In [1]:

```
import numpy as np
import pandas as pd

data=pd.read_csv('occupancy2.csv')
print(data.head())
print(data.info())
```

```
   index             date  Temperature   Humidity       Light          CO2  \
0      1  2/11/2015 14:48      21.7600  31.133333  437.333333  1029.666667
1      2  2/11/2015 14:49      21.7900  31.000000  437.333333  1000.000000
2      3  2/11/2015 14:50      21.7675  31.122500  434.000000  1003.750000
3      4  2/11/2015 14:51      21.7675  31.122500  439.000000  1009.500000
4      5  2/11/2015 14:51      21.7900  31.133333  437.333333  1005.666667

   HumidityRatio  Occupancy
0       0.005021          1
1       0.005009          1
2       0.005022          1
3       0.005022          1
4       0.005030          1
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9751 entries, 0 to 9750
Data columns (total 8 columns):
index            9751 non-null int64
date             9751 non-null object
Temperature      9751 non-null float64
Humidity         9751 non-null float64
Light            9751 non-null float64
CO2              9751 non-null float64
HumidityRatio    9751 non-null float64
Occupancy        9751 non-null int64
dtypes: float64(5), int64(2), object(1)
memory usage: 609.5+ KB
None
```

In [2]:

```
data
data['date']=pd.to_datetime(data['date'],format='%m/%d/%Y %H:%M')
data.set_index('date', inplace=True)  # set the date time as index
data=data.drop(columns=['index'])
print(data.head())

```

```
                     Temperature    Humidity       Light          CO2  \
date
2015-02-11 14:48:00      21.7600  31.133333  437.333333  1029.666667
2015-02-11 14:49:00      21.7900  31.000000  437.333333  1000.000000
2015-02-11 14:50:00      21.7675  31.122500  434.000000  1003.750000
2015-02-11 14:51:00      21.7675  31.122500  439.000000  1009.500000
2015-02-11 14:51:00      21.7900  31.133333  437.333333  1005.666667

                     HumidityRatio  Occupancy
date
2015-02-11 14:48:00       0.005021          1
2015-02-11 14:49:00       0.005009          1
2015-02-11 14:50:00       0.005022          1
2015-02-11 14:51:00       0.005022          1
2015-02-11 14:51:00       0.005030          1
```

In [3]:

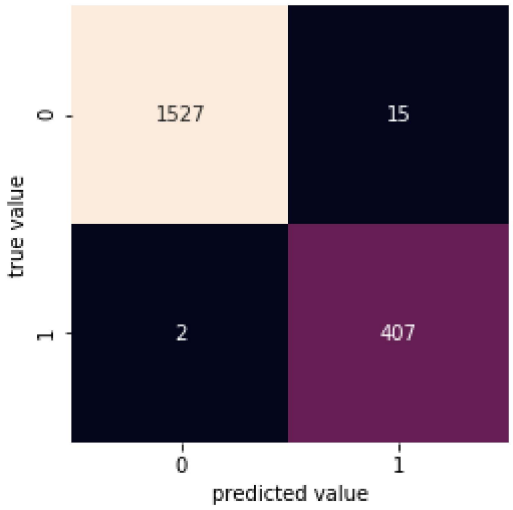
```
import numpy as np
import pandas as pd
from sklearn import datasets
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import SGDClassifier
from sklearn.metrics import accuracy_score
classifier=SGDClassifier(max_iter=1000)
#prepare data which is step 3
Y = data['Occupancy']  #price is the target, Y
X = np.array([data['Temperature'],data['Light']])  # X
X=X.T
xtrain,xtest,ytrain,ytest=train_test_split(X,Y,random_state=43,test_size=0.2)
classifier.fit(xtrain,ytrain)
ypredict=classifier.predict(xtest)
print(accuracy_score(ytest,ypredict))  #percentage of classification on the t
```

```
0.99128651973347
```

In [4]:

```
from sklearn.metrics import confusion_matrix
import  seaborn as sns
mat=confusion_matrix(ytest,ypredict)
sns.heatmap(mat,square=True,annot=True,cbar=False,fmt='d')
plt.xlabel('predicted value')
plt.ylabel('true value')
```

Out[4]: Text(91.68,0.5,'true value')



In [ ]:

```
1
```