```python
In [3]:  import matplotlib.pyplot as plt
         import pandas as pd
         import numpy as np
         %matplotlib inline
```

```python
In [41]:  df = pd.read_csv('creditcard.csv')

          # take a look at the dataset
          df.head()
```

Out[41]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739

5 rows × 31 columns

```python
In [42]:  # number of rows and columns
          df.shape
```

Out[42]: (284807, 31)

```python
In [43]:  # Index, Datatype and Memory information
          df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype
---  ------  --------------   -----
 0   Time    284807 non-null  float64
 1   V1      284807 non-null  float64
 2   V2      284807 non-null  float64
 3   V3      284807 non-null  float64
 4   V4      284807 non-null  float64
 5   V5      284807 non-null  float64
 6   V6      284807 non-null  float64
 7   V7      284807 non-null  float64
 8   V8      284807 non-null  float64
 9   V9      284807 non-null  float64
 10  V10     284807 non-null  float64
 11  V11     284807 non-null  float64
 12  V12     284807 non-null  float64
 13  V13     284807 non-null  float64
 14  V14     284807 non-null  float64
 15  V15     284807 non-null  float64
 16  V16     284807 non-null  float64
 17  V17     284807 non-null  float64
 18  V18     284807 non-null  float64
 19  V19     284807 non-null  float64
 20  V20     284807 non-null  float64
 21  V21     284807 non-null  float64
```

```
22  V22     284807 non-null  float64
23  V23     284807 non-null  float64
24  V24     284807 non-null  float64
25  V25     284807 non-null  float64
26  V26     284807 non-null  float64
27  V27     284807 non-null  float64
28  V28     284807 non-null  float64
29  Amount  284807 non-null  float64
30  Class   284807 non-null  int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

In [44]: `# Summary statistics for numerical columns`
`df.describe()`

Out[44]:

	Time	V1	V2	V3	V4	V5	
count	284807.000000	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.84
mean	94813.859575	3.919560e-15	5.688174e-16	-8.769071e-15	2.782312e-15	-1.552563e-15	2.0
std	47488.145955	1.958696e+00	1.651309e+00	1.516255e+00	1.415869e+00	1.380247e+00	1.33
min	0.000000	-5.640751e+01	-7.271573e+01	-4.832559e+01	-5.683171e+00	-1.137433e+02	-2.61
25%	54201.500000	-9.203734e-01	-5.985499e-01	-8.903648e-01	-8.486401e-01	-6.915971e-01	-7.6
50%	84692.000000	1.810880e-02	6.548556e-02	1.798463e-01	-1.984653e-02	-5.433583e-02	-2.7
75%	139320.500000	1.315642e+00	8.037239e-01	1.027196e+00	7.433413e-01	6.119264e-01	3.9
max	172792.000000	2.454930e+00	2.205773e+01	9.382558e+00	1.687534e+01	3.480167e+01	7.33

8 rows × 31 columns

In [45]: `# Returns the number of non-null values in each DataFrame column`
`df.count()`

```
Out[45]: Time     284807
         V1       284807
         V2       284807
         V3       284807
         V4       284807
         V5       284807
         V6       284807
         V7       284807
         V8       284807
         V9       284807
         V10      284807
         V11      284807
         V12      284807
         V13      284807
         V14      284807
         V15      284807
         V16      284807
         V17      284807
         V18      284807
         V19      284807
         V20      284807
         V21      284807
         V22      284807
         V23      284807
```

```
V24      284807
V25      284807
V26      284807
V27      284807
V28      284807
Amount   284807
Class    284807
dtype: int64
```

In [46]:
```python
# Returns the highest value in each column
df.max()
```

Out[46]:
```
Time     172792.000000
V1            2.454930
V2           22.057729
V3            9.382558
V4           16.875344
V5           34.801666
V6           73.301626
V7          120.589494
V8           20.007208
V9           15.594995
V10          23.745136
V11          12.018913
V12           7.848392
V13           7.126883
V14          10.526766
V15           8.877742
V16          17.315112
V17           9.253526
V18           5.041069
V19           5.591971
V20          39.420904
V21          27.202839
V22          10.503090
V23          22.528412
V24           4.584549
V25           7.519589
V26           3.517346
V27          31.612198
V28          33.847808
Amount    25691.160000
Class         1.000000
dtype: float64
```

In [47]:
```python
# Returns the correlation between columns in a DataFrame
pd.isnull(df).head()
```

Out[47]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25	V2
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	Fal

5 rows × 31 columns

```
In [48]: # Returns the number of non-null values in each DataFrame column and convert to array u
         arr = df.count().to_numpy()
         print(arr)
```

```
[284807 284807 284807 284807 284807 284807 284807 284807 284807 284807
 284807 284807 284807 284807 284807 284807 284807 284807 284807 284807
 284807 284807 284807 284807 284807 284807 284807 284807 284807 284807
 284807]
```

```
In [49]: # sorting using numpy
         sorted_arr = np.sort(arr)
         print(sorted_arr)
```
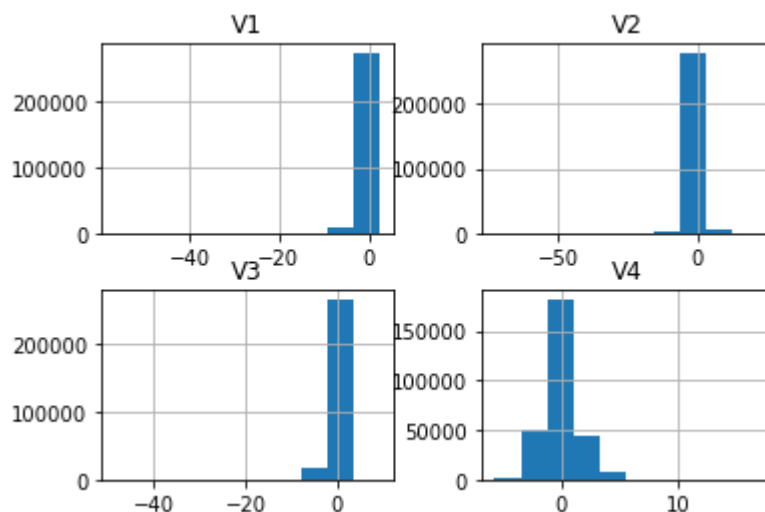
```
[284807 284807 284807 284807 284807 284807 284807 284807 284807 284807
 284807 284807 284807 284807 284807 284807 284807 284807 284807 284807
 284807 284807 284807 284807 284807 284807 284807 284807 284807 284807
 284807]
```

```
In [51]: cdf = df[['V1','V2','V3','V4']]
         cdf.head(9)
```

Out[51]:

	V1	V2	V3	V4
0	-1.359807	-0.072781	2.536347	1.378155
1	1.191857	0.266151	0.166480	0.448154
2	-1.358354	-1.340163	1.773209	0.379780
3	-0.966272	-0.185226	1.792993	-0.863291
4	-1.158233	0.877737	1.548718	0.403034
5	-0.425966	0.960523	1.141109	-0.168252
6	1.229658	0.141004	0.045371	1.202613
7	-0.644269	1.417964	1.074380	-0.492199
8	-0.894286	0.286157	-0.113192	-0.271526

```
In [52]: cdf.hist()
         plt.show()
```



```
In [ ]:
```