

```
In [81]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
import seaborn as sns
from sklearn.datasets import load_iris
from sklearn import tree
from sklearn.tree import export_graphviz
import graphviz
import pydotplus
import io
import random
from scipy import misc
```

```
-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-81-265b04ccbc0c> in <module>()
      10 from sklearn import tree
      11 from sklearn.tree import export_graphviz
----> 12 import graphviz
      13 import pydotplus
      14 import io
```

ModuleNotFoundError: No module named 'graphviz'

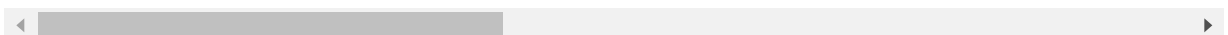
```
In [82]: cc = pd.read_csv("E:/creditcard.csv")  
df=pd.DataFrame(cc)  
df
```

Out[82]:

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | |
|--------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-------|
| 0 | 0.0 | -1.359807 | -0.072781 | 2.536347 | 1.378155 | -0.338321 | 0.462388 | 0.23 |
| 1 | 0.0 | 1.191857 | 0.266151 | 0.166480 | 0.448154 | 0.060018 | -0.082361 | -0.07 |
| 2 | 1.0 | -1.358354 | -1.340163 | 1.773209 | 0.379780 | -0.503198 | 1.800499 | 0.79 |
| 3 | 1.0 | -0.966272 | -0.185226 | 1.792993 | -0.863291 | -0.010309 | 1.247203 | 0.23 |
| 4 | 2.0 | -1.158233 | 0.877737 | 1.548718 | 0.403034 | -0.407193 | 0.095921 | 0.59 |
| 5 | 2.0 | -0.425966 | 0.960523 | 1.141109 | -0.168252 | 0.420987 | -0.029728 | 0.47 |
| 6 | 4.0 | 1.229658 | 0.141004 | 0.045371 | 1.202613 | 0.191881 | 0.272708 | -0.00 |
| 7 | 7.0 | -0.644269 | 1.417964 | 1.074380 | -0.492199 | 0.948934 | 0.428118 | 1.12 |
| 8 | 7.0 | -0.894286 | 0.286157 | -0.113192 | -0.271526 | 2.669599 | 3.721818 | 0.37 |
| 9 | 9.0 | -0.338262 | 1.119593 | 1.044367 | -0.222187 | 0.499361 | -0.246761 | 0.65 |
| 10 | 10.0 | 1.449044 | -1.176339 | 0.913860 | -1.375667 | -1.971383 | -0.629152 | -1.42 |
| 11 | 10.0 | 0.384978 | 0.616109 | -0.874300 | -0.094019 | 2.924584 | 3.317027 | 0.47 |
| 12 | 10.0 | 1.249999 | -1.221637 | 0.383930 | -1.234899 | -1.485419 | -0.753230 | -0.68 |
| 13 | 11.0 | 1.069374 | 0.287722 | 0.828613 | 2.712520 | -0.178398 | 0.337544 | -0.09 |
| 14 | 12.0 | -2.791855 | -0.327771 | 1.641750 | 1.767473 | -0.136588 | 0.807596 | -0.42 |
| 15 | 12.0 | -0.752417 | 0.345485 | 2.057323 | -1.468643 | -1.158394 | -0.077850 | -0.60 |
| 16 | 12.0 | 1.103215 | -0.040296 | 1.267332 | 1.289091 | -0.735997 | 0.288069 | -0.58 |
| 17 | 13.0 | -0.436905 | 0.918966 | 0.924591 | -0.727219 | 0.915679 | -0.127867 | 0.70 |
| 18 | 14.0 | -5.401258 | -5.450148 | 1.186305 | 1.736239 | 3.049106 | -1.763406 | -1.55 |
| 19 | 15.0 | 1.492936 | -1.029346 | 0.454795 | -1.438026 | -1.555434 | -0.720961 | -1.08 |
| 20 | 16.0 | 0.694885 | -1.361819 | 1.029221 | 0.834159 | -1.191209 | 1.309109 | -0.87 |
| 21 | 17.0 | 0.962496 | 0.328461 | -0.171479 | 2.109204 | 1.129566 | 1.696038 | 0.10 |
| 22 | 18.0 | 1.166616 | 0.502120 | -0.067300 | 2.261569 | 0.428804 | 0.089474 | 0.24 |
| 23 | 18.0 | 0.247491 | 0.277666 | 1.185471 | -0.092603 | -1.314394 | -0.150116 | -0.92 |
| 24 | 22.0 | -1.946525 | -0.044901 | -0.405570 | -1.013057 | 2.941968 | 2.955053 | -0.06 |
| 25 | 22.0 | -2.074295 | -0.121482 | 1.322021 | 0.410008 | 0.295198 | -0.959537 | 0.54 |
| 26 | 23.0 | 1.173285 | 0.353498 | 0.283905 | 1.133563 | -0.172577 | -0.916054 | 0.36 |
| 27 | 23.0 | 1.322707 | -0.174041 | 0.434555 | 0.576038 | -0.836758 | -0.831083 | -0.26 |
| 28 | 23.0 | -0.414289 | 0.905437 | 1.727453 | 1.473471 | 0.007443 | -0.200331 | 0.74 |
| 29 | 23.0 | 1.059387 | -0.175319 | 1.266130 | 1.186110 | -0.786002 | 0.578435 | -0.76 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 284777 | 172764.0 | 2.079137 | -0.028723 | -1.343392 | 0.358000 | -0.045791 | -1.345452 | 0.22 |

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | |
|--------|----------|------------|-----------|-----------|-----------|-----------|-----------|-------|
| 284778 | 172764.0 | -0.764523 | 0.588379 | -0.907599 | -0.418847 | 0.901528 | -0.760802 | 0.75 |
| 284779 | 172766.0 | 1.975178 | -0.616244 | -2.628295 | -0.406246 | 2.327804 | 3.664740 | -0.53 |
| 284780 | 172766.0 | -1.727503 | 1.108356 | 2.219561 | 1.148583 | -0.884199 | 0.793083 | -0.52 |
| 284781 | 172766.0 | -1.139015 | -0.155510 | 1.894478 | -1.138957 | 1.451777 | 0.093598 | 0.19 |
| 284782 | 172767.0 | -0.268061 | 2.540315 | -1.400915 | 4.846661 | 0.639105 | 0.186479 | -0.04 |
| 284783 | 172768.0 | -1.796092 | 1.929178 | -2.828417 | -1.689844 | 2.199572 | 3.123732 | -0.27 |
| 284784 | 172768.0 | -0.669662 | 0.923769 | -1.543167 | -1.560729 | 2.833960 | 3.240843 | 0.18 |
| 284785 | 172768.0 | 0.032887 | 0.545338 | -1.185844 | -1.729828 | 2.932315 | 3.401529 | 0.33 |
| 284786 | 172768.0 | -2.076175 | 2.142238 | -2.522704 | -1.888063 | 1.982785 | 3.732950 | -1.27 |
| 284787 | 172769.0 | -1.029719 | -1.110670 | -0.636179 | -0.840816 | 2.424360 | -2.956733 | 0.28 |
| 284788 | 172770.0 | 2.007418 | -0.280235 | -0.208113 | 0.335261 | -0.715798 | -0.751373 | -0.45 |
| 284789 | 172770.0 | -0.446951 | 1.302212 | -0.168583 | 0.981577 | 0.578957 | -0.605641 | 1.25 |
| 284790 | 172771.0 | -0.515513 | 0.971950 | -1.014580 | -0.677037 | 0.912430 | -0.316187 | 0.39 |
| 284791 | 172774.0 | -0.863506 | 0.874701 | 0.420358 | -0.530365 | 0.356561 | -1.046238 | 0.75 |
| 284792 | 172774.0 | -0.724123 | 1.485216 | -1.132218 | -0.607190 | 0.709499 | -0.482638 | 0.54 |
| 284793 | 172775.0 | 1.971002 | -0.699067 | -1.697541 | -0.617643 | 1.718797 | 3.911336 | -1.25 |
| 284794 | 172777.0 | -1.266580 | -0.400461 | 0.956221 | -0.723919 | 1.531993 | -1.788600 | 0.31 |
| 284795 | 172778.0 | -12.516732 | 10.187818 | -8.476671 | -2.510473 | -4.586669 | -1.394465 | -3.63 |
| 284796 | 172780.0 | 1.884849 | -0.143540 | -0.999943 | 1.506772 | -0.035300 | -0.613638 | 0.19 |
| 284797 | 172782.0 | -0.241923 | 0.712247 | 0.399806 | -0.463406 | 0.244531 | -1.343668 | 0.92 |
| 284798 | 172782.0 | 0.219529 | 0.881246 | -0.635891 | 0.960928 | -0.152971 | -1.014307 | 0.42 |
| 284799 | 172783.0 | -1.775135 | -0.004235 | 1.189786 | 0.331096 | 1.196063 | 5.519980 | -1.57 |
| 284800 | 172784.0 | 2.039560 | -0.175233 | -1.196825 | 0.234580 | -0.008713 | -0.726571 | 0.01 |
| 284801 | 172785.0 | 0.120316 | 0.931005 | -0.546012 | -0.745097 | 1.130314 | -0.235973 | 0.81 |
| 284802 | 172786.0 | -11.881118 | 10.071785 | -9.834783 | -2.066656 | -5.364473 | -2.606837 | -4.97 |
| 284803 | 172787.0 | -0.732789 | -0.055080 | 2.035030 | -0.738589 | 0.868229 | 1.058415 | 0.02 |
| 284804 | 172788.0 | 1.919565 | -0.301254 | -3.249640 | -0.557828 | 2.630515 | 3.031260 | -0.29 |
| 284805 | 172788.0 | -0.240440 | 0.530483 | 0.702510 | 0.689799 | -0.377961 | 0.623708 | -0.68 |
| 284806 | 172792.0 | -0.533413 | -0.189733 | 0.703337 | -0.506271 | -0.012546 | -0.649617 | 1.57 |

284807 rows × 31 columns



In [83]: `df.shape`

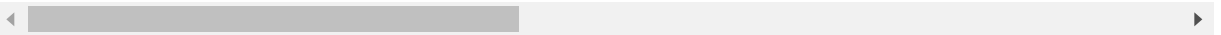
Out[83]: (284807, 31)

In [84]: `df.head()`

Out[84]:

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | V7 | |
|---|------|-----------|-----------|----------|-----------|-----------|-----------|-----------|--------|
| 0 | 0.0 | -1.359807 | -0.072781 | 2.536347 | 1.378155 | -0.338321 | 0.462388 | 0.239599 | 0.098 |
| 1 | 0.0 | 1.191857 | 0.266151 | 0.166480 | 0.448154 | 0.060018 | -0.082361 | -0.078803 | 0.085 |
| 2 | 1.0 | -1.358354 | -1.340163 | 1.773209 | 0.379780 | -0.503198 | 1.800499 | 0.791461 | 0.247 |
| 3 | 1.0 | -0.966272 | -0.185226 | 1.792993 | -0.863291 | -0.010309 | 1.247203 | 0.237609 | 0.377 |
| 4 | 2.0 | -1.158233 | 0.877737 | 1.548718 | 0.403034 | -0.407193 | 0.095921 | 0.592941 | -0.270 |

5 rows × 31 columns



In [85]: `df.isnull().any()`

Out[85]:

| | |
|--------|-------|
| Time | False |
| V1 | False |
| V2 | False |
| V3 | False |
| V4 | False |
| V5 | False |
| V6 | False |
| V7 | False |
| V8 | False |
| V9 | False |
| V10 | False |
| V11 | False |
| V12 | False |
| V13 | False |
| V14 | False |
| V15 | False |
| V16 | False |
| V17 | False |
| V18 | False |
| V19 | False |
| V20 | False |
| V21 | False |
| V22 | False |
| V23 | False |
| V24 | False |
| V25 | False |
| V26 | False |
| V27 | False |
| V28 | False |
| Amount | False |
| Class | False |
| dtype: | bool |

In [69]: `df.duplicated()`

```
Out[69]: 0      False
          1      False
          2      False
          3      False
          4      False
          5      False
          6      False
          7      False
          8      False
          9      False
         10      False
         11      False
         12      False
         13      False
         14      False
         15      False
         16      False
         17      False
         18      False
         19      False
         20      False
         21      False
         22      False
         23      False
         24      False
         25      False
         26      False
         27      False
         28      False
         29      False
          ...
        284777  False
        284778  False
        284779  False
        284780  False
        284781  False
        284782  False
        284783  False
        284784  False
        284785  False
        284786  False
        284787  False
        284788  False
        284789  False
        284790  False
        284791  False
        284792  False
        284793  False
        284794  False
        284795  False
        284796  False
        284797  False
        284798  False
        284799  False
        284800  False
        284801  False
        284802  False
```

```
284803    False
284804    False
284805    False
284806    False
Length: 284807, dtype: bool
```

```
In [70]: np.dtype(df.Class)
np.dtype(df.Amount)
np.dtype(df.Time)
```

```
Out[70]: dtype('float64')
```

```
In [71]: df.dtypes
```

```
Out[71]: Time      float64
V1      float64
V2      float64
V3      float64
V4      float64
V5      float64
V6      float64
V7      float64
V8      float64
V9      float64
V10     float64
V11     float64
V12     float64
V13     float64
V14     float64
V15     float64
V16     float64
V17     float64
V18     float64
V19     float64
V20     float64
V21     float64
V22     float64
V23     float64
V24     float64
V25     float64
V26     float64
V27     float64
V28     float64
Amount   float64
Class    int64
dtype: object
```

```
In [72]: df.columns
```

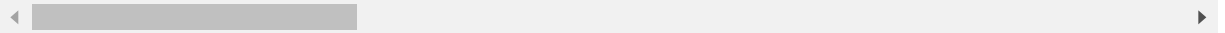
```
Out[72]: Index(['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
               'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
               'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount',
               'Class'],
              dtype='object')
```


In [73]: `df.describe()`

Out[73]:

| | Time | V1 | V2 | V3 | V4 | |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|
| count | 284807.000000 | 2.848070e+05 | 2.848070e+05 | 2.848070e+05 | 2.848070e+05 | 2.848070e+05 |
| mean | 94813.859575 | 3.919560e-15 | 5.688174e-16 | -8.769071e-15 | 2.782312e-15 | -1.551111e-15 |
| std | 47488.145955 | 1.958696e+00 | 1.651309e+00 | 1.516255e+00 | 1.415869e+00 | 1.381111e+00 |
| min | 0.000000 | -5.640751e+01 | -7.271573e+01 | -4.832559e+01 | -5.683171e+00 | -1.131111e+01 |
| 25% | 54201.500000 | -9.203734e-01 | -5.985499e-01 | -8.903648e-01 | -8.486401e-01 | -6.911111e-01 |
| 50% | 84692.000000 | 1.810880e-02 | 6.548556e-02 | 1.798463e-01 | -1.984653e-02 | -5.431111e-02 |
| 75% | 139320.500000 | 1.315642e+00 | 8.037239e-01 | 1.027196e+00 | 7.433413e-01 | 6.111111e-01 |
| max | 172792.000000 | 2.454930e+00 | 2.205773e+01 | 9.382558e+00 | 1.687534e+01 | 3.481111e+01 |

8 rows × 31 columns



In [74]: `df.Amount.min()`

Out[74]: 0.0

In [75]: `df.Amount.max()`

Out[75]: 25691.16

In [76]: `df.Time.min()`

Out[76]: 0.0

In [77]: `df.Time.max()`

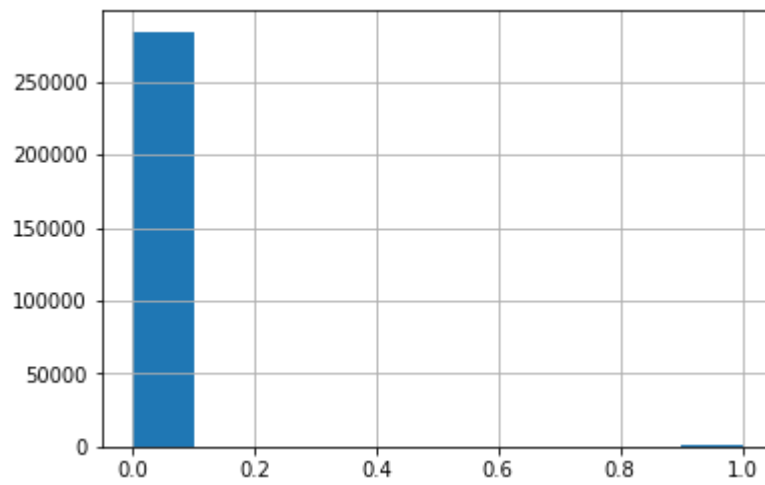
Out[77]: 172792.0

In [78]: `df.Class.mode()`

Out[78]: 0 0
dtype: int64

```
In [79]: df.Class.hist()
```

```
Out[79]: <matplotlib.axes._subplots.AxesSubplot at 0x15404389da0>
```



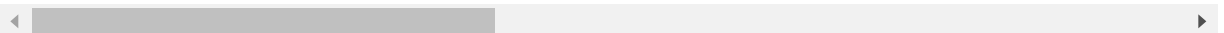
```
In [80]: df.loc[df['Class'] == 1] #to locate rows with card frauds
```

Out[80]:

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | |
|---------------|----------|-----------|-----------|------------|----------|-----------|-----------|-------|
| 541 | 406.0 | -2.312227 | 1.951992 | -1.609851 | 3.997906 | -0.522188 | -1.426545 | -2.53 |
| 623 | 472.0 | -3.043541 | -3.157307 | 1.088463 | 2.288644 | 1.359805 | -1.064823 | 0.325 |
| 4920 | 4462.0 | -2.303350 | 1.759247 | -0.359745 | 2.330243 | -0.821628 | -0.075788 | 0.562 |
| 6108 | 6986.0 | -4.397974 | 1.358367 | -2.592844 | 2.679787 | -1.128131 | -1.706536 | -3.49 |
| 6329 | 7519.0 | 1.234235 | 3.019740 | -4.304597 | 4.732795 | 3.624201 | -1.357746 | 1.713 |
| 6331 | 7526.0 | 0.008430 | 4.137837 | -6.240697 | 6.675732 | 0.768307 | -3.353060 | -1.63 |
| 6334 | 7535.0 | 0.026779 | 4.132464 | -6.560600 | 6.348557 | 1.329666 | -2.513479 | -1.68 |
| 6336 | 7543.0 | 0.329594 | 3.712889 | -5.775935 | 6.078266 | 1.667359 | -2.420168 | -0.81 |
| 6338 | 7551.0 | 0.316459 | 3.809076 | -5.615159 | 6.047445 | 1.554026 | -2.651353 | -0.74 |
| 6427 | 7610.0 | 0.725646 | 2.300894 | -5.329976 | 4.007683 | -1.730411 | -1.732193 | -3.96 |
| 6446 | 7672.0 | 0.702710 | 2.426433 | -5.234513 | 4.416661 | -2.170806 | -2.667554 | -3.87 |
| 6472 | 7740.0 | 1.023874 | 2.001485 | -4.769752 | 3.819195 | -1.271754 | -1.734662 | -3.05 |
| 6529 | 7891.0 | -1.585505 | 3.261585 | -4.137422 | 2.357096 | -1.405043 | -1.879437 | -3.51 |
| 6609 | 8090.0 | -1.783229 | 3.402794 | -3.822742 | 2.625368 | -1.976415 | -2.731689 | -3.43 |
| 6641 | 8169.0 | 0.857321 | 4.093912 | -7.423894 | 7.380245 | 0.973366 | -2.730762 | -1.49 |
| 6717 | 8408.0 | -1.813280 | 4.917851 | -5.926130 | 5.701500 | 1.204393 | -3.035138 | -1.71 |
| 6719 | 8415.0 | -0.251471 | 4.313523 | -6.891438 | 6.796797 | 0.616297 | -2.966327 | -2.43 |
| 6734 | 8451.0 | 0.314597 | 2.660670 | -5.920037 | 4.522500 | -2.315027 | -2.278352 | -4.68 |
| 6774 | 8528.0 | 0.447396 | 2.481954 | -5.660814 | 4.455923 | -2.443780 | -2.185040 | -4.71 |
| 6820 | 8614.0 | -2.169929 | 3.639654 | -4.508498 | 2.730668 | -2.122693 | -2.341017 | -4.23 |
| 6870 | 8757.0 | -1.863756 | 3.442644 | -4.468260 | 2.805336 | -2.118412 | -2.332285 | -4.26 |
| 6882 | 8808.0 | -4.617217 | 1.695694 | -3.114372 | 4.328199 | -1.873257 | -0.989908 | -4.57 |
| 6899 | 8878.0 | -2.661802 | 5.856393 | -7.653616 | 6.379742 | -0.060712 | -3.131550 | -3.10 |
| 6903 | 8886.0 | -2.535852 | 5.793644 | -7.618463 | 6.395830 | -0.065210 | -3.136372 | -3.10 |
| 6971 | 9064.0 | -3.499108 | 0.258555 | -4.489558 | 4.853894 | -6.974522 | 3.628382 | 5.431 |
| 8296 | 11080.0 | -2.125490 | 5.973556 | -11.034727 | 9.007147 | -1.689451 | -2.854415 | -7.81 |
| 8312 | 11092.0 | 0.378275 | 3.914797 | -5.726872 | 6.094141 | 1.698875 | -2.807314 | -0.59 |
| 8335 | 11131.0 | -1.426623 | 4.141986 | -9.804103 | 6.666273 | -4.749527 | -2.073129 | -10.0 |
| 8615 | 11629.0 | -3.891192 | 7.098916 | -11.426467 | 8.607557 | -2.065706 | -2.985288 | -8.13 |
| 8617 | 11635.0 | 0.919137 | 4.199633 | -7.535607 | 7.426940 | 1.118215 | -2.886722 | -1.34 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 251891 | 155548.0 | 1.878230 | 1.325630 | -2.333469 | 4.233151 | 1.355184 | -0.853508 | 0.716 |

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | |
|--------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-------|
| 251904 | 155554.0 | -1.040067 | 3.106703 | -5.409027 | 3.109903 | -0.887237 | -2.497522 | -2.07 |
| 252124 | 155662.0 | -1.928613 | 4.601506 | -7.124053 | 5.716088 | 1.026579 | -3.189073 | -2.26 |
| 252774 | 155965.0 | -1.201398 | 4.864535 | -8.328823 | 7.652399 | -0.167445 | -2.767695 | -3.17 |
| 254344 | 156685.0 | -0.129778 | 0.141547 | -0.894702 | -0.457662 | 0.810608 | -0.504723 | 1.373 |
| 254395 | 156710.0 | 0.202402 | 1.176270 | 0.346379 | 2.882138 | 1.407133 | -0.504355 | 1.438 |
| 255403 | 157207.0 | 1.170756 | 2.501038 | -4.986159 | 5.374160 | 0.997798 | -1.259004 | -1.23 |
| 255556 | 157284.0 | -0.242245 | 4.147186 | -5.672349 | 6.493741 | 1.591168 | -1.602523 | -0.95 |
| 258403 | 158638.0 | -5.976119 | -7.196980 | -5.388316 | 5.104799 | 4.676533 | -5.566870 | -4.29 |
| 261056 | 159844.0 | -0.408111 | 3.132944 | -3.098030 | 5.803893 | 0.890609 | -0.501474 | -0.44 |
| 261473 | 160034.0 | -2.349340 | 1.512604 | -2.647497 | 1.753792 | 0.406328 | -2.188494 | -0.68 |
| 261925 | 160243.0 | -2.783865 | 1.596824 | -2.084844 | 2.512986 | -1.446749 | -0.828496 | -0.73 |
| 262560 | 160537.0 | 0.567539 | 3.309385 | -6.631268 | 6.394574 | -0.054172 | -2.396535 | -2.79 |
| 262826 | 160665.0 | -0.417340 | 4.700055 | -7.521767 | 7.671884 | 0.260821 | -2.646693 | -2.85 |
| 263080 | 160791.0 | 2.132386 | 0.705608 | -3.530759 | 0.514779 | 1.527175 | -1.716268 | 1.132 |
| 263274 | 160870.0 | -0.644278 | 5.002352 | -8.252739 | 7.756915 | -0.216267 | -2.751496 | -3.35 |
| 263324 | 160895.0 | -0.848290 | 2.719882 | -6.199070 | 3.044437 | -3.301910 | -1.992117 | -3.73 |
| 263877 | 161154.0 | -3.387601 | 3.977881 | -6.978585 | 1.657766 | -1.100500 | -3.599487 | -3.68 |
| 268375 | 163181.0 | -5.238808 | 0.623013 | -5.784507 | 1.678889 | -0.364432 | -0.477295 | -4.27 |
| 272521 | 165132.0 | -7.503926 | -0.360628 | -3.830952 | 2.486103 | 2.497367 | 1.332437 | -6.78 |
| 274382 | 165981.0 | -5.766879 | -8.402154 | 0.056543 | 6.950983 | 9.880564 | -5.773192 | -5.74 |
| 274475 | 166028.0 | -0.956390 | 2.361594 | -3.171195 | 1.970759 | 0.474761 | -1.902598 | -0.05 |
| 275992 | 166831.0 | -2.027135 | -1.131890 | -1.135194 | 1.086963 | -0.010547 | 0.423797 | 3.790 |
| 276071 | 166883.0 | 2.091900 | -0.757459 | -1.192258 | -0.755458 | -0.620324 | -0.322077 | -1.08 |
| 276864 | 167338.0 | -1.374424 | 2.793185 | -4.346572 | 2.400731 | -1.688433 | 0.111136 | -0.92 |
| 279863 | 169142.0 | -1.927883 | 1.125653 | -4.518331 | 1.749293 | -1.566487 | -2.010494 | -0.88 |
| 280143 | 169347.0 | 1.378559 | 1.289381 | -5.004247 | 1.411850 | 0.442581 | -1.326536 | -1.41 |
| 280149 | 169351.0 | -0.676143 | 1.126366 | -2.213700 | 0.468308 | -1.120541 | -0.003346 | -2.23 |
| 281144 | 169966.0 | -3.113832 | 0.585864 | -5.399730 | 1.817092 | -0.840618 | -2.943548 | -2.20 |
| 281674 | 170348.0 | 1.991976 | 0.158476 | -2.583441 | 0.408670 | 1.151147 | -0.096695 | 0.223 |

492 rows × 31 columns



```
In [21]: df.groupby('Class').count()
```

Out[21]:

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Class | | | | | | | | | | |
| 0 | 284315 | 284315 | 284315 | 284315 | 284315 | 284315 | 284315 | 284315 | 284315 | 284315 |
| 1 | 492 | 492 | 492 | 492 | 492 | 492 | 492 | 492 | 492 | 492 |

2 rows × 30 columns



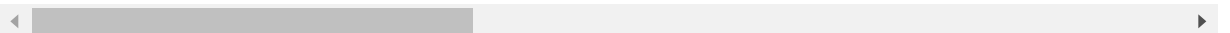
```
In [22]: df.sort_values(['Amount'],ascending=True, inplace=False)
```

Out[22]:

| | Time | V1 | V2 | V3 | V4 | V5 | V6 |
|---------------|----------|-----------|-----------|------------|-----------|------------|-----------|
| 15816 | 27255.0 | 1.248804 | 0.047208 | 0.423388 | -0.139515 | -0.592217 | -0.980654 |
| 77470 | 57062.0 | -1.188664 | -0.612034 | 2.422204 | -0.812786 | 0.318493 | -0.671637 |
| 190885 | 129019.0 | 1.868263 | 0.273764 | -0.288023 | 3.835852 | 0.268329 | 0.817380 |
| 87335 | 61640.0 | -0.848470 | 1.426562 | 2.137094 | 2.852036 | -0.366945 | 1.158146 |
| 174481 | 121931.0 | -1.184195 | 0.804518 | 2.240498 | 2.853175 | 1.038068 | 0.171728 |
| 261840 | 160203.0 | 2.037530 | -0.068534 | -1.566791 | 0.125108 | 0.301918 | -0.714361 |
| 60681 | 49446.0 | 1.051131 | 0.078365 | 1.371753 | 2.662810 | -0.347892 | 1.390227 |
| 115131 | 73758.0 | -0.956441 | 1.002075 | 0.870574 | -1.475914 | 0.029941 | -0.326977 |
| 190894 | 129024.0 | 1.895858 | 0.300037 | -0.177958 | 4.044251 | -0.037868 | 0.200553 |
| 115136 | 73760.0 | -0.907995 | 0.758230 | 1.819020 | -0.545179 | -0.795851 | -0.063335 |
| 214401 | 139661.0 | -1.457568 | 1.601630 | -0.497917 | -1.025701 | 0.255455 | -1.339380 |
| 174505 | 121941.0 | 1.821246 | 0.396278 | 0.239165 | 4.248032 | -0.368472 | -0.229155 |
| 174506 | 121941.0 | 1.882277 | 0.256156 | -0.164326 | 4.056145 | -0.081415 | 0.238211 |
| 174509 | 121943.0 | 0.097001 | 1.048308 | -0.990620 | -0.460902 | 1.297780 | -1.637058 |
| 60732 | 49471.0 | 1.070972 | 0.144385 | 1.490478 | 2.855083 | -0.557954 | 0.879469 |
| 140093 | 83533.0 | 1.125524 | 0.189854 | 1.508660 | 2.719937 | -0.882342 | 0.128390 |
| 230216 | 146239.0 | 1.899345 | 0.323415 | 0.046089 | 3.740338 | 0.050781 | 0.692388 |
| 36101 | 38396.0 | -2.310780 | 0.081239 | 2.401953 | 1.671214 | 0.276306 | 0.478155 |
| 4741 | 4185.0 | -0.843893 | 1.097478 | 2.608834 | 2.307204 | -0.211221 | 0.276982 |
| 35965 | 38339.0 | 1.211645 | 0.383396 | 0.438614 | 2.265713 | 0.343609 | 0.877276 |
| 214544 | 139716.0 | -2.006121 | 2.048709 | -0.963971 | -1.046216 | -0.118273 | -1.400030 |
| 200628 | 133509.0 | -1.075657 | 0.389560 | -2.728046 | -0.531685 | 1.447749 | -1.957976 |
| 65767 | 51735.0 | -0.683837 | 0.562573 | 1.792653 | -0.155097 | 0.387675 | -0.584327 |
| 249926 | 154640.0 | 1.992562 | -0.501778 | -0.507760 | 0.274683 | -0.592847 | -0.135490 |
| 214730 | 139799.0 | 2.008262 | -0.465498 | -0.245160 | 0.244059 | -0.904274 | -0.699963 |
| 114982 | 73700.0 | -0.504772 | 0.762501 | 1.535679 | -0.782013 | 0.135575 | -1.046358 |
| 131213 | 79522.0 | -0.749358 | -0.293553 | 1.884835 | -0.019693 | 0.661074 | 0.057788 |
| 164921 | 117076.0 | -0.323021 | 0.192965 | -0.050790 | -2.549011 | 1.350568 | -0.894664 |
| 140041 | 83504.0 | 1.305420 | 0.034120 | 0.300825 | -0.063400 | -0.623145 | -1.208488 |
| 19696 | 30475.0 | -0.555119 | 0.686358 | 2.672443 | 0.825653 | -0.255788 | 0.842306 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 199032 | 132779.0 | -4.389647 | -1.993089 | -12.627453 | 0.282311 | -29.730600 | 21.307738 |

| | Time | V1 | V2 | V3 | V4 | V5 | V6 |
|--------|----------|------------|------------|------------|-----------|-------------|-----------|
| 48529 | 43684.0 | -12.618406 | -16.550248 | -7.779531 | 2.245844 | -7.609465 | 3.514395 |
| 190137 | 128701.0 | -19.780626 | -25.663628 | -10.865410 | 6.046025 | -16.459773 | 9.410864 |
| 44513 | 41978.0 | -11.839743 | -17.711661 | -6.893572 | 4.385610 | -18.883164 | 10.857248 |
| 217781 | 141031.0 | -11.128890 | -22.632193 | -9.398485 | 5.163725 | -5.105794 | 6.483704 |
| 238412 | 149657.0 | -11.448790 | -21.749060 | -8.478933 | 9.477565 | -6.349227 | 3.148005 |
| 37000 | 38763.0 | -14.711825 | -23.250844 | -7.631400 | 5.975826 | -15.615302 | 8.060516 |
| 216065 | 140335.0 | -11.446245 | -17.964561 | -10.768507 | 4.912280 | -17.332406 | 11.025540 |
| 228525 | 145548.0 | -11.316807 | -22.225978 | -10.723514 | 5.827606 | -5.752157 | 1.413055 |
| 228723 | 145630.0 | -32.543140 | -50.383269 | -10.733302 | 9.571705 | 8.656575 | -6.377450 |
| 23128 | 32605.0 | -23.383129 | -18.098804 | -20.195367 | 4.249104 | -42.147898 | 22.529298 |
| 218985 | 141546.0 | -13.396920 | -19.230653 | -9.042012 | 5.678408 | -21.577019 | 12.128950 |
| 201638 | 133971.0 | -10.950173 | -13.359133 | -10.664755 | 1.157565 | -28.363785 | 17.019934 |
| 57209 | 47807.0 | -13.537461 | -24.426864 | -7.355943 | 6.896486 | -8.515928 | 2.354758 |
| 1632 | 1264.0 | -11.140706 | -9.612726 | -12.389545 | 6.013346 | -32.092129 | 21.393069 |
| 188560 | 128027.0 | -12.895214 | -20.492129 | -12.794629 | 4.691974 | -17.515077 | 15.568823 |
| 240739 | 150726.0 | -13.099603 | -25.434451 | -11.921912 | 4.976147 | -6.139222 | 3.203290 |
| 19760 | 30537.0 | -14.191832 | -25.313252 | -7.578781 | 7.730844 | -8.285413 | 3.994474 |
| 228158 | 145381.0 | -13.710746 | -21.177912 | -9.340628 | 8.385815 | -22.756016 | 14.435314 |
| 226690 | 144755.0 | -11.066027 | -5.312984 | -13.391432 | 2.674077 | -27.737487 | 16.410638 |
| 245474 | 152763.0 | -14.641710 | -28.554825 | -12.714462 | 5.878264 | -7.855074 | 2.471004 |
| 74699 | 55709.0 | -16.950064 | -16.417395 | -12.523381 | 6.555638 | -27.752964 | 18.072031 |
| 227921 | 145283.0 | -21.532478 | -34.704768 | -8.303035 | 10.264175 | 3.957175 | -3.229695 |
| 284249 | 172273.0 | -9.030538 | -11.112584 | -16.233798 | 3.592021 | -40.427726 | 23.917837 |
| 169457 | 119713.0 | -20.924897 | -37.943452 | -14.060281 | 10.473005 | -10.866639 | 6.256654 |
| 54018 | 46253.0 | -21.780665 | -38.305310 | -12.122469 | 9.752791 | -12.880794 | 4.256017 |
| 46841 | 42951.0 | -23.712839 | -42.172688 | -13.320825 | 9.925019 | -13.945538 | 5.564891 |
| 151296 | 95286.0 | -34.549296 | -60.464618 | -21.340854 | 16.875344 | -19.229075 | 6.335259 |
| 58465 | 48401.0 | -36.802320 | -63.344698 | -20.645794 | 16.715537 | -20.672064 | 7.694002 |
| 274771 | 166198.0 | -35.548539 | -31.850484 | -48.325589 | 15.304184 | -113.743307 | 73.301626 |

284807 rows × 31 columns



```
In [195]: import random
random.seed(3)
```

```
In [196]: no_frauds = len(df[df['Class'] == 1])
```

```
In [197]: non_fraud_indices = df[df.Class == 0].index
```

```
In [198]: random_indices = np.random.choice(non_fraud_indices,no_frauds, replace=False)
```

```
In [199]: fraud_indices = df[df.Class == 1].index
```

```
In [200]: under_sample_indices = np.concatenate([fraud_indices,random_indices])
```

```
In [201]: under_sample = df.loc[under_sample_indices]
```

```
In [202]: X_under = under_sample.loc[:,under_sample.columns != 'Class']
y_under = under_sample.loc[:,under_sample.columns == 'Class']
X_under_train, X_under_test, y_under_train, y_under_test = train_test_split(X_
under,y_under,test_size = 0.3, random_state = 0)
```

```
In [203]: c=DecisionTreeClassifier()
```

```
In [204]: dt=c.fit(X_under_train,y_under_train)
```

```
In [205]: y_under_pred=c.predict(X_under_test)
```

```
In [206]: y_under_pred
```

```
Out[206]: array([0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0,
1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1,
0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0,
1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0,
1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1,
0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0,
1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1,
0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0,
1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1,
1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0,
0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1,
0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1,
1, 0, 1, 1, 0, 1, 0, 0, 0, 1], dtype=int64)
```

```
In [207]: from sklearn.metrics import accuracy_score
```

```
In [208]: score= accuracy_score(y_under_test, y_under_pred)*100
```

```
In [209]: print("Accuracy using Decision tree",round(score,2),"%")
```

Accuracy using Decision tree 92.23 %

```
In [210]: from sklearn.metrics import confusion_matrix
cnf_matrix = confusion_matrix(y_under_test, y_under_pred)
```

```
In [233]: cnf_matrix = confusion_matrix(y_under_test, y_under_pred)
print (cnf_matrix)
```

```
[[137  12]
 [ 11 136]]
```

```
In [234]: print( 'Recall: ' + str(np.round(100*float((cnf_matrix[1][1]))/float((cnf_matrix[1][0]+cnf_matrix[1][1])),2))+'%')
```

Recall: 92.52%

```
In [235]: Precision=float((cnf_matrix[1][1]))/float((cnf_matrix[1][0]+cnf_matrix[0][1]))
print(Precision)
```

5.913043478260869

```
In [236]: Recall=float((cnf_matrix[1][1]))/float((cnf_matrix[1][0]+cnf_matrix[1][1]))
print(Recall)
```

0.9251700680272109

```
In [237]: F1_Score = 2*(Recall * Precision) / (Recall + Precision)
print(F1_Score)
```

1.5999999999999999

```
In [238]: #to check for whole data
```

```
X = df.iloc[:,1:29]
Y = df.Class
```

```
In [239]: X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size = 0.3)
```

```
In [240]: c=DecisionTreeClassifier()
```

```
In [241]: c.fit(X_train, Y_train)
```

```
Out[241]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort=False, random_state=None,
splitter='best')
```

```
In [242]: y_pred_full = c.predict(X_test)
```

```
In [243]: y_pred_full
```

```
Out[243]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
In [244]: from sklearn.metrics import accuracy_score
```

```
In [245]: score= accuracy_score(Y_test, y_pred_full)*100
```

```
In [246]: print("Accuracy using Decision tree",round(score,2),"%")  
Accuracy using Decision tree 99.91 %
```

```
In [247]: cnf_matrix = confusion_matrix(Y_test, y_pred_full)
```

```
In [248]: print (cnf_matrix)  
[[85251   41]  
 [   32  119]]
```

```
In [249]: Precision=float((cnf_matrix[1][1]))/float((cnf_matrix[1][0]+cnf_matrix[0][1]))  
print(Precision)  
1.63013698630137
```

```
In [250]: Recall=float((cnf_matrix[1][1]))/float((cnf_matrix[1][0]+cnf_matrix[1][1]))  
print(Recall)  
0.7880794701986755
```

```
In [251]: print( 'Recall: ' + str(np.round(100*float((cnf_matrix[1][1]))/float((cnf_matr  
ix[1][0]+cnf_matrix[1][1])),2))+'%')  
Recall: 78.81%
```

```
In [252]: F1_Score = 2*(Recall * Precision) / (Recall + Precision)  
print(F1_Score)  
1.0625
```