Ex11

December 28, 2023

[7]:
```
import pandas as pd
df = pd.read_csv ('/Users/thutranghoa/Code/Data_analysis/Data/rent.csv')
df
```

[7]:

	bathrooms	bedrooms	price	longitude	latitude	interest_level
0	1.5	3	3000	-73.9425	40.7145	2
1	1.0	2	5465	-73.9667	40.7947	1
2	1.0	1	2850	-74.0018	40.7388	3
3	1.0	1	3275	-73.9677	40.7539	1
4	1.0	4	3350	-73.9493	40.8241	1
…	…	…	…	…	…	…
49347	1.0	2	3200	-73.9790	40.7426	2
49348	1.0	1	3950	-74.0163	40.7102	1
49349	1.0	1	2595	-73.9900	40.7601	1
49350	1.0	0	3350	-74.0101	40.7066	1
49351	1.0	2	2200	-73.9172	40.8699	1

[49352 rows x 6 columns]

[8]:
```
from sklearn.model_selection import train_test_split

X = df.drop(['price'], axis=1)
y = df['price']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=44)
```

[9]:
```
X_train
```

[9]:

	bathrooms	bedrooms	longitude	latitude	interest_level
25070	2.0	3	-74.0059	40.7128	2
11403	1.0	1	-73.9560	40.7472	1
35438	1.0	1	-73.9982	40.7568	2
2903	1.0	0	-73.9705	40.7892	1
37416	1.0	0	-73.9864	40.7300	1
…	…	…	…	…	…
19183	1.0	1	-74.0055	40.7434	2
4180	2.0	2	-73.9841	40.5783	2
25773	1.0	1	-73.9677	40.7539	1

```
3491          1.0          2   -73.9809   40.7278                1
14100         1.0          3   -73.9525   40.8254                2

[39481 rows x 5 columns]
```

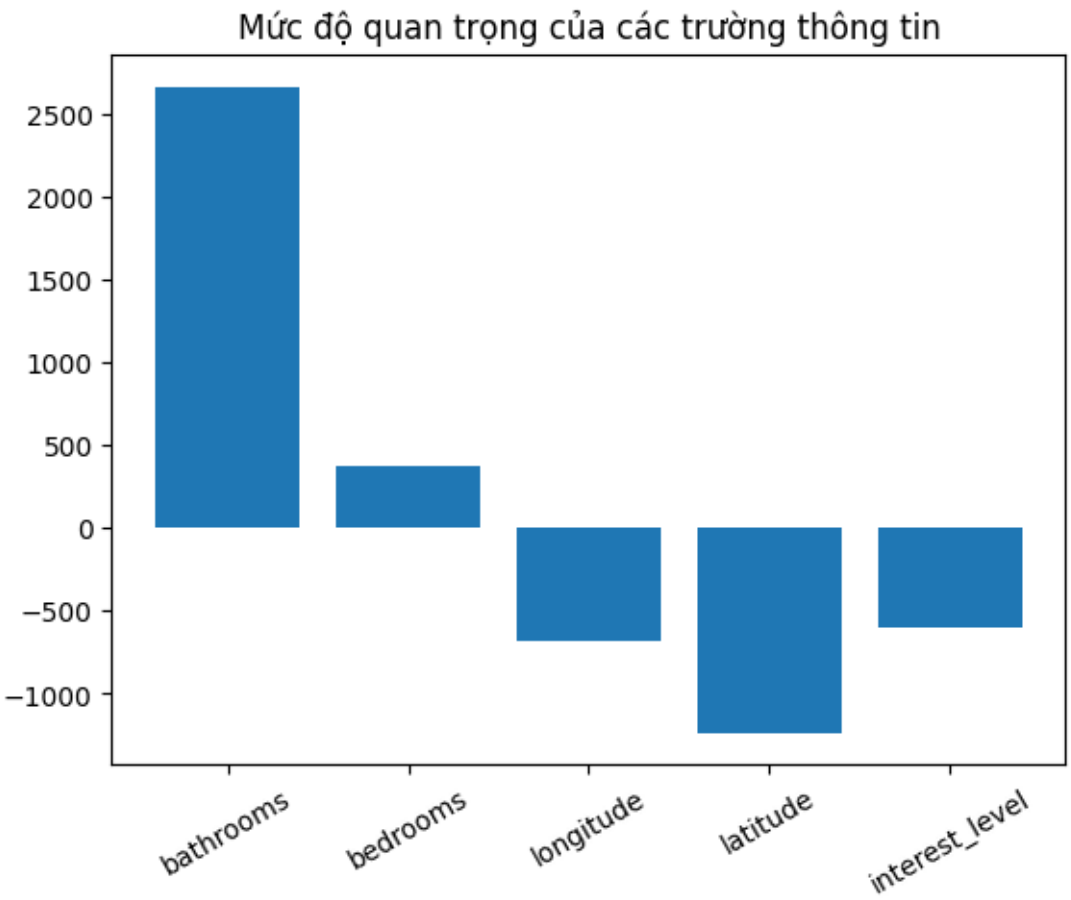
[10]:
```
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error , r2_score

regr = DecisionTreeRegressor()

regr.fit(X_train, y_train)

y_1 = regr.predict(X_test)

print ('R2_score DecisionTree = ', r2_score(y_test, y_1))
```

```
R2_score DecisionTree =  0.0027062821560310812
```

[11]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error , r2_score

LR = LinearRegression()
LR.fit(X_train, y_train)
predictions_LR = LR.predict(X_test)

print ('MSE of LinearRegression= ', mean_squared_error(y_test, predictions_LR))
print ('R2_score of Linear Regression= ', r2_score(y_test, predictions_LR))
```

```
MSE of LinearRegression=  2043420028.5015402
R2_score of Linear Regression=  0.0010253981749799301
```

[12]:
```
import matplotlib.pyplot as plt
A = list(X.columns)
importance = LR.coef_

# plot feature importance
feature = df.columns
plt.bar(A, importance)
plt.title ('Mức độ quan trọng của các trường thông tin')
plt.xticks(rotation = 30)
plt.show()
```



[13]:
```
print ("Feature bathrooms has strongest impact on price")
```

```
Feature bathrooms has strongest impact on price
```