In [2]:

```python
import numpy as np
import pandas as pd
from pandas import *
import matplotlib.pyplot as plt
```

In [4]:

```python
data = pd.read_csv('Downloads/creditcard.csv')
data.head()
```

Out[4]:

| V7	V8	V9	...	V21	V22	V23	V24	V25	
0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.18
-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.12
0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.13
0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.22
0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.50

In [5]:

```python
count_classes = pd.value_counts(data['Class'], sort = False)

count_classes.plot (kind='bar')
plt.title ("Operaciones fraudulentas sobre no fraudulentas")
plt.xlabel ("Fraudulentas")
plt.ylabel ("Frecuencia")
```
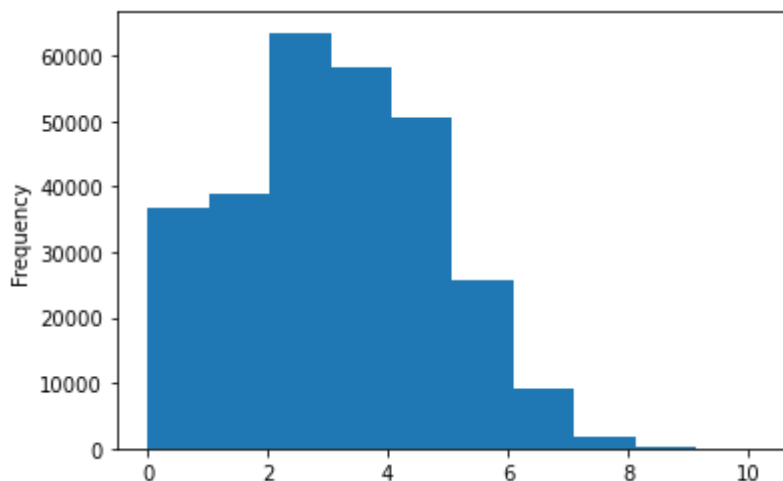
Out[5]:

```
Text(0, 0.5, 'Frecuencia')
```

In [6]:
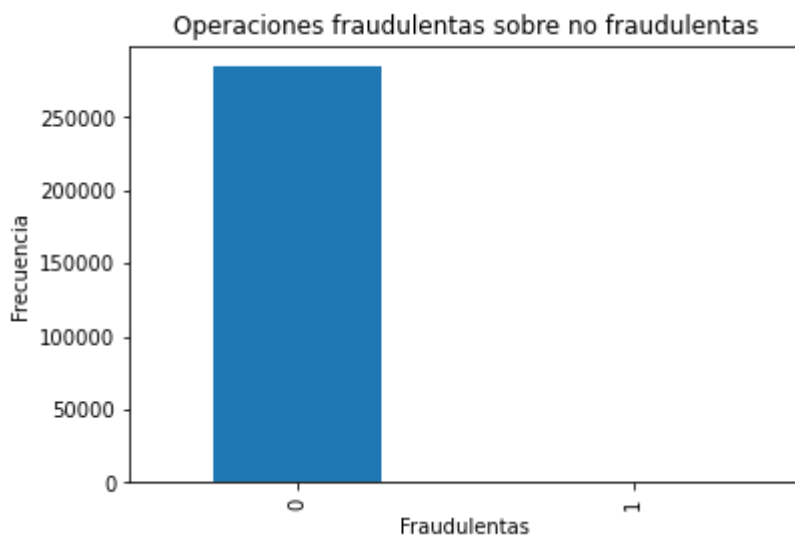
```python
data['logAmount'] = np.log(data['Amount']+1)

data['logAmount'].sort_values().plot.hist()
```

Out[6]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x20b6ea1b1c0>
```


In [7]:

```python
from sklearn.preprocessing import StandardScaler
data['normAmount'] = StandardScaler().fit_transform(data['Amount'].values.reshape (-1,1
))
data = data.drop (['Time', 'Amount','logAmount'], axis = 1);
```

In [8]:

```python
X = data.iloc[:, data.columns != 'Class']
y = data.iloc[:, data.columns == 'Class']
len(y[y.Class ==1]);
```

In [9]:

```python
number_records_fraud = len (data[data.Class==1])

fraud_indices = np.array (data[data.Class==1].index)
normal_indices = np.array (data[data.Class==0].index)
```

In [10]:

```python
random_normal_indices = np.random.choice (normal_indices, number_records_fraud, replace
= False )

under_sample_indices = np.concatenate ([fraud_indices, random_normal_indices])
```

In [11]:

```python
under_sample_data = data.iloc[under_sample_indices,:]

X_undersample = under_sample_data.iloc [:, under_sample_data.columns != 'Class'];
y_undersample = under_sample_data.iloc [:, under_sample_data.columns == 'Class'];
```

In [12]:

```python
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split (X,y, test_size = 0.3, random_state = 0)
X_train_under, X_test_under, y_train_under, y_test_under = train_test_split (X_undersample,y_undersample, test_size = 0.3, random_state = 0)
```

In [13]:

```python
from sklearn.metrics import recall_score
from sklearn.neural_network import MLPClassifier
MLPC = MLPClassifier(hidden_layer_sizes=(200,), max_iter=10000)
MLPC.fit(X_train_under, y_train_under)
y_pred = MLPC.predict(X_test)
recall_acc = recall_score (y_test,y_pred)
recall_acc
```

```
C:\Users\kurno\anaconda3\lib\site-packages\sklearn\utils\validation.py:73:
DataConversionWarning: A column-vector y was passed when a 1d array was ex
pected. Please change the shape of y to (n_samples, ), for example using r
avel().
  return f(**kwargs)
```

Out[13]:

0.9591836734693877

In [ ]: