Final project

March 8, 2023

```python
In [1]: import pandas as pd
        data = pd.read_csv('Banknote-authentication-dataset.csv')
        data
```

```
Out[1]:             V1        V2
        0      3.621600   8.66610
        1      4.545900   8.16740
        2      3.866000  -2.63830
        3      3.456600   9.52280
        4      0.329240  -4.45520
        5      4.368400   9.67180
        6      3.591200   3.01290
        7      2.092200  -6.81000
        8      3.203200   5.75880
        9      1.535600   9.17720
        10     1.224700   8.77790
        11     3.989900  -2.70660
        12     1.899300   7.66250
        13    -1.576800  10.84300
        14     3.404000   8.72610
        15     4.676500  -3.38950
        16     2.671900   3.06460
        17     0.803550   2.84730
        18     1.447900  -4.87940
        19     5.242300  11.02720
        20     5.786700   7.89020
        21     0.329200  -4.45520
        22     3.936200  10.16220
        23     0.935840   8.88550
        24     4.433800   9.88700
        25     0.705700  -5.49810
        26     1.143200  -3.74130
        27    -0.382140   8.39090
        28     6.563300   9.81870
        29     4.890600  -3.35840
        ...         ...       ...
        1342  -1.747900  -5.82300
```

```
1343 -0.959230  -6.71280
1344  1.345100   0.23589
1345  2.227900   4.09510
1346  1.257200   4.87310
1347 -5.385700   9.12140
1348 -2.978600   2.34450
1349 -1.585100  -2.15620
1350 -0.218880  -2.20380
1351  1.318300   1.90170
1352  1.489600   3.42880
1353  0.115920   3.22190
1354 -3.392400   3.35640
1355 -6.163200   8.70960
1356 -4.078600   2.92390
1357 -2.589900  -0.39110
1358 -1.011600  -0.19038
1359  0.066129   2.49140
1360 -0.247450   1.93680
1361 -1.573200   1.06360
1362 -2.166800   1.59330
1363 -1.166700  -1.42370
1364 -2.839100  -6.63000
1365 -4.504600  -5.81260
1366 -2.410000   3.74330
1367  0.406140   1.34920
1368 -1.388700  -4.87730
1369 -3.750300 -13.45860
1370 -3.563700  -8.38270
1371 -2.541900  -0.65804

[1372 rows x 2 columns]
```

```python
In [10]: import numpy as np
         variance_mean = np.mean(data['V1'])
         skewness_mean = np.mean(data['V2'])
         print(f"V1 mean = {variance_mean}")
         print(f"V2 mean = {skewness_mean}")
```

```
V1 mean = 0.43373525728862977
V2 mean = 1.9223531209912539
```

```python
In [12]: v1_dev = np.std(data["V1"])
         v2_dev = np.std(data["V2"])

         print(f"standard deviation for v1 and v2 is {v1_dev} and {v2_dev}")
```
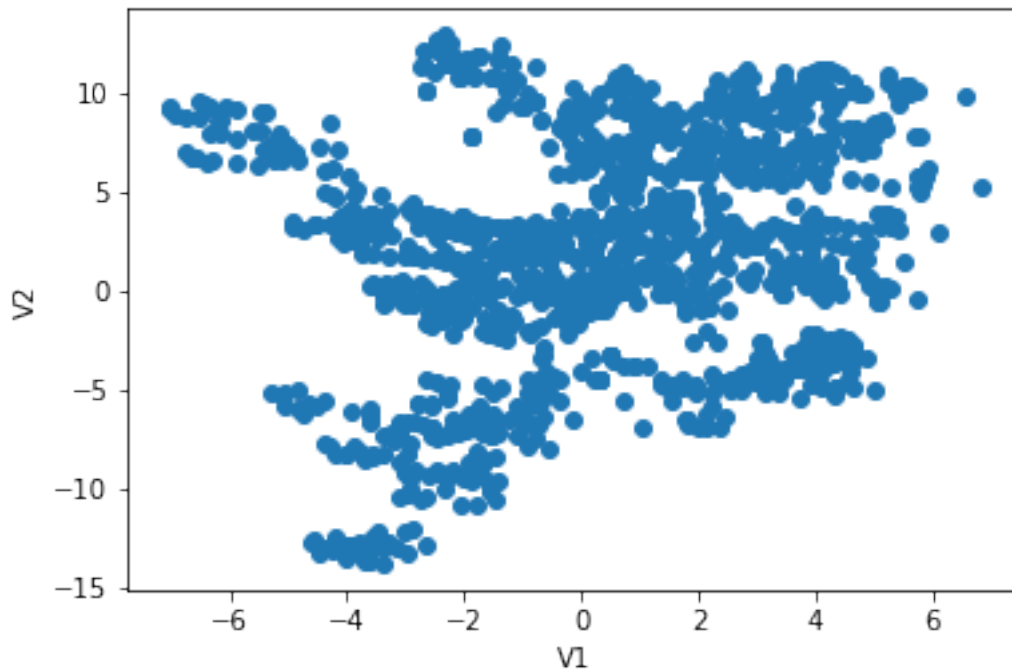
```
standard deviation for v1 and v2 is 2.841726405206097 and 5.866907488271993
```

```
In [23]: import matplotlib.pyplot as plt
         plt.xlabel('V1')
         plt.ylabel('V2')
         plt.scatter(data['V1'],data['V2'])
```

Out[23]: <matplotlib.collections.PathCollection at 0x7fb9a292a2b0>


```
In [27]: from sklearn.cluster import KMeans
         v1 = data["V1"]
         v2 = data['V2']
         v1_v2 = np.column_stack((v1,v2))
         print(v1_v2)

         km_result = KMeans(n_clusters = 3).fit(v1_v2)
         km_result.cluster_centers_
         clusters = km_result.cluster_centers_

         plt.scatter(v1,v2)
         plt.scatter(clusters[:,0], clusters[:,1], s= 1000)
```

```
[[  3.6216    8.6661 ]
 [  4.5459    8.1674 ]
 [  3.866    -2.6383 ]
 ...
 [ -3.7503  -13.4586 ]
```

```
 [ -3.5637    -8.3827 ]
 [ -2.5419    -0.65804]]
```