

projet_prb

May 17, 2020

1 Apprentissage

```
[5]:   0   1   2   3   4   5   6   7   8   9   ...  55  56  57  58  59  60  61  \
0   0   1   6  15  12   1   0   0   0   7   ...  0   0   0   6  14   7   1
1   0   0  10  16   6   0   0   0   0   7   ...  0   0   0  10  16  15   3
2   0   0   8  15  16  13   0   0   0   1   ...  0   0   0   9  14   0   0
3   0   0   0   3  11  16   0   0   0   0   ...  0   0   0   0   1  15   2
4   0   0   5  14   4   0   0   0   0   0   ...  0   0   0   4  12  14   7
```

```
      62  63  64
0   0   0   0
1   0   0   0
2   0   0   7
3   0   0   4
4   0   0   6
```

[5 rows x 65 columns]

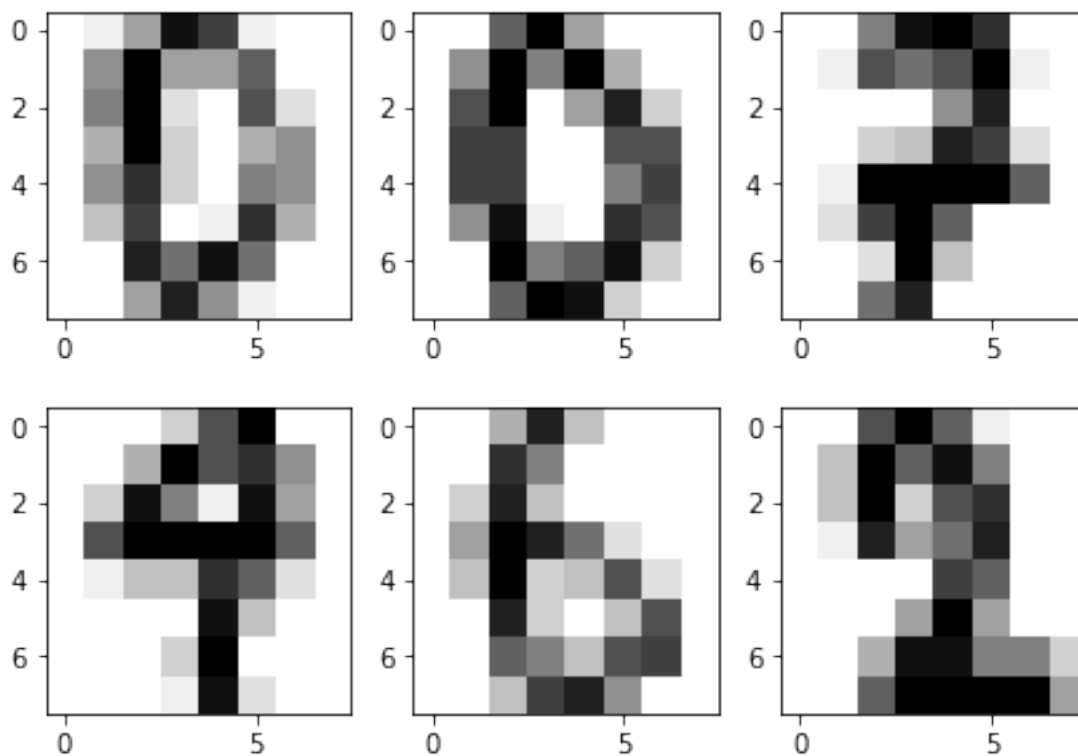
```
[6]:      0      1      2      3      4  \
count  3823.0  3823.000000  3823.000000  3823.000000  3823.000000
mean     0.0     0.301334     5.481821     11.805912     11.451478
std     0.0     0.866986     4.631601     4.259811     4.537556
min     0.0     0.000000     0.000000     0.000000     0.000000
25%     0.0     0.000000     1.000000    10.000000     9.000000
50%     0.0     0.000000     5.000000    13.000000    13.000000
75%     0.0     0.000000     9.000000    15.000000    15.000000
max     0.0     8.000000    16.000000    16.000000    16.000000

      5      6      7      8      9   ...  \
count  3823.000000  3823.000000  3823.000000  3823.000000  3823.000000  ...
mean     5.505362     1.387392     0.142297     0.002093     1.960502  ...
std     5.613060     3.371444     1.051598     0.088572     3.052353  ...
min     0.000000     0.000000     0.000000     0.000000     0.000000  ...
25%     0.000000     0.000000     0.000000     0.000000     0.000000  ...
50%     4.000000     0.000000     0.000000     0.000000     0.000000  ...
75%    10.000000     0.000000     0.000000     0.000000     3.000000  ...
max    16.000000    16.000000    16.000000     5.000000    15.000000  ...
```

	55	56	57	58	59 \
count	3823.000000	3823.000000	3823.000000	3823.000000	3823.000000
mean	0.148313	0.000262	0.283024	5.855872	11.942977
std	0.767761	0.016173	0.928046	4.980012	4.334508
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	1.000000	10.000000
50%	0.000000	0.000000	0.000000	5.000000	13.000000
75%	0.000000	0.000000	0.000000	10.000000	15.000000
max	12.000000	1.000000	10.000000	16.000000	16.000000

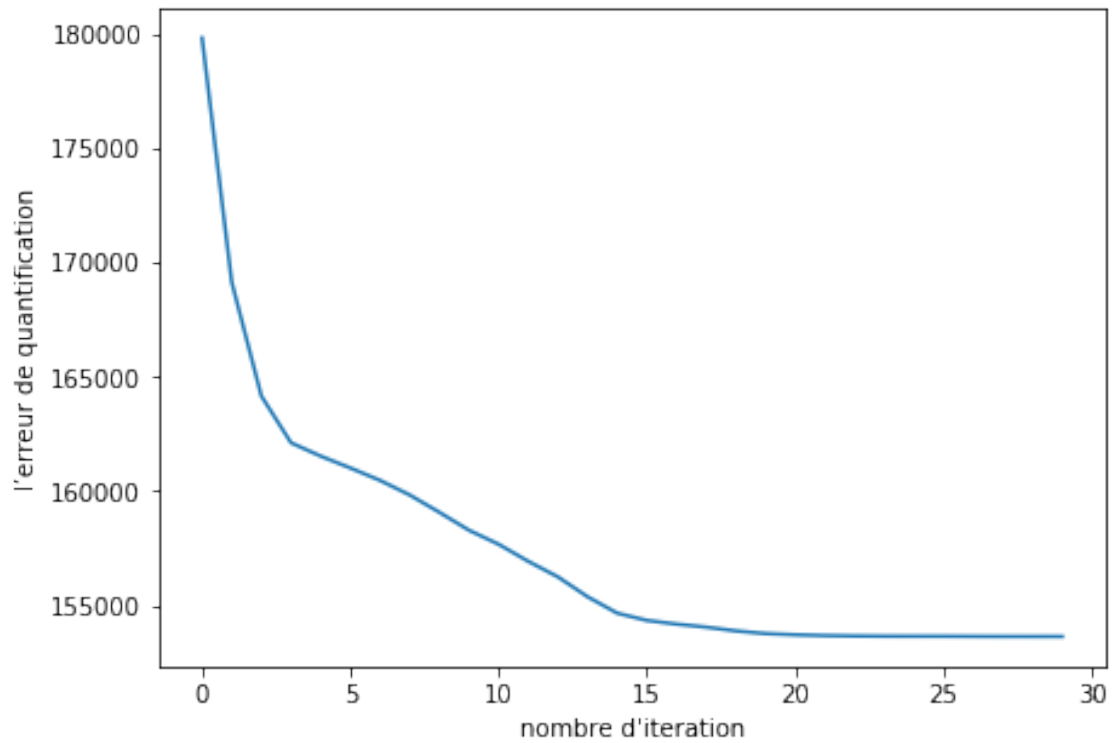
	60	61	62	63	64
count	3823.000000	3823.000000	3823.000000	3823.000000	3823.000000
mean	11.461156	6.700497	2.105676	0.202197	4.497253
std	4.991934	5.775815	4.028266	1.150694	2.869831
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	9.000000	0.000000	0.000000	0.000000	2.000000
50%	13.000000	6.000000	0.000000	0.000000	4.000000
75%	16.000000	12.000000	2.000000	0.000000	7.000000
max	16.000000	16.000000	16.000000	16.000000	9.000000

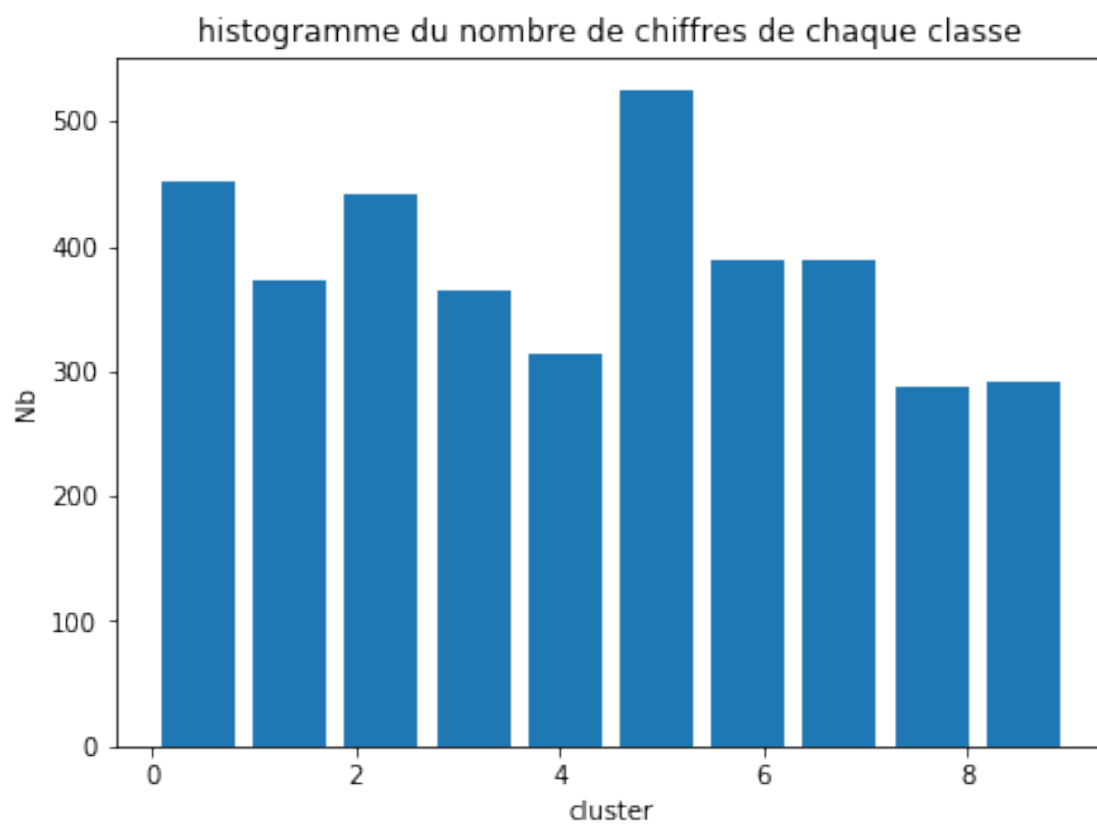
[8 rows x 65 columns]

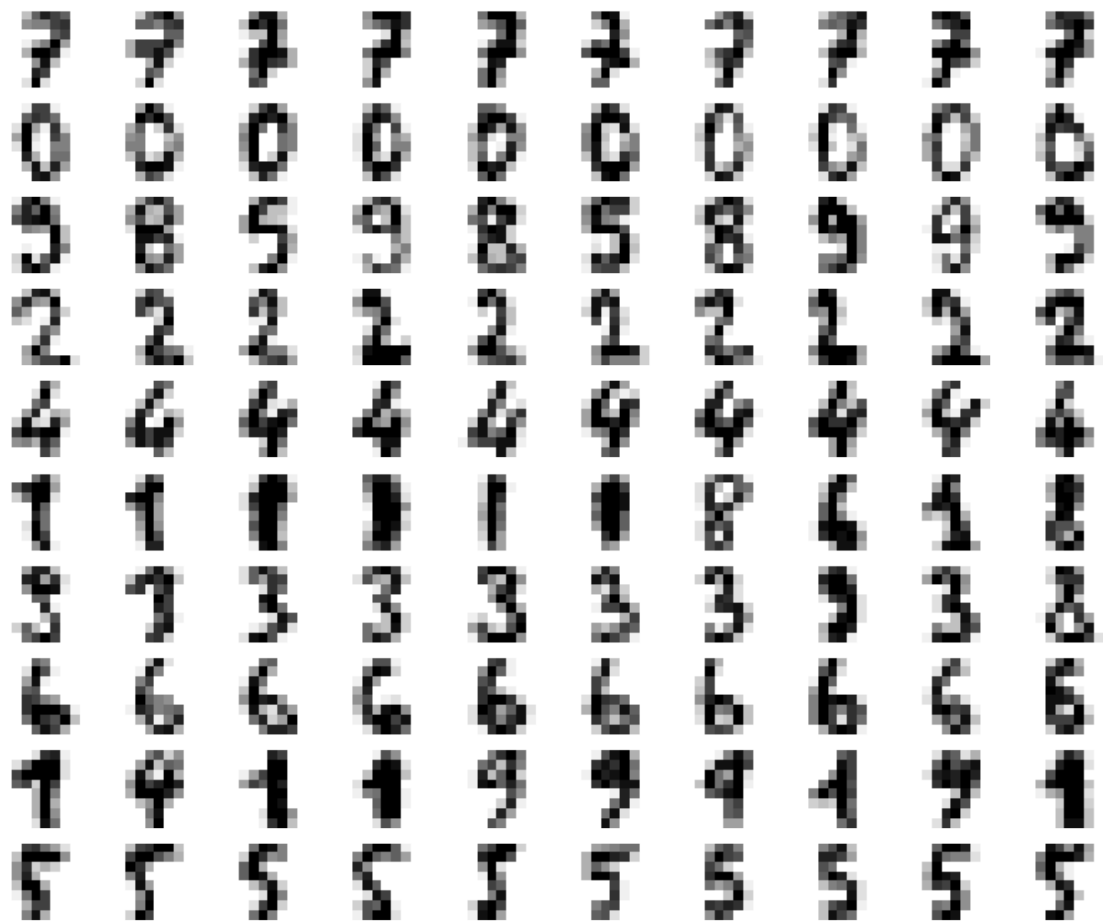


```
/Users/zuoyou/opt/anaconda3/lib/python3.7/site-packages/scipy/cluster/vq.py:139:  
RuntimeWarning: Some columns have standard deviation zero. The values of these  
columns will not change.  
RuntimeWarning)
```

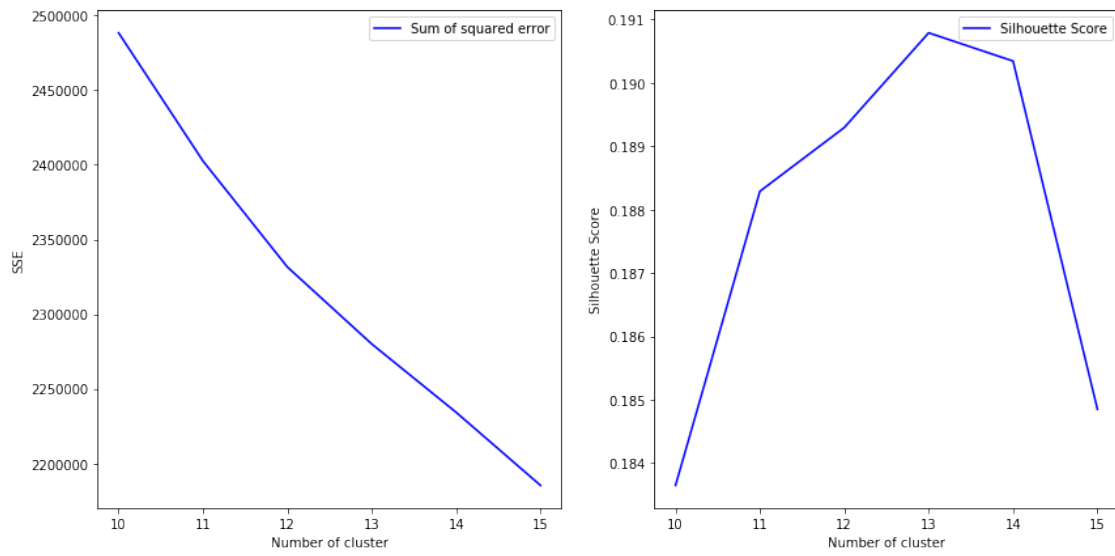
```
[9]: Text(0, 0.5, 'l'erreur de quantification')
```







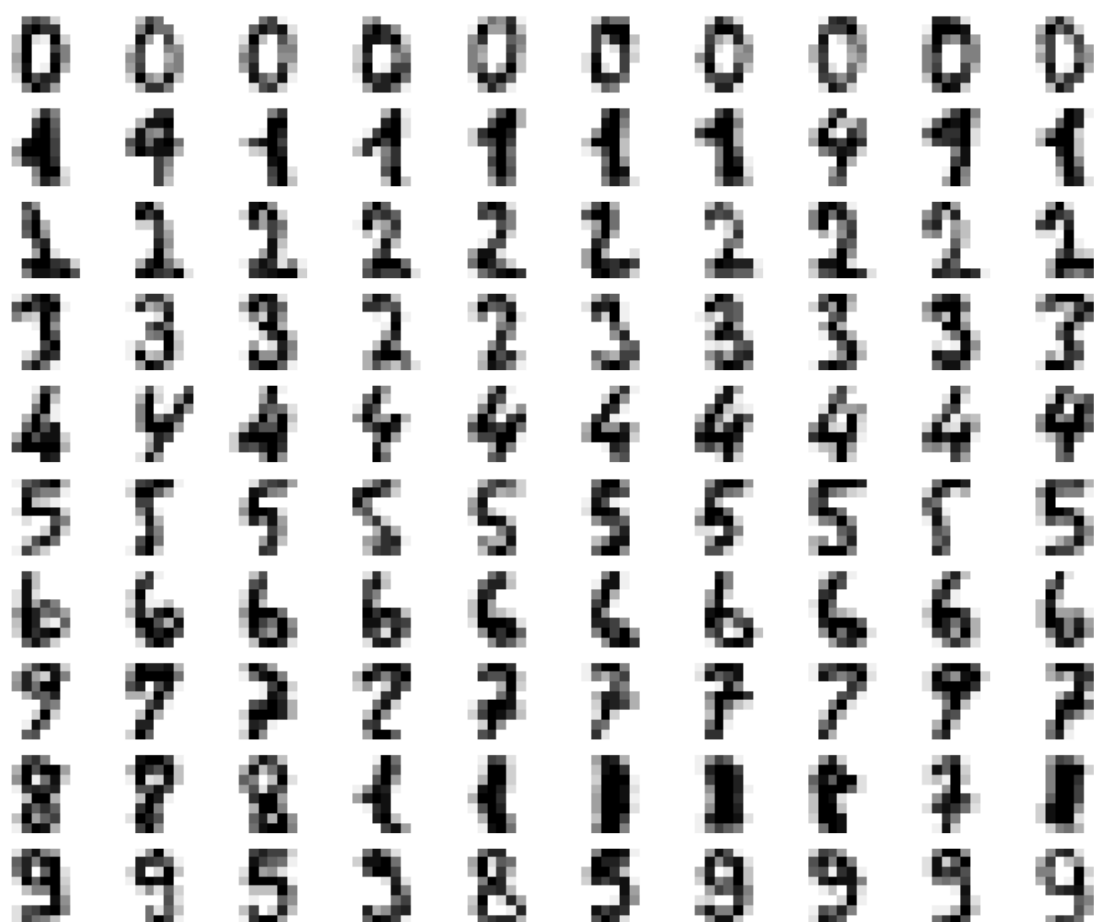
```
For n_clusters = 10 The average silhouette_score is : 0.18364722930013658
For n_clusters = 11 The average silhouette_score is : 0.18828833990205618
For n_clusters = 12 The average silhouette_score is : 0.18929383393509136
For n_clusters = 13 The average silhouette_score is : 0.19078825639251884
For n_clusters = 14 The average silhouette_score is : 0.19034308938748307
For n_clusters = 15 The average silhouette_score is : 0.18484933200940232
```



le meilleur $k = 10$

2 Test

[7, 0, 9, 2, 4, 8, 3, 6, 1, 5]



Confusion matrix

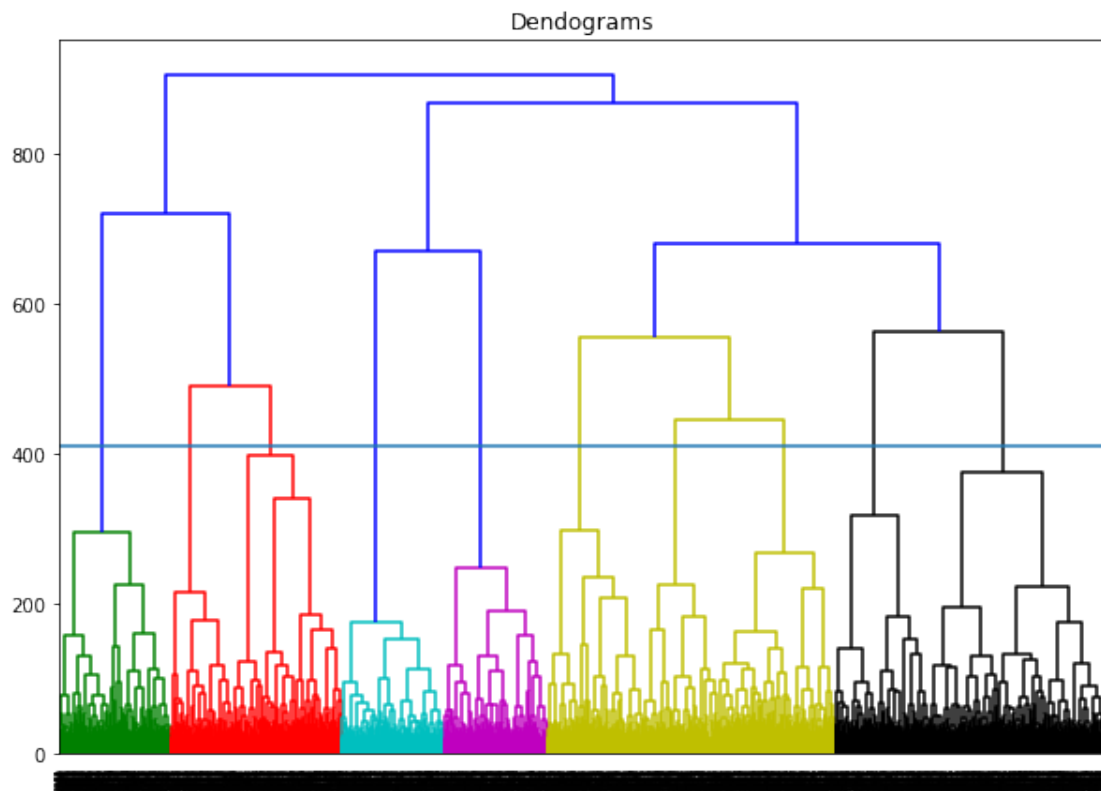
```
[[176  0  0  0  2  0  0  0  0  0]
 [  0 57 21  2  0  0  4  0 98  0]
 [  1  2 150  5  0  0  0  3 13  3]
 [  0  0  1 151  0  1  0  8  7 15]
 [  0  5  0  0 162  0  0  6  8  0]
 [  0  0  0  0  1 144  1  0  0 36]
 [  1  0  0  0  1  0 176  0  3  0]
 [  0  8  0  0  1  0  0 168  2  0]
 [  0  8  1  5  0  4  2  2 120 32]
 [  0 23  0  4  0  5  0  5  2 141]]
```

Classification report

	precision	recall	f1-score	support
0	0.99	0.99	0.99	178
1	0.55	0.31	0.40	182
2	0.87	0.85	0.86	177
3	0.90	0.83	0.86	183
4	0.97	0.90	0.93	181

5	0.94	0.79	0.86	182
6	0.96	0.97	0.97	181
7	0.88	0.94	0.91	179
8	0.47	0.69	0.56	174
9	0.62	0.78	0.69	180
accuracy			0.80	1797
macro avg	0.82	0.80	0.80	1797
weighted avg	0.82	0.80	0.80	1797

3 Clustering Hiérarchique



[20] : 0.09835207951870259

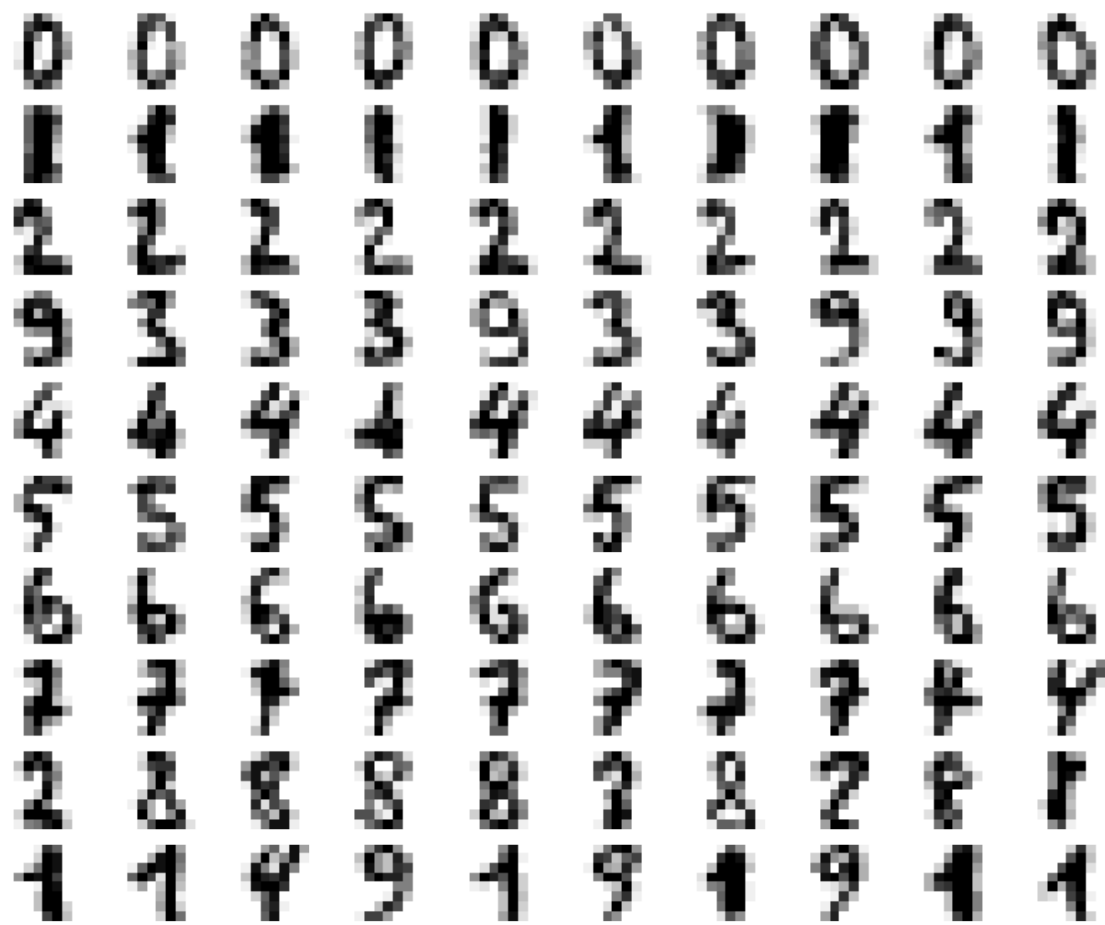
[21] : 0.06905571540674862

[22] : 0.01595605545383207

[28] : 0.1998430551922574

[29]: 0.1745470931891432

[9, 3, 5, 2, 8, 0, 6, 7, 4, 1]



Confusion matrix

```
[[176  0  0  0  2  0  0  0  0  0]
 [  0 105 24  0  0  1  2  0  0 50]
 [  1  3 141  7  0  0  0  2 22  1]
 [  0  1  0 162  0  2  0  9  9  0]
 [  0  4  0  0 127  0  3  3  5 39]
 [  0  0  0  13  1 166  1  0  0  1]
 [  1  3  0  0  1  1 175  0  0  0]
 [  0  0  0  0  0  0  0 156  2 21]
 [  0 18  1  4  0  1  1  1 137 11]
 [  0  0  0 143  0  2  0  3  4 28]]
```

Classification report

	precision	recall	f1-score	support
0	0.99	0.99	0.99	178

1	0.78	0.58	0.66	182
2	0.85	0.80	0.82	177
3	0.49	0.89	0.63	183
4	0.97	0.70	0.81	181
5	0.96	0.91	0.94	182
6	0.96	0.97	0.96	181
7	0.90	0.87	0.88	179
8	0.77	0.79	0.78	174
9	0.19	0.16	0.17	180
accuracy			0.76	1797
macro avg	0.79	0.76	0.77	1797
weighted avg	0.78	0.76	0.76	1797