In [1]:
```
import pandas as pd
df = pd.read_csv('Housing.csv')
df
```

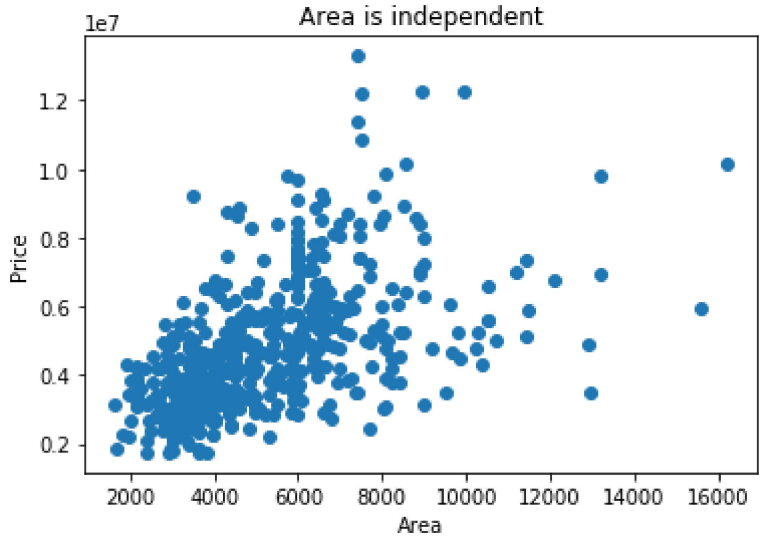
Out[1]:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwa
0	13300000	7420	4	2	3	yes	no	no	
1	12250000	8960	4	4	4	yes	no	no	
2	12250000	9960	3	2	2	yes	no	yes	
3	12215000	7500	4	2	2	yes	no	yes	
4	11410000	7420	4	1	2	yes	yes	yes	
5	10850000	7500	3	3	1	yes	no	yes	
6	10150000	8580	4	3	4	yes	no	no	
7	10150000	16200	5	3	2	yes	no	no	
8	9870000	8100	4	1	2	yes	yes	yes	
9	9800000	5750	3	2	4	yes	yes	no	
10	9800000	13200	3	1	2	yes	no	yes	

In [2]:
```
df.isnull().sum()
```

Out[2]:
```
price              0
area               0
bedrooms           0
bathrooms          0
stories            0
mainroad           0
guestroom          0
basement           0
hotwaterheating    0
airconditioning    0
parking            0
prefarea           0
furnishingstatus   0
dtype: int64
```

In [19]:
```
import matplotlib.pyplot as plt
plt.scatter(df['area'],df['price'])
plt.ylabel("Price")
plt.xlabel("Area")
plt.title("Area is independent")
plt.show()
```

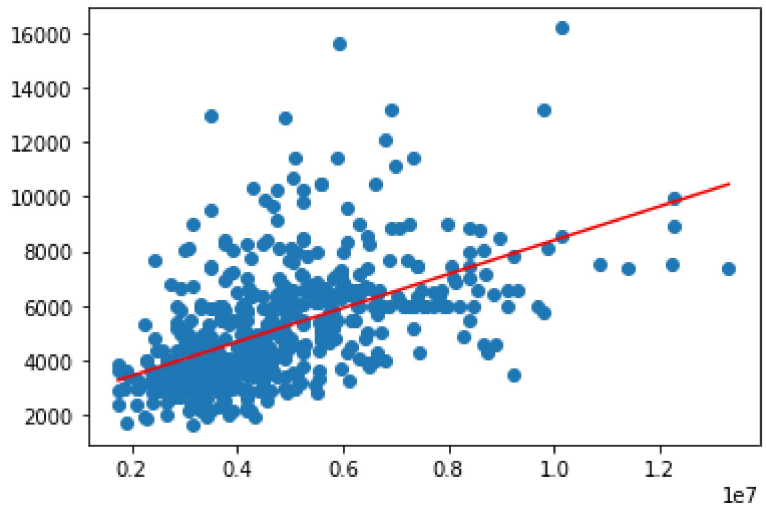


In [13]:
```
from scipy import stats
x = df['price']
y = df['area']
slope, intercept, r, p, std_err = stats.linregress(x, y)

def myfunc(x):
    return slope * x + intercept

mymodel = list(map(myfunc, x))

plt.scatter(x, y)
plt.plot(x, mymodel , color="r")
plt.show()
```



In [22]:
```
from sklearn.model_selection import train_test_split
x = df['area']
y = df['price']
x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.4, rando
```

In [24]:
```
x_train
```

Out[24]:

284	7770
33	5960
427	2145
8	8100
173	5300
95	4100
468	2835
195	4410
74	4040
359	3600
367	3630
79	6000
477	4960
330	4050
274	6450
538	3649
411	2145
10	13200
415	4785
263	3968
174	3800
335	3816
188	5720
36	7482
209	6720
77	6500
30	7475
58	7680
29	5500
296	4600
...	
0	7420
462	2160
316	5900
269	3900
56	11440
480	3480
35	7000
194	8150
103	6350
214	4350
15	6000
187	6100
417	3640
403	12944
57	9000
297	3640
12	6550
192	6600
459	3500
49	7440
524	3264
429	4775
6	8580
39	6000
347	3350
460	8100

```
237     5010
31      7000
488     5200
40      6550
Name: area, Length: 327, dtype: int64
```

In [27]:
```
import numpy as np
x_train = np.array(x_train).reshape(-1,1)
x_train
```

Out[27]:
```
array([[ 7770],
       [ 5960],
       [ 2145],
       [ 8100],
       [ 5300],
       [ 4100],
       [ 2835],
       [ 4410],
       [ 4040],
       [ 3600],
       [ 3630],
       [ 6000],
       [ 4960],
       [ 4050],
       [ 6450],
       [ 3649],
       [ 2145],
       [13200],
       [ 4785],
```

In [28]:
```
import numpy as np
x_test = np.array(x_test).reshape(-1,1)
x_test
```

Out[28]:
```
array([[ 4900],
       [ 4040],
       [ 3584],
       [ 1905],
       [ 6420],
       [11175],
       [ 7424],
       [ 8250],
       [ 3240],
       [ 6930],
       [ 4500],
       [ 3000],
       [ 6525],
       [ 6000],
       [ 3060],
       [ 3480],
       [ 9000],
       [ 3000],
       [ 3000],
```

In [29]:
```
from sklearn.linear_model import LinearRegression
lr =LinearRegression()
lr.fit(x_train,y_train)
```

Out[29]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=Fals
e)
```

In [30]:
```
c = lr.intercept_
c
```

Out[30]:
```
2313418.6659722454
```

In [31]:
```
m = lr.coef_
m
```

Out[31]:
```
array([481.71534686])
```

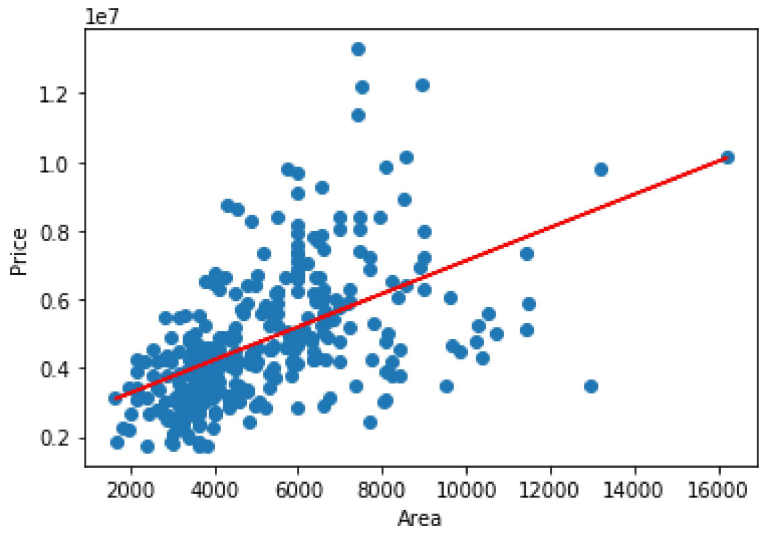
In [32]:
```
y_pred_train = m*x_train + c
y_pred_train
```

```
       [ 4481137.72684001],
       [ 3655959.33766968],
       [ 3252763.59234828],
       [ 3903079.31060861],
       [ 4240280.05341026],
       [ 3975336.61263753],
       [ 5545728.64339951],
       [ 5796702.33911331],
       [ 4408880.42481108],
       [ 6407999.11427802],
       [ 6889714.46113753],
       [ 5131453.44510034],
       [ 4394428.9644053 ],
       [ 4336623.12278216],
       [ 4066862.52854084],
       [ 4418514.73174827],
       [ 6287570.27756315],
       [ 3670410.79807547],
       [ 6215312.97553422],
       [ 4876144.3112648 ],
```

In [33]:
```
# same as above
y_pred_train1 = lr.predict(x_train)
y_pred_train1
```

Out[33]:
```
array([ 6056346.91107059,  5184442.13325488,  3346698.08498588,
        6215312.97553422,  4866510.00432761,  4288451.58809621,
        3679081.67431894,  4437783.34562266,  4259548.66728464,
        4047593.91466646,  4062045.37507224,  5203710.74712927,
        4702726.78639538,  4264365.82075323,  5420482.65321604,
        4071197.96666257,  3346698.08498588,  8672061.24451769,
        4618426.60069497,  4224865.16231075,  4143936.98403836,
        4151644.42958811,  5068830.4500086 ,  5917612.89117505,
        5550545.79686811,  5444568.42055902,  5914240.88374703,
        6012992.52985323,  4962853.07369951,  4529309.26152596,
        4673823.86558381,  3541792.80046398,  3468572.06774133,
        5196485.01692637,  4962853.07369951,  4481137.72684001,
        4485954.8803086 ,  4721995.40026976,  5203710.74712927,
        4697909.63292679,  4240280.05341026,  4018690.99385489,
        3816370.5481739 ,  4365526.04359373,  4124668.37016398,
        3816370.5481739 ,  3960885.15223175,  4168022.75138133,
        4432966.19215406,  5492739.95524497,  3758564.70655076,
        6591050.94608463,  4134302.67710117,  4664189.55864662,
        3975336.61263753,  6143055.6735053 ,  5121819.13816315,
```

In [36]:
```
import matplotlib.pyplot as plt
plt.scatter(x_train,y_train)
plt.plot(x_train,y_pred_train,color="r")
plt.ylabel("Price")
plt.xlabel("Area")
plt.show()
```



In [40]:
```
import numpy as np

input_data = float(input("Enter the Area: "))

# Reshape the input_data to be a 2D array
input_data = np.array(input_data).reshape(1, -1)

# Make prediction
predicted_price = lr.predict(input_data)

# Display the predicted price
print("Predicted Price:", predicted_price)
```

```
Enter the Area: 5000
Predicted Price: [4721995.40026976]
```

In [ ]:
```

```