資料科學 Report of Test 3

1106022 陳柏嘉

1. 首先觀察資料集有沒有 na 值，計算 na 比例 & 補植

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
****************************************
```

Age 遺失比例：$\frac{177}{891} = 19.865\%$

先從 Name 得到稱謂 (Mr. , Mrs. , Miss. , …)：

```
# 取出稱謂 Mr., Mrs., Miss.：'空格' + 字母 + '.'
df['Title'] = df['Name'].str.extract(' ([A-Za-z]+)\.', expand = False)
```

再補上各稱謂範圍裡年齡的中位數：

```
# 計算個別 Title 的年齡中位數，補值
df['Age'].fillna(df.groupby('Title')['Age'].transform('median'),
inplace = True)
```

Cabin 遺失比例：$\frac{687}{891} = 77.104\%$

雖然遺失比例高達77%應該直接丟棄，但還是有23%的真實資料，且其實可以從 Pclass (票種) 資訊推測出船艙區域 (越高級的票會住在越高級的區域)。

因為只需大略知道船艙區域，所以取 Cabin 的第一個字母：

```
# 取出船艙的第一個字母，表示船艙所在的區域
df['Cabin'] = df['Cabin'].str[:1]
```

再根據各票種的船艙區域中位數補上 na 值：

```
# 計算個別 Pclass 的船艙區域中位數，補值
df['Cabin'].fillna(df.groupby('Pclass')['Cabin'].transform('median').as
type(int), inplace = True)
```

Embarked 遺失比例：$\frac{2}{891} = 0.224\%$

Embarked (登船地點) 因為比例不高，所以隨意挑一種方式補上就可以了：

```
# 'Embarked' 登船地點，補上與下一筆相同值，若最後一筆是 na 則補上與前一筆相同值
df['Embarked'].fillna(method='bfill', inplace=True)
df['Embarked'].fillna(method='pad', inplace=True)
```

2. 資料清理 & 刪減

將 SibSp (兄弟姐妹/配偶) 和 Parch (父母/孩子) 合併為 FamilyNum (家人數)：

```
# 合併 Sibsp & Parch = FamilyNum
df['FamilyNum'] = df['SibSp'] + df['Parch']
```

將不需要的資料丟棄：

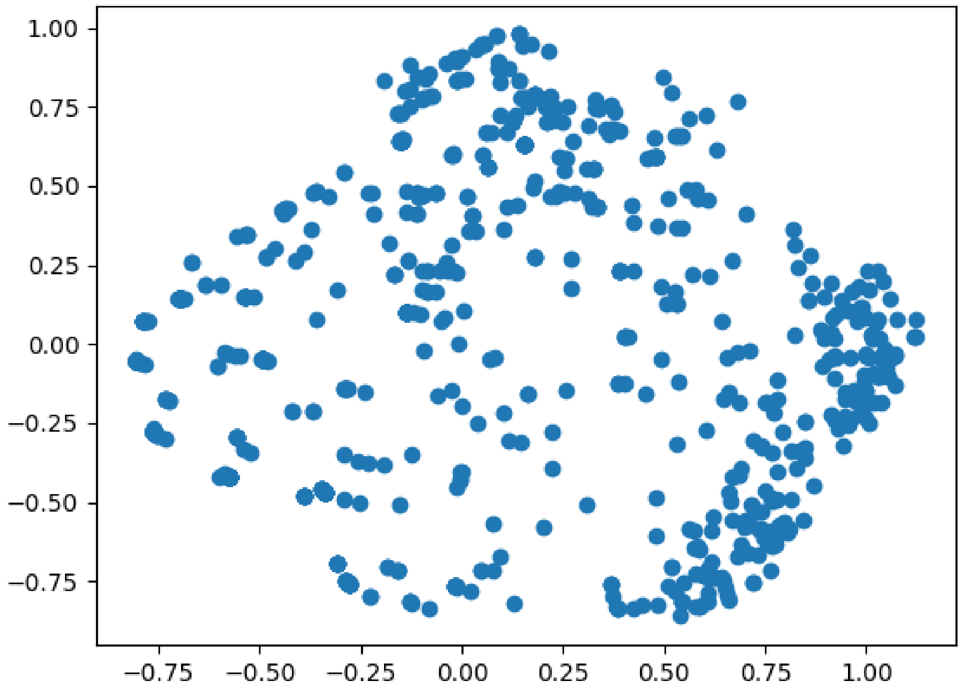
```
# 刪除 PassengerId, Name, SibSp, Parch, Ticket, Title
col_drop = ['PassengerId', 'Name', 'SibSp', 'Parch', 'Ticket', 'Title']
df = df.drop(col_drop, axis = 1)
```

整理後的樣子：

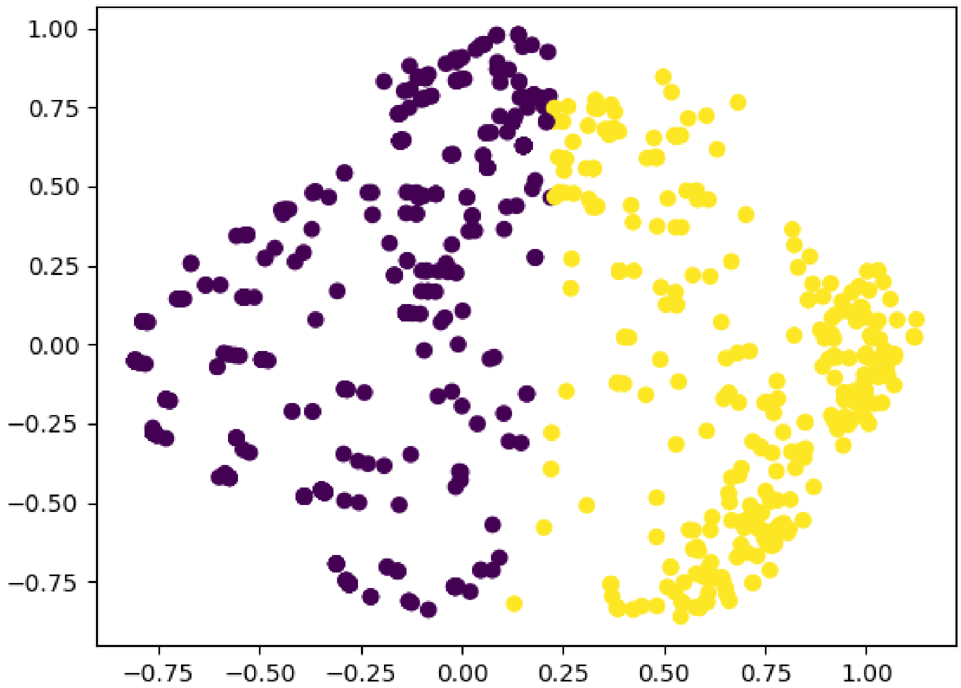
```
****************************************
     Survived  Pclass  Sex  Age     Fare  Cabin  Embarked  FamilyNum
0           0       3    0  2.0   7.2500    6.0         0          1
1           1       1    1  3.0  71.2833    3.0         1          1
2           1       3    1  2.0   7.9250    6.0         0          0
3           1       1    1  3.0  53.1000    3.0         0          1
4           0       3    0  3.0   8.0500    6.0         0          0
..        ...     ...  ...  ...      ...    ...       ...        ...
886         0       2    0  2.0  13.0000    5.5         0          0
887         1       1    1  1.0  30.0000    2.0         0          0
888         0       3    1  2.0  23.4500    6.0         0          3
889         1       1    0  2.0  30.0000    3.0         1          0
890         0       3    0  3.0   7.7500    6.0         2          0

[891 rows x 8 columns]
****************************************
```

3. 標準化 & 數據轉 2 維 (如 code 所示)

數據圖；



經過 Spectral Clustering 分類的圖 (n = 2)：

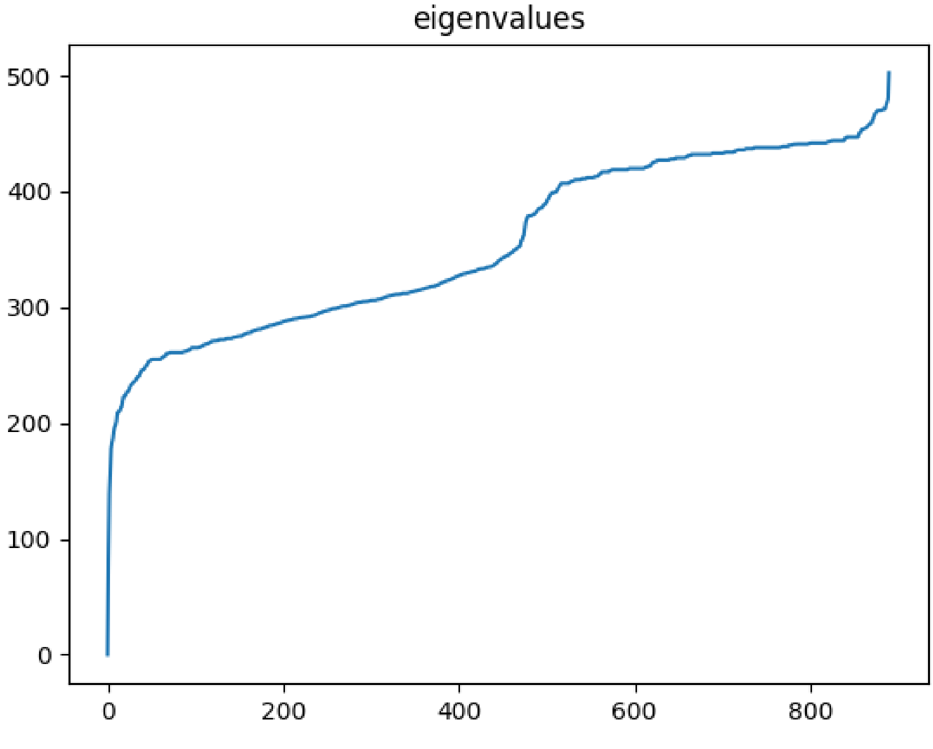


4. 計算 eigenvalues & eigenvectors，並將 eigenvalues 由大到小排序：

```
eigval:  [  0.        82.86428 139.1262  160.78013 178.01018 183.       186.61471
 192.82337 196.54681 198.33294 201.6839  209.       209.       209.95922
 210.83402 213.47722 214.       219.13769 223.       223.       223.6951
 225.29404 226.90413 227.26602 228.20247 230.       232.47352 232.96793
 233.84589 234.97919 235.41982 236.089   237.16837 239.       239.03986
 240.13362 240.80882 243.03831 244.22536 245.74786 245.95856 246.11667
 247.30779 248.29597 249.70971 250.40671 252.       253.42884 253.84752
 254.18614 254.93088 255.       255.       255.       255.       255.
 255.       255.       255.       255.       255.00458 255.50744 256.57179
 257.09425 257.43352 257.58892 258.94292 260.       260.       260.1709
 260.68006 261.       261.       261.       261.       261.       261.
 261.       261.       261.       261.       261.       261.       261.
 261.       261.       261.48176 261.66116 262.       262.       262.15309
 262.77259 262.99949 263.       263.5909  264.52728 265.       265.
 265.       265.       265.       265.       265.       265.01622 265.26572
 265.67857 265.90045 266.29985 266.62422 267.15654 267.798   267.97827
 268.       268.       268.57023 268.72292 269.13462 269.86789 270.09087
 270.68976 270.80268 271.       271.       271.       271.       271.09513
 271.17269 272.       272.       272.       272.       272.       272.
 272.00501 272.48757 273.       273.       273.       273.       273.
 273.       273.14564 273.34339 273.88781 274.10385 274.12915 274.27429
 274.4166  274.68907 275.       275.       275.       275.       275.20585
 276.12668 276.34546 276.84748 277.0917  277.60319 277.68489 278.
 278.       278.17048 278.22578 278.89448 279.64503 279.71812 280.
 280.20932 280.42294 280.76355 280.92126 281.       281.       281.17398
 281.27775 281.73238 281.95004 282.06412 282.52434 282.79859 283.
 283.12431 283.90038 284.       284.10561 284.19281 284.34474 284.5163
 284.76346 285.02783 285.1985  285.55019 285.92892 285.98305 286.21953
 286.40839 286.67953 286.90755 287.3486  287.52464 287.74959 287.79774
 288.       288.35132 288.58447 288.74108 288.89525 289.       289.26017
 289.35095 289.68411 290.       290.       290.       290.17899 290.3176
 290.58891 290.70529 290.82324 291.       291.       291.07593 291.25718
 291.38057 291.40178 291.50723 291.73871 291.79128 291.93672 291.99436
 292.04572 292.42295 292.53326 292.88936 292.94725 293.02299 293.44967
 293.79255 294.32225 294.50789 295.       295.24878 295.40062 295.62712
 296.24502 296.55522 296.73432 296.88968 297.       297.13624 297.48765
 297.57904 297.94078 298.10769 298.54081 298.61999 298.82137 298.98052
 299.       299.04738 299.29759 299.6077  299.94638 300.15129 300.20833
 300.56343 300.74046 300.94382 301.       301.       301.       301.
 301.37364 301.52814 301.77337 301.83592 302.22577 302.4053  302.68019
 302.80672 303.25133 303.58918 303.59735 303.93686 304.       304.
 304.       304.07959 304.3301  304.71711 304.76899 304.95924 305.
 305.       305.       305.08151 305.30288 305.60251 305.84907 305.91953
 305.99517 306.       306.       306.       306.       306.26629 306.61188
 306.78969 306.8593  306.99487 307.       307.29043 307.86046 308.05315
 308.18682 308.57347 308.89227 309.23157 309.47013 309.56434 309.73744
 309.99607 310.0949  310.21494 310.41915 310.66437 310.77398 311.
 311.       311.00892 311.09023 311.24849 311.25769 311.35907 311.77181
 311.98812 312.       312.       312.       312.       312.       312.32314
 312.58716 312.76045 312.97537 313.21504 313.2765  313.4699  313.77252
 313.85198 314.1778  314.19427 314.52131 314.73001 314.82506 314.99611
```

```
315.3037  315.54164 315.74562 315.84297 316.02168 316.44018 316.66706
316.75924 317.      317.15025 317.58891 317.8408  317.99572 318.
318.      318.      318.4427  318.88203 319.      319.36419 319.49195
320.05178 320.53172 321.041   321.25847 321.42359 321.85684 322.30014
322.76386 323.00198 323.03207 323.39233 323.49796 323.80566 323.99414
324.49048 324.93311 325.1414  325.78176 326.10725 326.45343 326.63628
326.88651 327.37332 327.57398 327.70464 328.37264 328.41022 328.69377
329.      329.      329.07383 329.54173 329.62953 329.82851 330.
330.1363  330.54897 330.66499 330.96925 331.      331.      331.3791
331.78568 332.17098 332.32685 332.89042 332.95238 333.      333.
333.08427 333.19822 333.37798 333.65422 334.02182 334.48139 334.60108
334.84764 335.      335.      335.07984 335.67943 336.      336.16342
337.12635 337.4702  338.31315 339.2645  339.79545 340.46352 341.01191
341.72759 342.17415 342.59614 342.82647 343.63132 343.97006 344.33245
345.      345.      345.26948 346.38608 346.54522 347.22842 347.6195
348.08096 349.25026 349.6041  349.96131 350.71789 351.45172 351.84584
352.5495  353.10518 357.64109 358.15824 361.09038 362.      367.85868
373.05237 376.      377.72755 379.      379.      379.      379.
379.50247 380.      380.15989 380.51535 381.54242 381.8437  383.19971
384.24585 385.01    385.43793 385.50638 385.58056 386.61896 387.23031
388.58653 389.02792 389.91748 390.46841 393.      393.67234 395.81122
397.      398.      399.      399.      399.03133 399.43944 399.59845
399.9482  401.15287 402.70438 404.06497 404.9432  406.51387 407.
407.      407.      407.      407.      407.      407.      407.
407.      407.39312 408.44129 408.64344 409.      409.      409.33385
410.      410.      410.      410.      410.      410.      410.
410.15831 411.      411.      411.      411.      411.      411.
412.      412.      412.      412.      412.      412.      412.
412.01364 412.36377 413.      413.      413.      413.16751 413.6376
414.98094 415.15338 416.18324 417.      417.      417.      417.
417.      417.      417.      417.03889 417.11678 418.01752 418.2796
418.59854 419.      419.      419.      419.      419.      419.
419.      419.      419.      419.      419.      419.      419.
419.      419.      419.      419.      419.      419.51984 420.
420.      420.      420.      420.      420.      420.      420.
420.      420.      420.      420.      420.      420.      420.
420.      420.      420.17031 420.72971 421.      421.10585 421.42474
422.      422.      422.28862 422.72359 424.53419 425.      425.34441
425.75968 426.      426.03855 426.63543 426.99003 427.      427.
427.      427.      427.      427.      427.      427.      427.
427.      427.      427.      427.45977 428.      428.      428.
428.      428.      428.11263 429.      429.      429.      429.
429.      429.      429.      429.      429.      429.      429.08012
429.36331 429.72742 430.02938 431.      431.      431.04623 431.12707
432.      432.      432.      432.      432.      432.      432.
432.      432.      432.      432.      432.      432.      432.
432.      432.      432.      432.      432.      432.      432.
432.      432.04679 432.59243 433.      433.      433.      433.
433.      433.      433.      433.      433.      433.      433.
433.      433.      433.42769 433.69939 433.98115 434.      434.
434.      434.      434.      434.      434.      434.      434.
```

```
 435.       435.       435.34746 435.62755 435.98809 436.       436.
 436.       436.       436.       436.       436.       436.39259 437.
 437.       437.       437.       437.       437.       437.       437.
 437.       437.61857 438.       438.       438.       438.       438.
 438.       438.       438.       438.       438.       438.       438.
 438.       438.       438.       438.       438.       438.       438.
 438.       438.       438.       438.       438.       438.       438.
 438.       438.       438.       438.19277 438.31514 438.74915 439.
 439.       439.       439.       439.       439.       439.       440.
 440.       440.       440.07661 440.45461 440.70627 440.78258 441.
 441.       441.       441.       441.       441.       441.       441.
 441.       441.       441.       441.       441.       441.       441.
 441.29965 441.50536 441.90151 442.       442.       442.       442.
 442.       442.       442.       442.       442.       442.       442.
 442.       442.       442.       442.       442.       442.       442.
 442.45297 442.78334 442.98185 443.28518 443.38454 443.69806 443.88413
 444.       444.       444.       444.       444.       444.       444.
 444.       444.       444.       444.       444.       444.       444.26378
 446.1036  446.14319 447.       447.       447.       447.       447.
 447.       447.       447.       447.       447.       447.       447.
 447.       447.58907 450.47866 451.       452.       453.73646 454.
 454.       454.41896 454.5775  455.11143 455.91609 456.52873 457.79625
 458.47711 458.77132 458.92301 461.37697 462.87217 464.94384 467.10166
 468.       469.14203 469.97903 470.       470.       470.       470.07679
 470.44035 471.       471.       471.01137 471.91919 475.72019 477.51784
 480.44976 502.55623]
```



5. 列出最小的 nonzero eigenvalue & corresponding eigenvector：

```
The smallest nonzero eigenvalue and the corresponding eigenvector:
eigenvalue:
 82.86428
eigenvector:
 [[-0.034  0.053 -0.011  0.013  0.009  0.005 -0.017  0.     0.003 -0.
   0.007  0.009 -0.    -0.006 -0.001 -0.001 -0.     0.001 -0.    -0.
   0.    -0.    -0.001  0.    -0.     0.    -0.     0.     0.001 -0.001
  -0.     0.     0.001 -0.001  0.001 -0.001 -0.    -0.     0.001 -0.
  -0.003  0.004 -0.001  0.005 -0.007  0.     0.     0.    -0.007  0.003
   0.     0.     0.    -0.003 -0.004 -0.001 -0.    -0.    -0.005  0.005
   0.    -0.    -0.001  0.    -0.    -0.001  0.002  0.     0.    -0.001
   0.     0.     0.001  0.    -0.    -0.     0.    -0.     0.     0.
   0.001 -0.    -0.     0.     0.    -0.    -0.001 -0.001 -0.    -0.
   0.001 -0.    -0.002 -0.    -0.    -0.002  0.     0.    -0.001  0.002
  -0.002  0.002  0.    -0.001  0.001 -0.001  0.     0.001 -0.001 -0.
   0.001 -0.     0.     0.     0.     0.     0.    -0.    -0.    -0.
   0.    -0.    -0.     0.    -0.    -0.    -0.     0.     0.001  0.004
   0.001  0.001  0.013  0.009 -0.     0.003 -0.007 -0.002  0.003  0.007
  -0.    -0.008 -0.003 -0.001  0.005  0.004 -0.001 -0.     0.011 -0.
   0.     0.001 -0.005  0.004 -0.     0.005  0.005 -0.002 -0.002  0.001
   0.     0.002 -0.    -0.001  0.002  0.007 -0.    -0.    -0.001  0.003
   0.     0.     0.     0.001 -0.    -0.002  0.001 -0.001  0.     0.
  -0.    -0.    -0.    -0.     0.    -0.     0.    -0.    -0.     0.
   0.     0.     0.     0.003 -0.     0.     0.     0.    -0.     0.
  -0.001  0.     0.     0.001 -0.     0.    -0.002 -0.004  0.016 -0.001
   0.    -0.004  0.005  0.011 -0.    -0.    -0.003  0.    -0.     0.
  -0.     0.    -0.     0.     0.     0.    -0.008 -0.004 -0.014  0.009
   0.021 -0.024  0.    -0.    -0.003  0.005 -0.007  0.002 -0.     0.
  -0.004 -0.    -0.    -0.     0.     0.     0.    -0.    -0.     0.002
  -0.015  0.    -0.001 -0.008  0.002  0.003  0.007 -0.02   0.018 -0.004
   0.016 -0.011  0.007  0.    -0.     0.    -0.    -0.     0.    -0.
  -0.     0.    -0.     0.002  0.003  0.01   0.007  0.006 -0.007 -0.007
   0.011  0.015 -0.007 -0.     0.012  0.013  0.007  0.007 -0.003 -0.
  -0.001 -0.     0.     0.     0.     0.    -0.     0.    -0.     0.
  -0.    -0.    -0.001  0.005  0.005 -0.003 -0.017  0.004  0.009  0.001
  -0.008  0.001 -0.006  0.009 -0.021  0.     0.     0.     0.003 -0.08
  -0.019 -0.     0.    -0.     0.     0.     0.     0.    -0.     0.
  -0.    -0.    -0.    -0.     0.    -0.005  0.008  0.004 -0.001  0.
   0.007  0.    -0.001 -0.02  -0.001  0.011 -0.     0.031 -0.007  0.032
   0.027  0.007  0.002 -0.     0.004 -0.011  0.     0.    -0.     0.
  -0.    -0.     0.    -0.     0.     0.    -0.    -0.    -0.    -0.
   0.     0.002  0.004  0.007 -0.     0.    -0.    -0.018 -0.009  0.012
   0.008 -0.029 -0.031 -0.01  -0.022 -0.009  0.014 -0.006  0.016  0.
  -0.    -0.    -0.     0.     0.     0.    -0.    -0.    -0.     0.
   0.    -0.    -0.    -0.     0.     0.004  0.    -0.    -0.004  0.001
   0.009 -0.007 -0.004 -0.001 -0.013 -0.002  0.006 -0.008 -0.002  0.002
   0.     0.    -0.065 -0.02   0.046 -0.005  0.011  0.003  0.013  0.006
  -0.027  0.001 -0.002 -0.018  0.003 -0.     0.     0.     0.003 -0.012
   0.008  0.001 -0.004  0.005  0.001  0.003  0.    -0.032 -0.009  0.023
   0.     0.    -0.     0.    -0.     0.018  0.    -0.    -0.     0.
  -0.     0.    -0.    -0.     0.     0.     0.    -0.    -0.001 -0.011
```

```
 -0.038  0.005 -0.001 -0.006 -0.009  0.001  0.002 -0.     0.007 -0.
 -0.    -0.     0.    -0.     0.     0.085  0.    -0.    -0.     0.029
 -0.031  0.04  -0.015 -0.     0.002  0.032  0.015 -0.012  0.004  0.007
  0.033 -0.009 -0.003 -0.     0.005 -0.004  0.    -0.004  0.002 -0.001
  0.007  0.006 -0.008 -0.002  0.008  0.001 -0.001  0.     0.     0.
 -0.017 -0.003  0.     0.009 -0.355 -0.    -0.    -0.     0.099 -0.
 -0.005 -0.175 -0.     0.    -0.003  0.009  0.005  0.     0.    -0.
 -0.    -0.186  0.172  0.     0.     0.26  -0.    -0.706  0.238 -0.181
  0.112 -0.042  0.004  0.02   0.026 -0.036 -0.054 -0.    -0.001 -0.001
  0.007 -0.009 -0.003 -0.004 -0.015 -0.004 -0.007  0.002  0.01  -0.004
 -0.01   0.003 -0.015  0.011 -0.01   0.013 -0.009 -0.004  0.012 -0.005
  0.009  0.012  0.007 -0.006 -0.013  0.01  -0.011  0.002  0.004 -0.
  0.    -0.004 -0.003 -0.009  0.011  0.004  0.     0.     0.     0.071
  0.044 -0.024  0.058  0.017 -0.004  0.006  0.007  0.005  0.     0.
  0.    -0.    -0.042 -0.002  0.032  0.011 -0.001  0.01  -0.033 -0.024
 -0.022 -0.007  0.003  0.06   0.027 -0.02  -0.005 -0.005 -0.037 -0.001
  0.002 -0.007  0.005  0.014  0.005 -0.003 -0.015  0.01   0.006 -0.002
  0.011 -0.011 -0.001 -0.007 -0.017 -0.     0.002 -0.     0.     0.026
  0.011 -0.007  0.009 -0.005  0.042  0.    -0.    -0.    -0.    -0.
  0.     0.    -0.    -0.     0.    -0.    -0.    -0.     0.     0.
 -0.    -0.    -0.    -0.    -0.009  0.    -0.054 -0.    -0.     0.048
 -0.017  0.006  0.004  0.023  0.005  0.     0.003 -0.003 -0.032 -0.001
  0.014  0.     0.    -0.     0.    -0.     0.    -0.    -0.    -0.
 -0.    -0.    -0.    -0.004  0.034  0.003  0.009 -0.007  0.003 -0.005
  0.03  -0.     0.     0.007  0.005 -0.002 -0.006 -0.013 -0.001 -0.024
 -0.008 -0.023  0.004 -0.035  0.     0.     0.     0.     0.    -0.
 -0.     0.     0.     0.     0.023 -0.003  0.     0.    -0.     0.
 -0.    -0.    -0.    -0.     0.     0.     0.     0.    -0.     0.
 -0.    -0.     0.    -0.     0.    -0.     0.    -0.    -0.     0.
  0.    -0.     0.     0.    -0.     0.    -0.     0.    -0.    -0.
 -0.     0.    -0.     0.    -0.     0.    -0.    -0.    -0.    -0.
  0.     0.    -0.    -0.     0.    -0.     0.    -0.    -0.     0.
  0.    -0.    -0.     0.     0.    -0.    -0.     0.    -0.     0.
  0.     0.    -0.    -0.    -0.     0.     0.     0.     0.    -0.
 -0.     0.     0.    -0.     0.     0.    -0.     0.    -0.    -0.
 -0.     0.     0.    -0.    -0.     0.    -0.    -0.    -0.     0.
  0.    -0.     0.    -0.    -0.    -0.    -0.    -0.     0.     0.
  0.     0.    -0.    -0.    -0.     0.    -0.     0.     0.    -0.
 -0.    -0.    -0.     0.    -0.    -0.     0.     0.    -0.     0.
  0.    -0.    -0.    -0.     0.    -0.    -0.    -0.     0.     0.
 -0.     0.    -0.    -0.    -0.    -0.     0.    -0.     0.     0.
 -0.    -0.    -0.    -0.    -0.    -0.     0.     0.    -0.     0.
 -0.   ]]
```

Code 與註解：

```
import pandas as pd
import numpy as np
from sklearn.cluster import SpectralClustering
from sklearn.preprocessing import StandardScaler, normalize
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA

df = pd.read_csv('titanic/train.csv')

print(df.isnull().sum())

# 取出稱謂 Mr., Mrs., Miss.：'空格' + 字母 + '.'
df['Title'] = df['Name'].str.extract(' ([A-Za-z]+)\.', expand = False)

def title_map(x):   # title mapping 'Mr' = 0 , 'Miss' = 1 , 'Mrs' = 2
and other = 3
    title = 0
    if x == 'Mr':
        title = 0
    elif x == 'Miss' :
        title = 1
    elif x == 'Mrs':
        title = 2
    else:
        title = 3
    return title
df['Title'] = df['Title'].map(title_map)
# Name 已轉為 Title 紀錄，其餘部分與模型訓練無關
df.drop('Name', axis = 1 , inplace = True)

# Sex Mapping
sex_mapping = { 'male': 0 , 'female': 1 }
df['Sex'] = df['Sex'].map(sex_mapping)

# 計算個別 Title 的年齡中位數，補值
df['Age'].fillna(df.groupby('Title')['Age'].transform('median'),
inplace = True)
```

```python
# Age Mapping，取十分位
# print(df.Age.max()) # MAX 值為 80
df.loc[ df['Age'] < 10 , 'Age' ] = 0
df.loc[ (df['Age'] >= 10) & (df['Age'] < 20) , 'Age' ] = 1
df.loc[ (df['Age'] >= 20) & (df['Age'] < 30) , 'Age' ] = 2
df.loc[ (df['Age'] >= 30) & (df['Age'] < 40) , 'Age' ] = 3
df.loc[ (df['Age'] >= 40) & (df['Age'] < 50) , 'Age' ] = 4
df.loc[ (df['Age'] >= 50) & (df['Age'] < 60) , 'Age' ] = 5
df.loc[ (df['Age'] >= 60) & (df['Age'] < 70) , 'Age' ] = 6
df.loc[ (df['Age'] >= 70) & (df['Age'] < 80) , 'Age' ] = 7
df.loc[ (df['Age'] >= 80) , 'Age'] = 8

# 'Embarked' 登船地點，補上與下一筆相同值，若最後一筆是 na 則補上與前一筆相同
值
df['Embarked'].fillna(method='bfill', inplace=True)
df['Embarked'].fillna(method='pad', inplace=True)

# Embarked Mapping
embarked_mapping = { 'S': 0 , 'C': 1 , 'Q': 2 }
df['Embarked'] = df['Embarked'].map(embarked_mapping)

# 取出船艙的第一個字母，表示船艙所在的區域
df['Cabin'] = df['Cabin'].str[:1]

# Cabin Mapping
cabin_mapping = { 'A': 0 , 'B': 2 , 'C': 3 , 'D': 4 , 'E': 5 ,
                  'F': 6 , 'G': 7 , 'T': 8 }
df['Cabin'] = df['Cabin'].map(cabin_mapping)

# 計算個別 Pclass 的船艙區域中位數，補值
df['Cabin'].fillna(df.groupby('Pclass')['Cabin'].transform('median'),
inplace = True)

# 合併 Sibsp & Parch = FamilyNum
df['FamilyNum'] = df['SibSp'] + df['Parch']

# 刪除 PassengerId, SibSp, Parch, Ticket, Title
```

```python
col_drop = ['PassengerId', 'SibSp', 'Parch', 'Ticket', 'Title']
df = df.drop(col_drop, axis = 1)

print('*************************************')
print(df.head(3))
print('*************************************')

target = df['Survived']
dataset = df.drop('Survived', axis = 1)

X = dataset.values
Y = target.values

# 平均 & 變異數標準化
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_normalized = normalize(X_scaled)

# 主成分分析，數據轉 2 維
pca = PCA(n_components = 2)
X_2 = pca.fit_transform(X_normalized)

# 座標圖
plt.scatter(X_2[:,0], X_2[:,1])
plt.show()

# 建立 SpectralClustering 模型
sc = SpectralClustering(n_clusters=2)
sc.fit(X_2, Y)
labels = sc.labels_
plt.scatter(X_2[:,0], X_2[:,1], c=labels)
plt.show()



from sklearn.metrics import pairwise_distances

# 計算各點之間的距離，距離 < mean 才視為相連，生成 A 矩陣
A = pairwise_distances(X_normalized, metric='euclidean')
```

```python
A_mean = A.mean()
print('\nA mean:', A_mean)
vectorizer = np.vectorize(lambda x: 1 if (x > 0) & (x < A_mean) else 0)
A = np.vectorize(vectorizer)(A)
print('A:', A)

# Laplacian Matrix
from scipy.sparse import csgraph
L = csgraph.laplacian(A, normed=False)
print('L: ', L)

# 計算 eigenvalues & eigenvectors
eigval, eigvec = np.linalg.eig(L)
eigval = eigval.astype(float).round(5)
eigvec = eigvec.astype(float).round(5)
eigval = np.sort(eigval)
print('eigval: ', eigval)
# print('eigvec: ', eigvec)

# 取得最小的 nonzero eigenvalue
minval = np.min(eigval[np.nonzero(eigval)])

# 取得 corresponding eigenvector
def near(a, b, rtol = 1e-5, atol = 1e-8):
    return np.abs(a - b) < (atol + rtol * np.abs(b))

nonozero_eigvec = eigvec[near(eigval, minval)].astype(float).round(3)

print('\nThe smallest nonzero eigenvalue and the corresponding
eigenvector:')
print('eigenvalue: \n', minval)
print('eigenvector: \n',nonozero_eigvec)

plt.plot(eigval)
plt.title('eigenvalues')
plt.show()
```