```

```
[93]: from sklearn.metrics import mean_squared_error
```

```
[94]: mean_squared_error(y_test,y_pred)

[94]: 1.44945749105138e-24

[95]: #model 2

[96]: y = df[['price']]

[103]: x = df[['bathrooms' , 'bedrooms', 'basement']]

[105]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)

[106]: lr2 = LinearRegression()

[107]: lr2.fit(x_train,y_train)

[107]: LinearRegression()

[109]: y_pred=lr2.predict(x_test)

[110]: mean_squared_error(y_test,y_pred)

[110]: 2849450783944.8496

[ ]:
```