```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error

data=pd.read_csv("//home//alanjohn//Downloads//u3.csv")
data
```

```
        UR        EE
0     5.48  16635535
1     5.83  16545652
2     5.79  15881197
3    20.51  11336911
4    17.43  12988845
..     ...       ...
262   7.29  30726310
263   6.83  35372506
264  14.87  33298644
265   9.35  35707239
266   9.98  33962549

[267 rows x 2 columns]
```

```
data.tail()
```

```
        UR        EE
262   7.29  30726310
263   6.83  35372506
264  14.87  33298644
265   9.35  35707239
266   9.98  33962549
```

```
data.isnull()
```

```
        UR     EE
0    False  False
1    False  False
2    False  False
3    False  False
4    False  False
..     ...    ...
262  False  False
263  False  False
264  False  False
265  False  False
266  False  False
```

```
[267 rows x 2 columns]
```

```
data.isnull().sum()
```

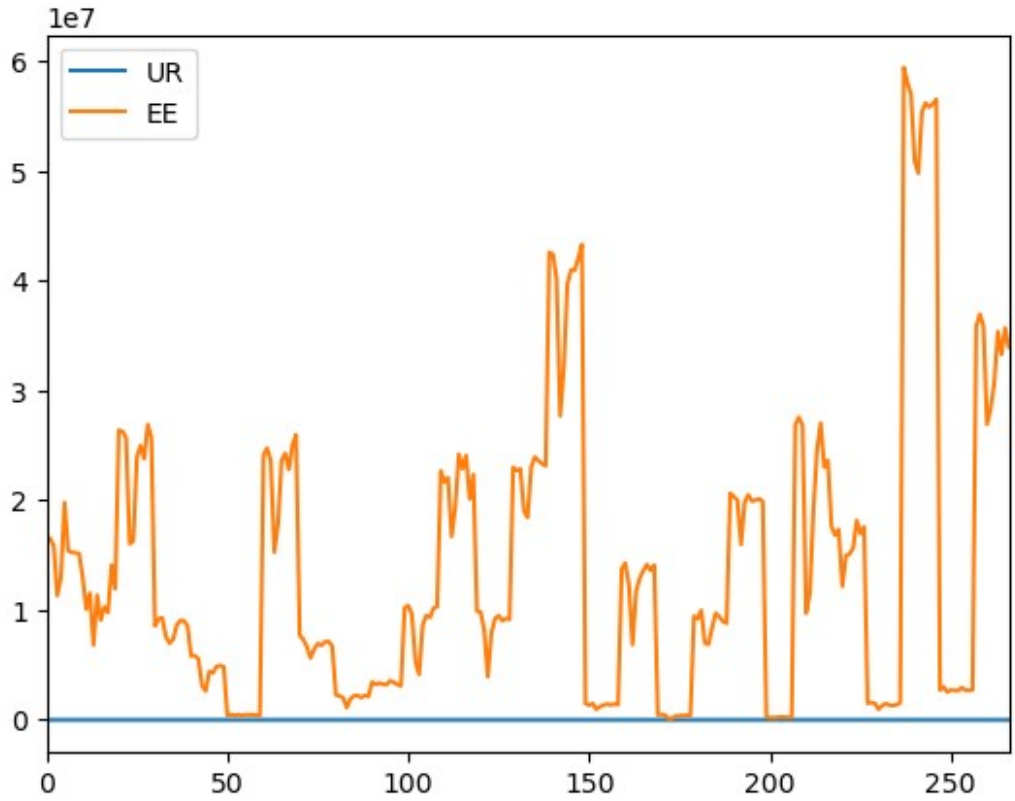
```
UR    0
EE    0
dtype: int64
```

```
data.info
```

```
<bound method DataFrame.info of         UR        EE
0     5.48  16635535
1     5.83  16545652
2     5.79  15881197
3    20.51  11336911
4    17.43  12988845
..     ...       ...
262   7.29  30726310
263   6.83  35372506
264  14.87  33298644
265   9.35  35707239
266   9.98  33962549

[267 rows x 2 columns]>
```

```
data.plot()
```

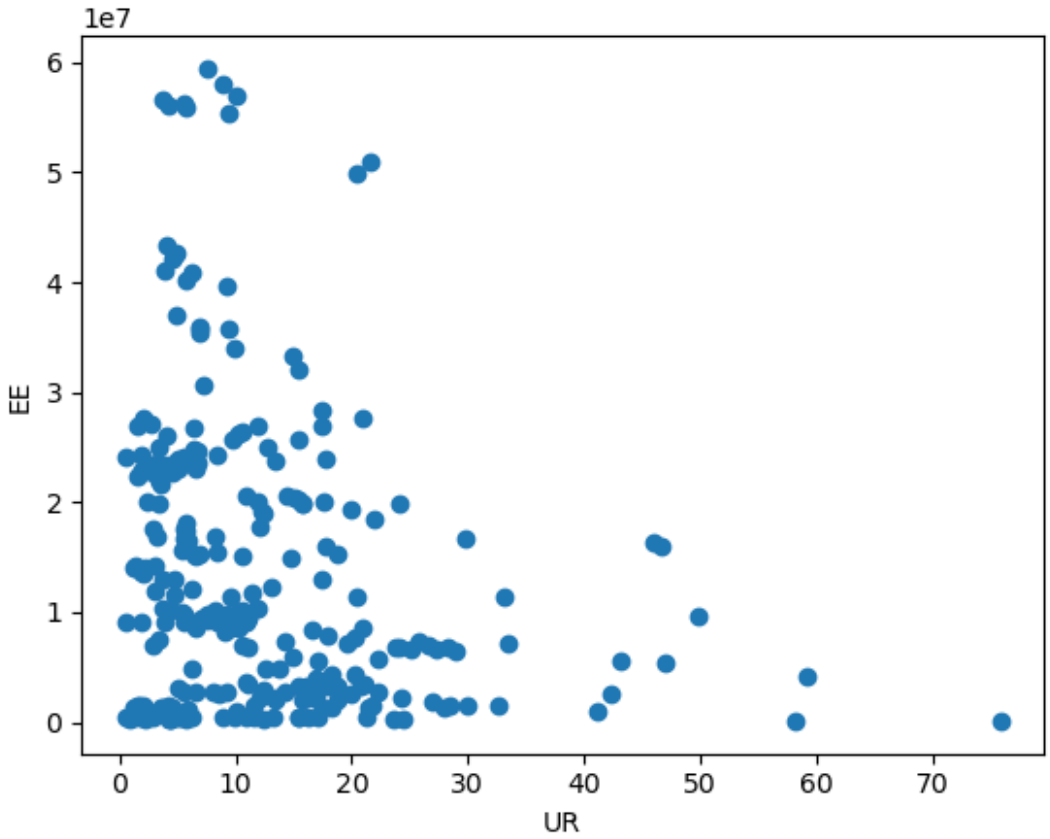
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd72dac77c0>
```




```
data.describe()
```

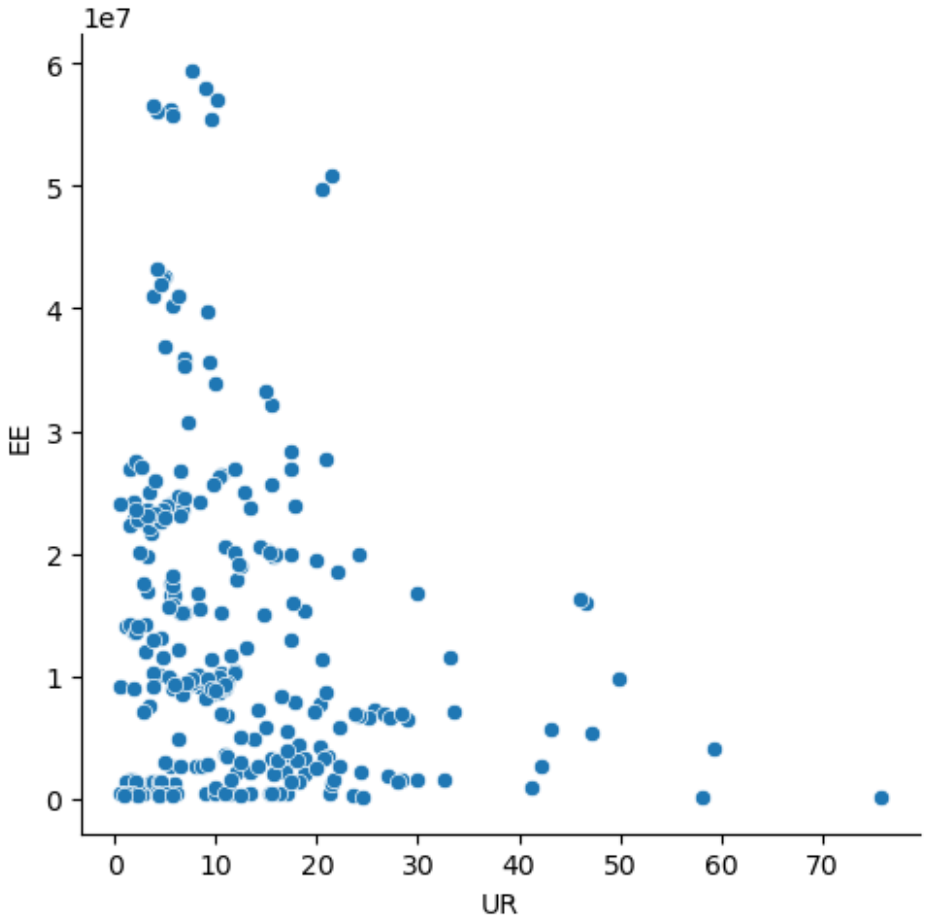
	UR	EE
count	267.000000	2.670000e+02
mean	12.236929	1.396211e+07
std	10.803283	1.336632e+07
min	0.500000	1.175420e+05
25%	4.845000	2.838930e+06
50%	9.650000	9.732417e+06
75%	16.755000	2.187869e+07
max	75.850000	5.943376e+07

```
plt.scatter(data['UR'],data['EE'])
plt.xlabel('UR')
plt.ylabel('EE')
plt.show()
```




```
sns.relplot(x='UR',y='EE',data=data)
```

```
<seaborn.axisgrid.FacetGrid at 0x7fd72da62490>
```




```
x=data.drop('UR',axis=1)
y=data['UR']

x.shape, y.shape
```

```
((267, 1), (267,))
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=0,test
_size=0.5)
x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

```
((133, 1), (134, 1), (133,), (134,))
```

```
lr=LinearRegression()
lr.fit(x_train,y_train)

LinearRegression()

pred=lr.predict(x_test)
pred
```

```
array([13.32888285, 13.96672421, 10.15565442, 12.26185362,
9.92230041,
```

```
       14.69177706, 14.36168744, 14.3743044 , 14.93810405,
14.18555544,
       14.48810727, 14.90268456, 12.42918992, 14.7319799 ,
10.1940606 ,
       13.35258816,  7.67703592, 12.78692116, 14.80067869,
14.88544347,
       14.92573197, 14.92944121, 12.30504903, 12.25624996,
10.70887098,
       13.32902159, 11.11180567, 14.52030961, 14.75615307,
4.90300641,
       10.76606645, 14.7323263 ,  4.98006652, 13.25602887,
10.73462815,
       13.13774346, 13.24159371, 13.36069551,  5.02905971,
10.27119409,
       13.47817859, 12.28169404, 13.76137389,  8.57815558,
13.72220794,
       11.44526098, 10.39232745, 11.35202427, 11.79583947,
12.91679359,
       10.68142063, 14.93420799, 12.24964219, 14.73770983,
13.1585107 ,
       13.19594412, 14.4345327 , 14.50372838, 14.40080994,
13.82598481,
       11.32037688, 11.40620776, 14.77110186, 14.94186529,
13.40521846,
       14.35757784, 13.29093989, 14.23427027, 12.46001753,
14.92901484,
       14.90202524, 14.59428009, 13.66835536, 14.706993  ,
11.40809916,
       14.48468887, 14.27204635, 13.61893224, 13.31061129,
14.73076011,
       10.81540996,  4.96317967, 10.29185625, 14.69032432,
10.93187218,
       13.14050736,  7.67081515, 13.73443095, 11.04022224,
10.87299244,
       13.3639864 , 14.70340896, 12.65927228, 10.86657844,
10.56325188,
       10.90886127, 14.89921931, 10.64903699, 10.52200896,
11.2963152 ,
       12.52487938, 10.99093607,  6.10295941, 13.93864643,
13.75600158,
        7.38615784, 14.72651302, 12.95347757, 13.27983961,
12.0730897 ,
       13.44943259,  8.38932577, 12.46234847, 10.67571136,
14.89082825,
        4.63211835, 14.73268605, 14.20832716, 12.84440747,
10.59619422,
       11.8418237 , 14.90518861, 11.97588997, 14.5651314 ,
14.8874102 ,
       12.14415166, 13.37316056, 13.43686781,  7.90056215,
```

```
14.5754924 ,
       12.64804163, 14.71464887, 14.49951335, 13.92955936])

p=lr.predict(x_train)
p

array([12.94795904, 14.89833078, 12.80470936, 13.5616714 ,
10.77539341,
       10.87170782, 14.89625256, 10.92769562, 12.45175416,
11.73377374,
       14.41701806, 11.57145249, 10.8505664 , 14.71257938,
8.92390834,
       11.99504881, 11.42684427, 14.41925906, 14.89756888,
14.39436067,
       13.74979025, 13.43730023, 11.51784407, 12.00980596,
14.40222492,
       14.33872373, 10.34731509, 14.11008058, 14.08947915,
14.5807949 ,
        9.04214815,  7.81360285, 12.11475863,  4.8252104 ,
9.25262323,
       13.70228346, 13.11581094, 10.89996851, 11.88405525,
8.61318344,
       12.22422109, 10.51829937, 14.63458818, 14.8845797 ,
13.50828383,
       14.44920419, 14.93127027, 14.74263155, 11.95176168,
10.18500025,
       14.39845335, 14.70645033, 14.48961682, 13.17585775,
13.20858031,
       14.02235617, 13.66659843, 14.88830995, 14.89099317,
10.75929091,
       12.0258139 , 13.22220673, 11.83638176, 13.30089982,
13.98343421,
       13.27672647, 11.58125052, 14.7288721 , 13.2305262 ,
13.3657834 ,
       11.6877359 , 14.88339482, 12.67438421,  4.3875499 ,
13.34244782,
       13.72743475, 10.18121283, 13.74849994, 10.65991216,
14.92532591,
       12.88330253, 14.70854993, 14.93318304, 13.33492623,
12.18008817,
       13.32676954, 13.23923392,  5.90470948,  9.50027437,
13.31583365,
       12.54892718, 10.04116659, 14.70395875,  7.48481619,
10.39664934,
       14.89922715, 14.49253923, 13.15630959, 14.61246509,
8.67279855,
       14.90370701, 13.77955782, 11.38396089, 10.06230035,
10.89668225,
        7.25774724, 10.17481397, 13.78970867, 13.70560516,
10.70463671,
       12.54786251, 14.58425266, 14.48856801, 14.90716709,
```

```
8.59296047,
       13.45038095, 14.79952871, 11.44148496, 14.89686077,
7.41831352,
       14.57406335, 10.7602578 , 13.60240782, 14.61106132,
5.1093986 ,
       12.9274139 , 12.27303013, 11.42264384, 14.49274546,
12.12581742,
       11.38795348, 14.10876105, 14.95161898])
```

```
y_train
```

```
14      9.55
50      8.89
220     6.25
123    17.88
62      6.66
       ...
251     8.01
192    17.70
117     2.41
47     13.79
172    75.85
Name: UR, Length: 133, dtype: float64
```

```
y_test
```

```
180    10.97
73     43.22
214     2.65
8       6.40
261    17.41
       ...
86     24.31
10      4.66
155     2.10
252     8.61
41     14.84
Name: UR, Length: 134, dtype: float64
```

```
pred
```

```
array([13.32888285, 13.96672421, 10.15565442, 12.26185362,
9.92230041,
       14.69177706, 14.36168744, 14.3743044 , 14.93810405,
14.18555544,
       14.48810727, 14.90268456, 12.42918992, 14.7319799 ,
10.1940606 ,
       13.35258816,  7.67703592, 12.78692116, 14.80067869,
14.88544347,
       14.92573197, 14.92944121, 12.30504903, 12.25624996,
10.70887098,
```

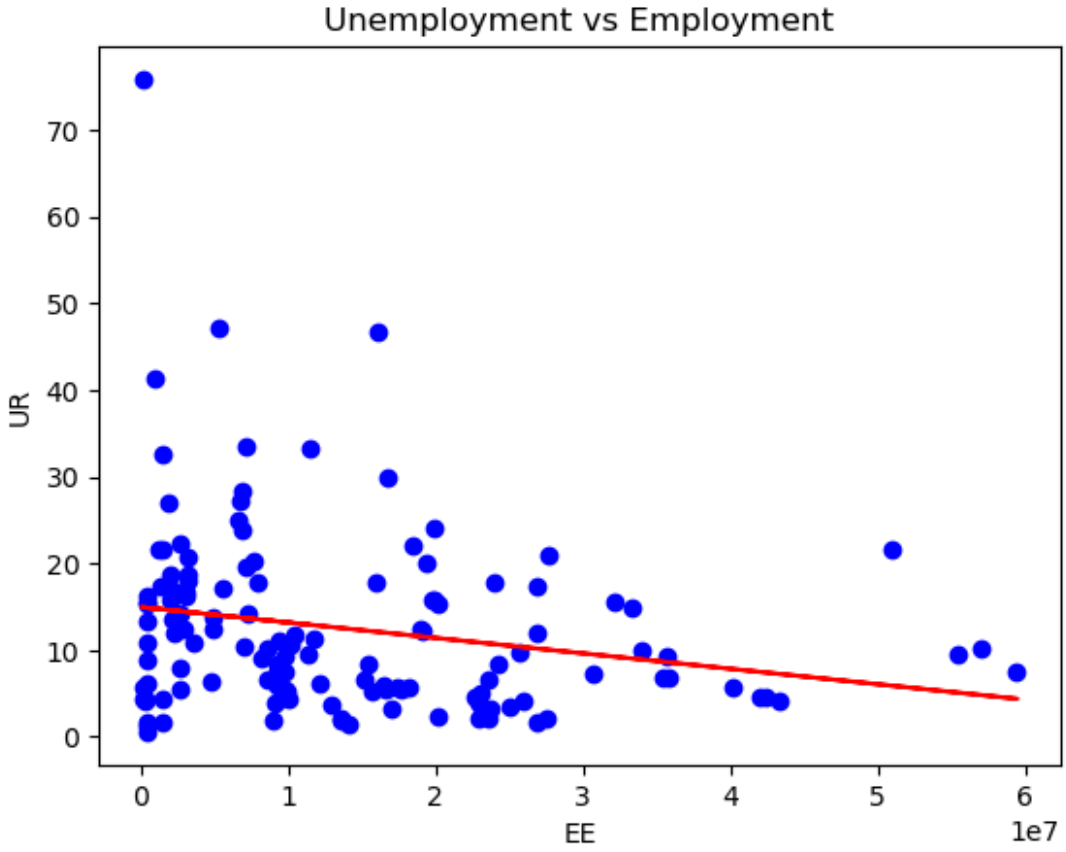
```
       13.32902159, 11.11180567, 14.52030961, 14.75615307,
4.90300641,
       10.76606645, 14.7323263 ,  4.98006652, 13.25602887,
10.73462815,
       13.13774346, 13.24159371, 13.36069551,  5.02905971,
10.27119409,
       13.47817859, 12.28169404, 13.76137389,  8.57815558,
13.72220794,
       11.44526098, 10.39232745, 11.35202427, 11.79583947,
12.91679359,
       10.68142063, 14.93420799, 12.24964219, 14.73770983,
13.1585107 ,
       13.19594412, 14.4345327 , 14.50372838, 14.40080994,
13.82598481,
       11.32037688, 11.40620776, 14.77110186, 14.94186529,
13.40521846,
       14.35757784, 13.29093989, 14.23427027, 12.46001753,
14.92901484,
       14.90202524, 14.59428009, 13.66835536, 14.706993  ,
11.40809916,
       14.48468887, 14.27204635, 13.61893224, 13.31061129,
14.73076011,
       10.81540996,  4.96317967, 10.29185625, 14.69032432,
10.93187218,
       13.14050736,  7.67081515, 13.73443095, 11.04022224,
10.87299244,
       13.3639864 , 14.70340896, 12.65927228, 10.86657844,
10.56325188,
       10.90886127, 14.89921931, 10.64903699, 10.52200896,
11.2963152 ,
       12.52487938, 10.99093607,  6.10295941, 13.93864643,
13.75600158,
        7.38615784, 14.72651302, 12.95347757, 13.27983961,
12.0730897 ,
       13.44943259,  8.38932577, 12.46234847, 10.67571136,
14.89082825,
        4.63211835, 14.73268605, 14.20832716, 12.84440747,
10.59619422,
       11.8418237 , 14.90518861, 11.97588997, 14.5651314 ,
14.8874102 ,
       12.14415166, 13.37316056, 13.43686781,  7.90056215,
14.5754924 ,
       12.64804163, 14.71464887, 14.49951335, 13.92955936])
```

```
diff=y_test-pred

pd.DataFrame(np.c_[y_test,pred,diff],columns=['actual','predicted','di
fference'])
```

```
     actual  predicted  difference
0     10.97  13.328883   -2.358883
```

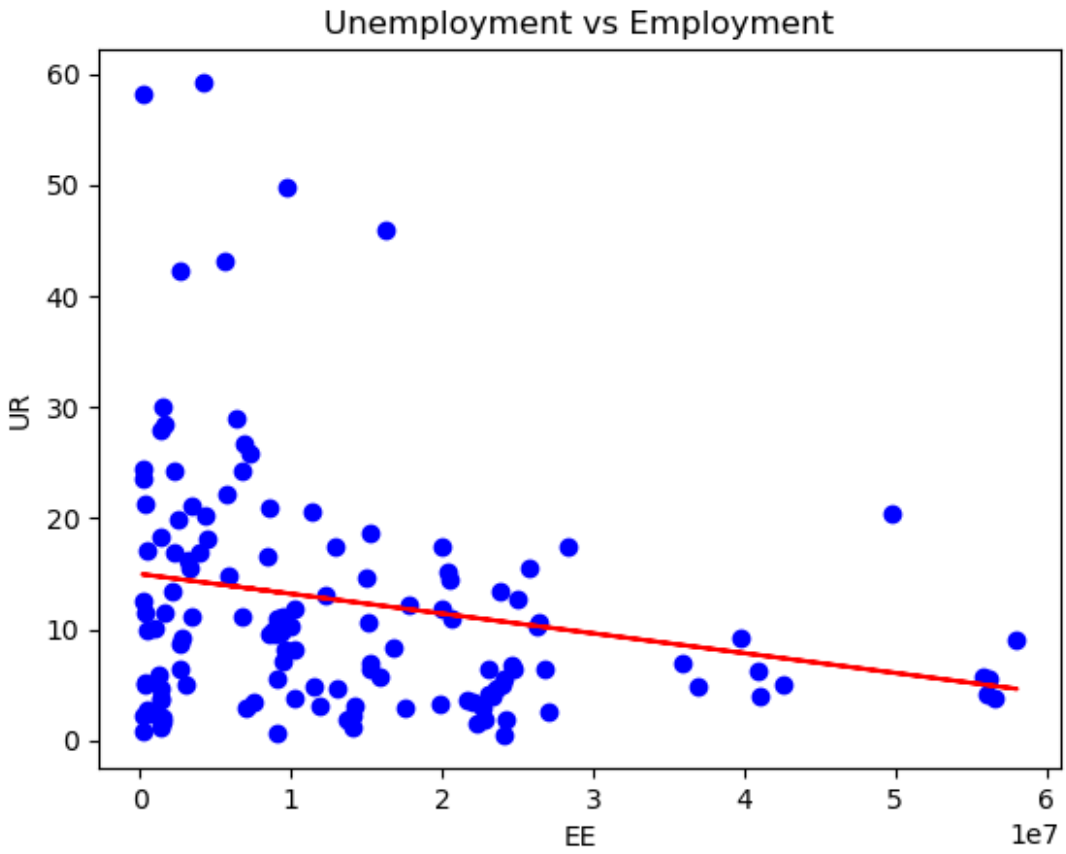
```
1     43.22  13.966724   29.253276
2      2.65  10.155654   -7.505654
3      6.40  12.261854   -5.861854
4     17.41   9.922300    7.487700
..      ...        ...         ...
129   24.31  14.575492    9.734508
130    4.66  12.648042   -7.988042
131    2.10  14.714649  -12.614649
132    8.61  14.499513   -5.889513
133   14.84  13.929559    0.910441

[134 rows x 3 columns]
```

```
plt.scatter(x_train,y_train,color='blue')
plt.plot(x_train,lr.predict(x_train),color='red')
plt.ylabel('UR')
plt.xlabel('EE')
plt.title('Unemployment vs Employment')
plt.show()
```




```
plt.scatter(x_test,y_test,color='blue')
plt.plot(x_test,lr.predict(x_test),color='red')
plt.ylabel('UR')
plt.xlabel('EE')
```

```
plt.title('Unemployment vs Employment')
plt.show()
```




```
lr.score(x_test,y_test)
```

```
0.0680437327236556
```

```
lr.score(x_train,y_train)
```

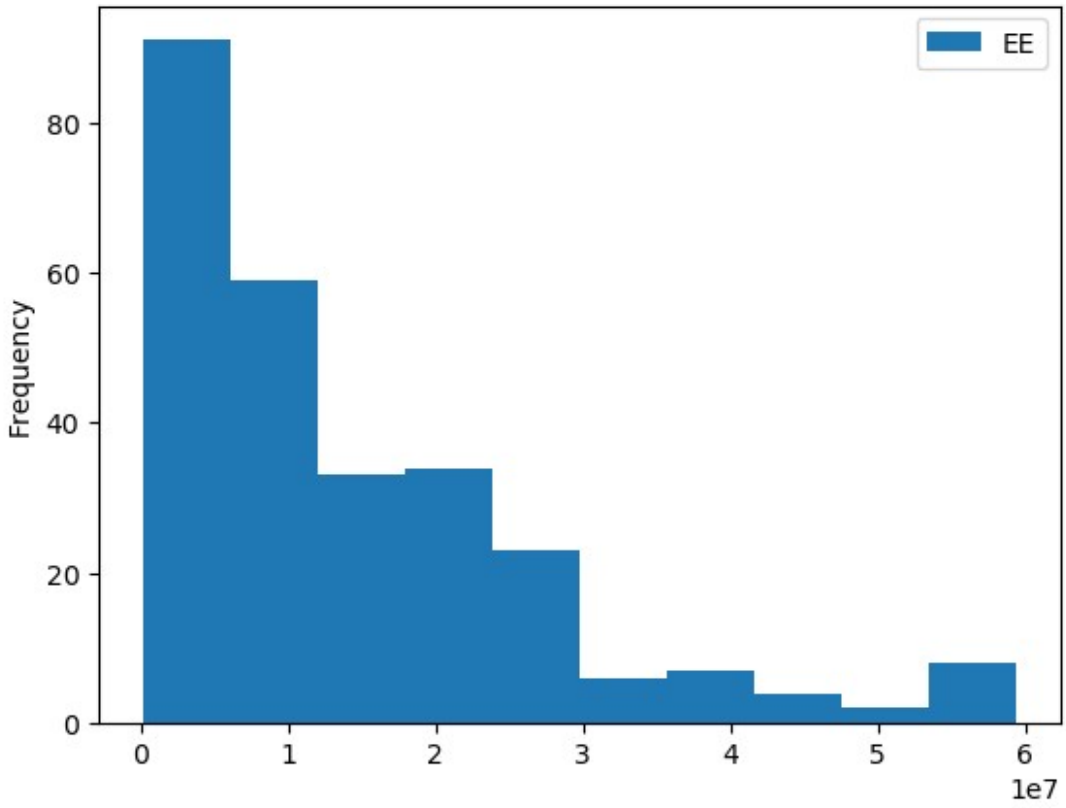
```
0.04827648559941844
```

```
rm=np.sqrt(mean_squared_error(y_test,pred))
r2=r2_score(y_test,pred)
rm,r2
```

```
(10.573640312731808, 0.0680437327236556)
```

```
data.plot(kind='hist',x='UR',y='EE')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd72d8add30>
```


EE
Frequency
80
60
40
20
0
0
1
2
3
4
5
6
1e7