In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn
import seaborn as sn
```

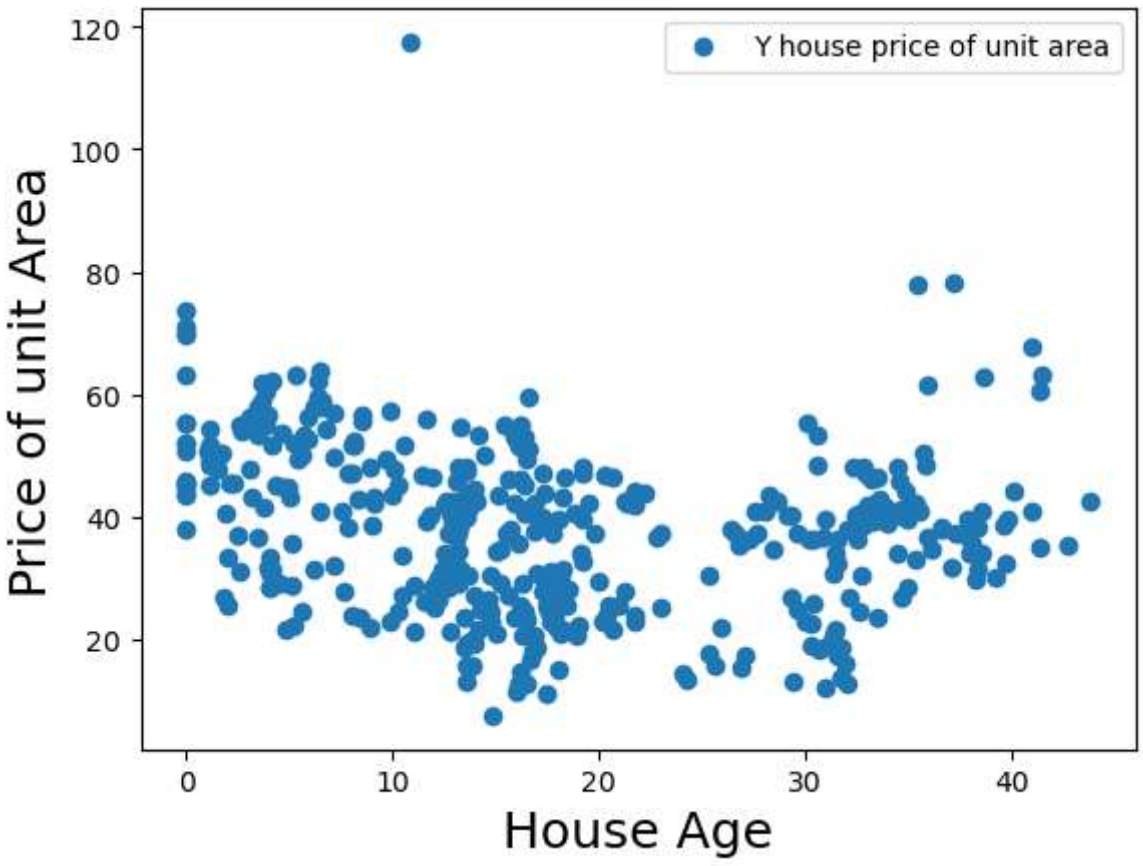
In [2]:
```
df = pd.read_csv("Real estate.csv")
```

In [3]:
```
df.head()
```

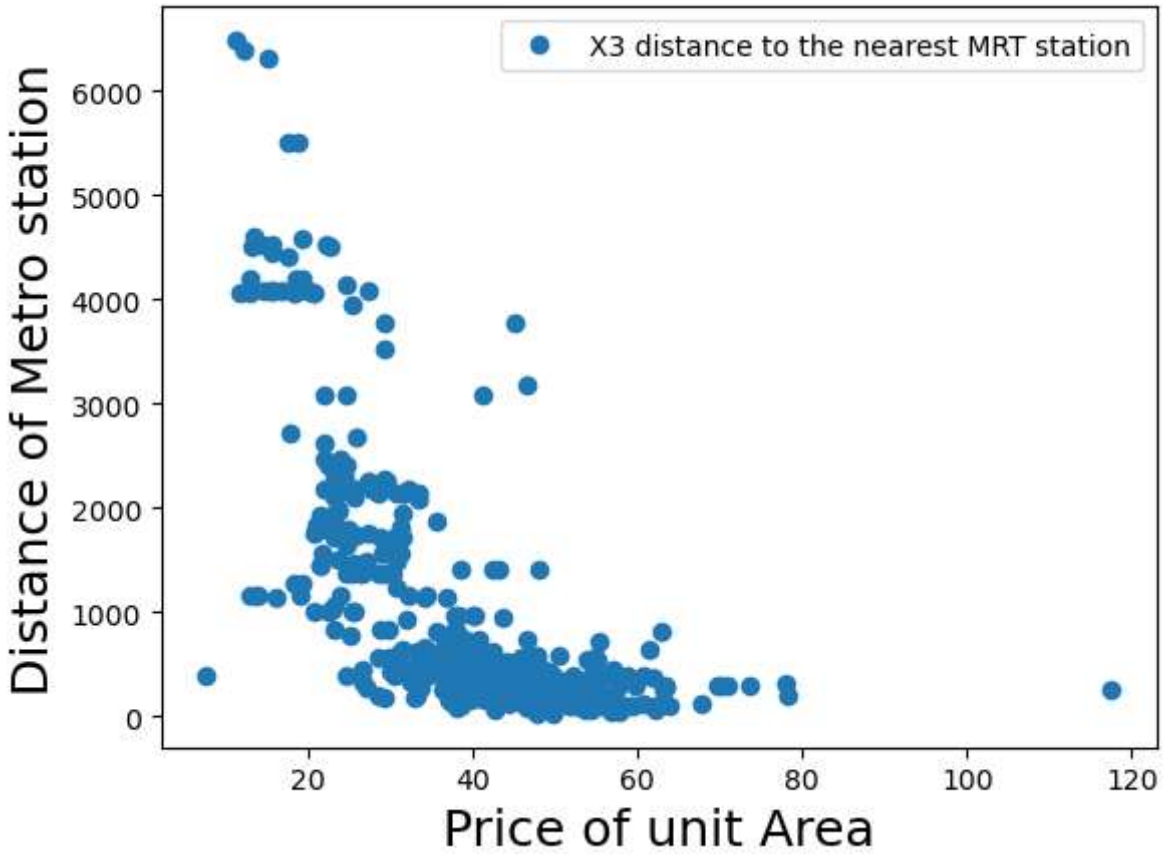
Out[3]:

	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude	Y house price of unit area
0	32.0	84.87882	10	24.98298	121.54024	37.9
1	19.5	306.59470	9	24.98034	121.53951	42.2
2	13.3	561.98450	5	24.98746	121.54391	47.3
3	13.3	561.98450	5	24.98746	121.54391	54.8
4	5.0	390.56840	5	24.97937	121.54245	43.1

In [4]:
```
df.plot(x='X2 house age',y='Y house price of unit area',style= 'o')
plt.xlabel('House Age',fontsize=18)
plt.ylabel('Price of unit Area', fontsize=18)
plt.show()
```

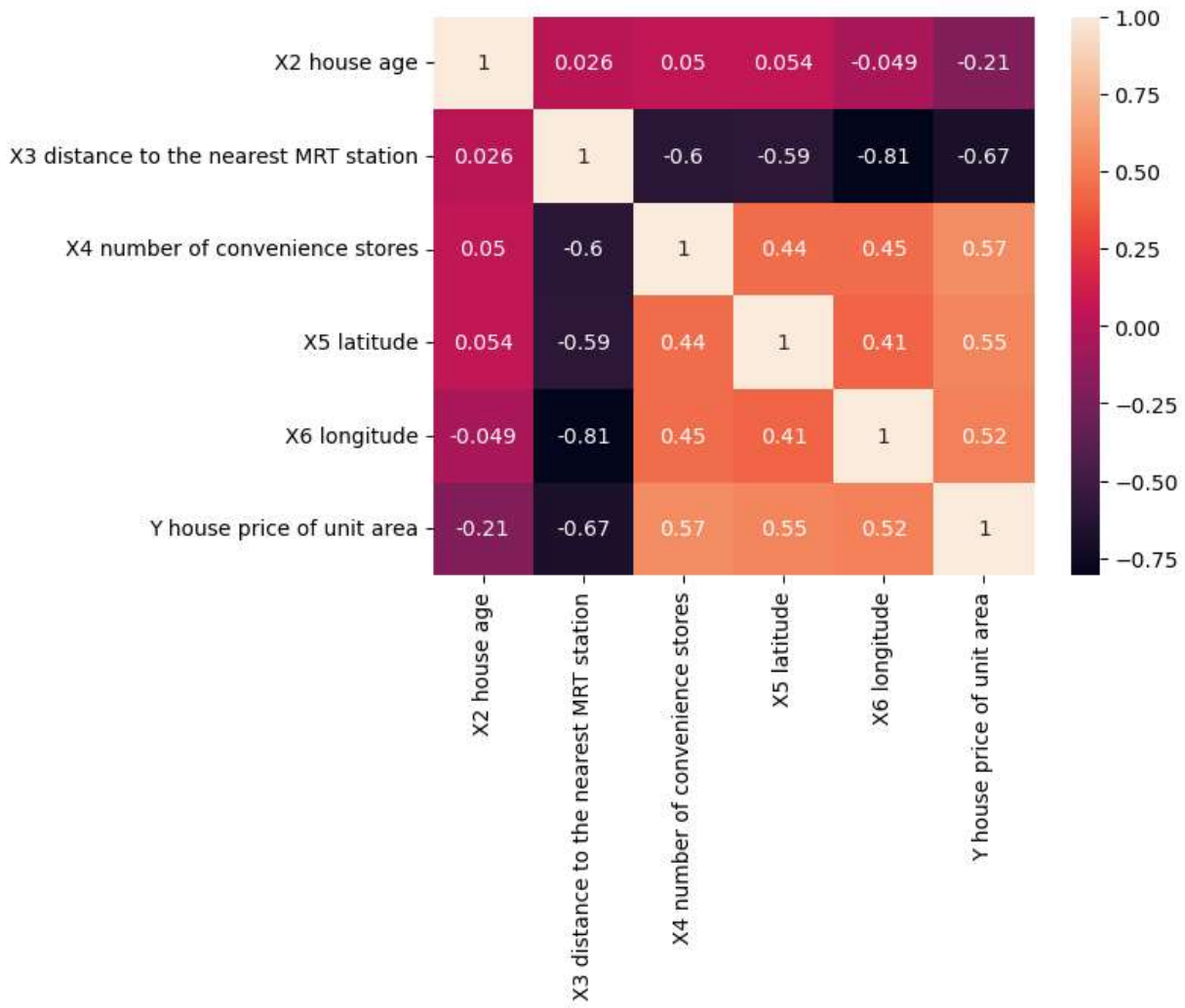


In [5]:
```
df.plot(x='Y house price of unit area',y='X3 distance to the nearest MRT station'
plt.xlabel('Price of unit Area',fontsize=18)
plt.ylabel('Distance of Metro station', fontsize=18)
plt.show()
```



In [6]:
```
sn.heatmap(df.corr(),annot=True)
```

Out[6]:
```
<AxesSubplot:>
```



In [7]:
```
X = df[['X3 distance to the nearest MRT station', 'X4 number of convenience store
y= pd.DataFrame(df['Y house price of unit area'])
```

In [8]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.25,random_sta
#The Data is divided as 75 % train and 25% test data by satyam dubey
```

In [9]:
```
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(310, 2)
(104, 2)
(310, 1)
(104, 1)
```

In [10]:
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

Out[10]:
```
LinearRegression()
```

In [11]:
```
# Fit the model to the data
c= float(regressor.intercept_)
print(c)
```

```
38.92392257991259
```

In [12]:
```
m= (regressor.coef_)
print(type(m))
print(m)
```

```
<class 'numpy.ndarray'>
[[-0.00540602  1.23950631]]
```

In [13]:
```
regressor.fit(X_train, y_train)
```

Out[13]:
```
LinearRegression()
```

In [14]:
```
y_pred = regressor.predict(X_test)
y_pred
```

Out[14]:
```
array([[31.70792727],
       [45.31443889],
       [22.28324334],
       [16.93987427],
       [38.68210172],
       [32.1197565 ],
       [30.85953565],
       [37.58481098],
       [38.26671887],
       [15.78034054],
       [16.3079356 ],
       [44.31859011],
       [33.22020989],
       [43.03728865],
       [49.59047218],
       [33.00253581],
       [38.68210172],
       [40.99796513],
       [30.8796082 ],
       [43.55735866],
       [50.87090006],
       [30.530999  ],
       [46.19740646],
       [44.95722967],
       [42.6820214 ],
       [32.51479259],
       [30.88418709],
       [44.55484871],
       [43.01003379],
       [16.85647021],
       [31.41203083],
       [39.04709478],
       [39.31016899],
       [44.55484871],
       [47.81543456],
       [16.23297236],
       [42.51165448],
       [41.63873305],
       [31.34459685],
       [16.93987427],
       [44.22995908],
       [32.72684574],
       [49.59047218],
       [31.03368516],
       [32.14441335],
       [34.20998646],
       [31.51900718],
       [41.11109961],
       [44.54709336],
       [38.06356457],
       [16.23297236],
       [48.17101657],
       [49.2040178 ],
       [16.85647021],
```

```
       [48.24701235],
       [48.68372087],
       [38.68210172],
       [37.34804006],
       [38.32955087],
       [42.46044218],
       [32.67036365],
       [39.24370876],
       [33.85180524],
       [16.23297236],
       [37.79099024],
       [48.27728606],
       [47.28913709],
       [46.8498211 ],
       [16.85647021],
       [37.76243511],
       [43.55735866],
       [14.87108254],
       [38.41644951],
       [42.98242903],
       [49.59047218],
       [45.65280165],
       [41.87319115],
       [37.34804006],
       [39.14989594],
       [43.55735866],
       [41.66365708],
       [37.92148656],
       [50.84702822],
       [39.47100684],
       [44.55484871],
       [49.59047218],
       [43.01003379],
       [33.47853989],
       [50.7246182 ],
       [16.93987427],
       [30.88418709],
       [44.77700866],
       [48.17848228],
       [32.81049104],
       [46.09125428],
       [40.97460627],
       [30.530999  ],
       [42.46044218],
       [44.95722967],
       [37.92148656],
       [21.77965641],
       [43.55735866],
       [30.27340961],
       [34.81838529]])
```

In [15]:
```
from sklearn.metrics import mean_squared_error, r2_score
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error: ", mse)
print("R^2 Score: ", r2)
```

```
Mean Squared Error:  74.99709556370581
R^2 Score:  0.5752529465522461
```

In [19]:
```
hello=y_test
hello.shape
```

Out[19]:
```
(104, 1)
```

In [20]:
```
hello['y_pred']= y_pred
```

In [21]:
```
hello.shape
```

Out[21]:
```
(104, 2)
```

In [22]:
```
hello.head()
```

Out[22]:

	Y house price of unit area	y_pred
388	27.3	31.707927
102	54.4	45.314439
187	22.0	22.283243
162	11.6	16.939874
90	45.4	38.682102