run-me

September 3, 2020

0.1 About this lab

This lab can be used to analyse *Flags* dataset and answer questions in *Final Project* quiz. Everything is ready: file `flags.csv` is already in your working dir and `pandas` is already installed. Let's go!

0.1.1 About Flags dataset

Source: https://archive.ics.uci.edu/ml/datasets/Flags

Creators: Collected primarily from the "Collins Gem Guide to Flags": Collins Publishers (1986).

Donor: Richard S. Forsyth.

```
In [4]: !ls

flags.csv  run-me.ipynb
```

```
In [2]: import pandas as pd
        import scipy.stats as stats

In [3]: df = pd.read_csv("flags.csv")

In [8]: df
```

Out[8]:

	name	landmass	zone	area	population	language	\
0	Afghanistan	5	1	648	16	10	
1	Albania	3	1	29	3	6	
2	Algeria	4	1	2388	20	8	
3	American-Samoa	6	3	0	0	1	
4	Andorra	3	1	0	0	6	
5	Angola	4	2	1247	7	10	
6	Anguilla	1	4	0	0	1	
7	Antigua-Barbuda	1	4	0	0	1	
8	Argentina	2	3	2777	28	2	
9	Argentine	2	3	2777	28	2	
10	Australia	6	2	7690	15	1	
11	Austria	3	1	84	8	4	
12	Bahamas	1	4	19	0	1	
13	Bahrain	5	1	1	0	8	

14	Bangladesh	5	1	143	90	6
15	Barbados	1	4	0	0	1
16	Belgium	3	1	31	10	6
17	Belize	1	4	23	0	1
18	Benin	4	1	113	3	3
19	Bermuda	1	4	0	0	1
20	Bhutan	5	1	47	1	10
21	Bolivia	2	3	1099	6	2
22	Botswana	4	2	600	1	10
23	Brazil	2	3	8512	119	6
24	British-Virgin-Isles	1	4	0	0	1
25	Brunei	5	1	6	0	10
26	Bulgaria	3	1	111	9	5
27	Burkina	4	4	274	7	3
28	Burma	5	1	678	35	10
29	Burundi	4	2	28	4	10
..	...	...	...	...	...	...
164	Swaziland	4	2	17	1	10
165	Sweden	3	1	450	8	6
166	Switzerland	3	1	41	6	4
167	Syria	5	1	185	10	8
168	Taiwan	5	1	36	18	7
169	Tanzania	4	2	945	18	10
170	Thailand	5	1	514	49	10
171	Togo	4	1	57	2	3
172	Tonga	6	2	1	0	10
173	Trinidad-Tobago	2	4	5	1	1
174	Tunisia	4	1	164	7	8
175	Turkey	5	1	781	45	9
176	Turks-Cocos-Islands	1	4	0	0	1
177	Tuvalu	6	2	0	0	1
178	UAE	5	1	84	1	8
179	Uganda	4	1	236	13	10
180	UK	3	4	245	56	1
181	Uruguay	2	3	178	3	2
182	US-Virgin-Isles	1	4	0	0	1
183	USA	1	4	9363	231	1
184	USSR	5	1	22402	274	5
185	Vanuatu	6	2	15	0	6
186	Vatican-City	3	1	0	0	6
187	Venezuela	2	4	912	15	2
188	Vietnam	5	1	333	60	10
189	Western-Samoa	6	3	3	0	1
190	Yugoslavia	3	1	256	22	6
191	Zaire	4	2	905	28	10
192	Zambia	4	2	753	6	10
193	Zimbabwe	4	2	391	8	10

	religion	bars	stripes	colours	...	saltires	quarters	sunstars	\
0	2	0	3	5	...	0	0	1	
1	6	0	0	3	...	0	0	1	
2	2	2	0	3	...	0	0	1	
3	1	0	0	5	...	0	0	0	
4	0	3	0	3	...	0	0	0	
5	5	0	2	3	...	0	0	1	
6	1	0	1	3	...	0	0	0	
7	1	0	1	5	...	0	0	1	
8	0	0	3	2	...	0	0	0	
9	0	0	3	3	...	0	0	1	
10	1	0	0	3	...	1	1	6	
11	0	0	3	2	...	0	0	0	
12	1	0	3	3	...	0	0	0	
13	2	0	0	2	...	0	0	0	
14	2	0	0	2	...	0	0	0	
15	1	3	0	3	...	0	0	0	
16	0	3	0	3	...	0	0	0	
17	1	0	2	8	...	0	0	0	
18	5	0	0	2	...	0	0	1	
19	1	0	0	6	...	1	1	0	
20	3	0	0	4	...	0	0	0	
21	0	0	3	3	...	0	0	0	
22	5	0	5	3	...	0	0	0	
23	0	0	0	4	...	0	0	22	
24	1	0	0	6	...	1	1	0	
25	2	0	0	4	...	0	0	0	
26	6	0	3	5	...	0	0	1	
27	5	0	2	3	...	0	0	1	
28	3	0	0	3	...	0	1	14	
29	5	0	0	3	...	1	0	3	
..	...	...	...	...	...	...	...	...	
164	1	0	5	7	...	0	0	0	
165	1	0	0	2	...	0	0	0	
166	1	0	0	2	...	0	0	0	
167	2	0	3	4	...	0	0	2	
168	3	0	0	3	...	0	1	1	
169	5	0	0	4	...	0	0	0	
170	3	0	5	3	...	0	0	0	
171	7	0	5	4	...	0	1	1	
172	1	0	0	2	...	0	1	0	
173	1	0	0	3	...	0	0	0	
174	2	0	0	2	...	0	0	1	
175	2	0	0	2	...	0	0	1	
176	1	0	0	6	...	1	1	0	
177	1	0	0	5	...	1	1	9	
178	2	1	3	4	...	0	0	0	
179	5	0	6	5	...	0	0	0	

180	1	0	0	3	...	1	0	0
181	0	0	9	3	...	0	1	1
182	1	0	0	6	...	0	0	0
183	1	0	13	3	...	0	1	50
184	6	0	0	2	...	0	0	1
185	1	0	0	4	...	0	0	0
186	0	2	0	4	...	0	0	0
187	0	0	3	7	...	0	0	7
188	6	0	0	2	...	0	0	1
189	1	0	0	3	...	0	1	5
190	6	0	3	4	...	0	0	1
191	5	0	0	4	...	0	0	0
192	5	3	0	4	...	0	0	0
193	5	0	7	5	...	0	0	1

	crescent	triangle	icon	animate	text	topleft	botright
0	0	0	1	0	0	black	green
1	0	0	0	1	0	red	red
2	1	0	0	0	0	green	white
3	0	1	1	1	0	blue	red
4	0	0	0	0	0	blue	red
5	0	0	1	0	0	red	black
6	0	0	0	1	0	white	blue
7	0	1	0	0	0	black	red
8	0	0	0	0	0	blue	blue
9	0	0	0	0	0	blue	blue
10	0	0	0	0	0	white	blue
11	0	0	0	0	0	red	red
12	0	1	0	0	0	blue	blue
13	0	0	0	0	0	white	red
14	0	0	0	0	0	green	green
15	0	0	1	0	0	blue	blue
16	0	0	0	0	0	black	red
17	0	0	1	1	1	red	red
18	0	0	0	0	0	green	green
19	0	0	1	1	0	white	red
20	0	0	0	1	0	orange	red
21	0	0	0	0	0	red	green
22	0	0	0	0	0	blue	blue
23	0	0	0	0	1	green	green
24	0	0	1	1	1	white	blue
25	0	1	1	1	1	white	gold
26	0	0	1	1	0	white	red
27	0	0	0	0	0	red	green
28	0	0	1	1	0	blue	red
29	0	0	0	0	0	white	white
..	...	...	...	...	...	...	...
164	0	0	1	0	0	blue	blue

```
165         0         0     0         0    0     blue      blue
166         0         0     0         0    0      red       red
167         0         0     0         0    0      red     black
168         0         0     0         0    0     blue       red
169         0         1     0         0    0    green      blue
170         0         0     0         0    0      red       red
171         0         0     0         0    0      red     green
172         0         0     0         0    0    white       red
173         0         1     0         0    0    white     white
174         1         0     0         0    0      red       red
175         1         0     0         0    0      red       red
176         0         0     1         1    0    white      blue
177         0         0     0         0    0    white      blue
178         0         0     0         0    0      red     black
179         0         0     0         1    0    black       red
180         0         0     0         0    0    white       red
181         0         0     0         0    0    white     white
182         0         0     1         1    1    white     white
183         0         0     0         0    0     blue       red
184         0         0     1         0    0      red       red
185         0         1     0         1    0    black     green
186         0         0     1         0    0     gold     white
187         0         0     1         1    0     gold       red
188         0         0     0         0    0      red       red
189         0         0     0         0    0     blue       red
190         0         0     0         0    0     blue       red
191         0         0     1         1    0    green     green
192         0         0     0         1    0    green     brown
193         0         1     1         1    0    green     green

[194 rows x 30 columns]
```

In [9]:
```
df.corr()
```

Out[9]:

	landmass	zone	area	population	language	religion	\
landmass	1.000000	-0.615759	0.013279	0.105940	0.496400	0.325073	
zone	-0.615759	1.000000	-0.025431	-0.124111	-0.618387	-0.320537	
area	0.013279	-0.025431	1.000000	0.519030	-0.027869	0.087941	
population	0.105940	-0.124111	0.519030	1.000000	0.062804	0.172731	
language	0.496400	-0.618387	-0.027869	0.062804	1.000000	0.422315	
religion	0.325073	-0.320537	0.087941	0.172731	0.422315	1.000000	
bars	-0.139290	0.085446	0.001243	-0.041392	-0.049862	-0.060000	
stripes	-0.063375	0.002359	0.028454	0.030100	0.105409	0.103222	
colours	-0.062856	0.173567	-0.162227	-0.115052	-0.197883	-0.017720	
red	0.093400	-0.032269	-0.005470	-0.024648	-0.011048	0.121721	
green	0.026226	0.069404	-0.081328	-0.047918	0.006046	0.128843	
blue	-0.177366	0.206066	-0.080439	-0.060222	-0.262201	-0.154891	
gold	-0.080465	0.164392	0.014500	-0.002184	-0.189429	0.023334	

```
white      -0.004134 -0.035296 -0.098135   -0.092640 -0.046823 -0.167605
black      -0.058243  0.026847 -0.119825   -0.103552  0.094465  0.074055
orange      0.020744  0.052239 -0.050718    0.050497 -0.064407 -0.051215
circles     0.065696 -0.042542 -0.005652    0.051257  0.098476  0.123260
crosses    -0.031103  0.101441 -0.057888   -0.076305 -0.210966 -0.225238
saltires   -0.003429  0.179667 -0.025492   -0.059061 -0.296106 -0.133373
quarters   -0.004537  0.153300 -0.001968   -0.041349 -0.220535 -0.135713
sunstars   -0.034843  0.121754  0.340572    0.191377 -0.126241 -0.075081
crescent    0.168438 -0.159272 -0.005049   -0.015544  0.193375  0.053145
triangle   -0.033154  0.060416 -0.075366   -0.078156 -0.064845  0.056868
icon        0.030359 -0.048692  0.026593    0.017966  0.028336  0.055714
animate    -0.118856  0.106034 -0.051539   -0.082259 -0.133756 -0.002741
text       -0.110743  0.037602  0.024423   -0.025194 -0.082977 -0.055158

                bars   stripes   colours       red    ...      circles  \
landmass   -0.139290 -0.063375 -0.062856  0.093400    ...     0.065696
zone        0.085446  0.002359  0.173567 -0.032269    ...    -0.042542
area        0.001243  0.028454 -0.162227 -0.005470    ...    -0.005652
population -0.041392  0.030100 -0.115052 -0.024648    ...     0.051257
language   -0.049862  0.105409 -0.197883 -0.011048    ...     0.098476
religion   -0.060000  0.103222 -0.017720  0.121721    ...     0.123260
bars        1.000000 -0.273371 -0.064574 -0.017093    ...    -0.096650
stripes    -0.273371  1.000000  0.038278  0.014210    ...    -0.087479
colours    -0.064574  0.038278  1.000000  0.311759    ...     0.178062
red        -0.017093  0.014210  0.311759  1.000000    ...     0.108639
green       0.116905 -0.085367  0.428388  0.031169    ...     0.011643
blue       -0.128510  0.126104  0.294779 -0.204010    ...     0.048217
gold        0.037118 -0.031987  0.500073  0.081769    ...     0.011643
white      -0.198697  0.125907  0.214334 -0.033483    ...     0.081854
black      -0.029076  0.091814  0.375794  0.142230    ...     0.079483
orange      0.046841 -0.106477  0.477655 -0.055783    ...     0.182707
circles    -0.096650 -0.087479  0.178062  0.108639    ...     1.000000
crosses    -0.144432 -0.190550  0.253827  0.102772    ...     0.030979
saltires   -0.140070 -0.190735  0.378810  0.122031    ...     0.151485
quarters   -0.150604  0.091941  0.187861  0.119914    ...     0.053062
sunstars   -0.114667  0.215486  0.031008  0.019726    ...     0.071881
crescent    0.021743 -0.010242 -0.173644 -0.091458    ...     0.054474
triangle   -0.132973 -0.005719  0.143252  0.025760    ...    -0.051356
icon       -0.002598 -0.107426  0.505614  0.097513    ...     0.145507
animate     0.016260 -0.157922  0.544576  0.133648    ...     0.205118
text        0.067709 -0.111564  0.398483  0.017507    ...     0.133001

             crosses  saltires  quarters  sunstars  crescent  triangle  \
landmass   -0.031103 -0.003429 -0.004537 -0.034843  0.168438 -0.033154
zone        0.101441  0.179667  0.153300  0.121754 -0.159272  0.060416
area       -0.057888 -0.025492 -0.001968  0.340572 -0.005049 -0.075366
population -0.076305 -0.059061 -0.041349  0.191377 -0.015544 -0.078156
language   -0.210966 -0.296106 -0.220535 -0.126241  0.193375 -0.064845
```

```
religion   -0.225238 -0.133373 -0.135713 -0.075081  0.053145  0.056868
bars       -0.144432 -0.140070 -0.150604 -0.114667  0.021743 -0.132973
stripes    -0.190550 -0.190735  0.091941  0.215486 -0.010242 -0.005719
colours     0.253827  0.378810  0.187861  0.031008 -0.173644  0.143252
red         0.102772  0.122031  0.119914  0.019726 -0.091458  0.025760
green      -0.043077  0.126624 -0.085607 -0.068738 -0.007137  0.039840
blue        0.273647  0.242192  0.289396  0.156916 -0.116522  0.006603
gold        0.010665  0.126624  0.009430 -0.019270 -0.096461  0.159207
white       0.191904  0.142195  0.197159  0.099590  0.037274 -0.011029
black      -0.114232 -0.073192 -0.154542 -0.069278 -0.148364  0.328242
orange      0.161914  0.239252  0.073556 -0.069191 -0.031026 -0.114468
circles     0.030979  0.151485  0.053062  0.071881  0.054474 -0.051356
crosses     1.000000  0.707599  0.421509  0.008528 -0.095344 -0.156367
saltires    0.707599  1.000000  0.462187  0.085256 -0.078406 -0.077259
quarters    0.421509  0.462187  1.000000  0.231980 -0.033034 -0.138260
sunstars    0.008528  0.085256  0.231980  1.000000  0.013965 -0.011677
crescent   -0.095344 -0.078406 -0.033034  0.013965  1.000000 -0.034188
triangle   -0.156367 -0.077259 -0.138260 -0.011677 -0.034188  1.000000
icon        0.144294  0.182112  0.073006 -0.078309 -0.091225 -0.062367
animate     0.139532  0.194235  0.093789  0.005640 -0.122981  0.021261
text        0.078399  0.162464  0.069320  0.029121 -0.073506 -0.066415

                icon   animate      text
landmass    0.030359 -0.118856 -0.110743
zone       -0.048692  0.106034  0.037602
area        0.026593 -0.051539  0.024423
population  0.017966 -0.082259 -0.025194
language    0.028336 -0.133756 -0.082977
religion    0.055714 -0.002741 -0.055158
bars       -0.002598  0.016260  0.067709
stripes    -0.107426 -0.157922 -0.111564
colours     0.505614  0.544576  0.398483
red         0.097513  0.133648  0.017507
green       0.190550  0.250157  0.243872
blue        0.118539  0.028249  0.106271
gold        0.380732  0.353249  0.243872
white       0.058387  0.049165  0.128483
black       0.130330  0.161053  0.030093
orange      0.293670  0.482238  0.212075
circles     0.145507  0.205118  0.133001
crosses     0.144294  0.139532  0.078399
saltires    0.182112  0.194235  0.162464
quarters    0.073006  0.093789  0.069320
sunstars   -0.078309  0.005640  0.029121
crescent   -0.091225 -0.122981 -0.073506
triangle   -0.062367  0.021261 -0.066415
icon        1.000000  0.448456  0.386365
animate     0.448456  1.000000  0.410678
```

```
        text        0.386365  0.410678  1.000000

        [26 rows x 26 columns]
```

```
In [9]: df.dtypes
```

```
Out[9]: name          object
        landmass       int64
        zone           int64
        area           int64
        population     int64
        language       int64
        religion       int64
        bars           int64
        stripes        int64
        colours        int64
        red            int64
        green          int64
        blue           int64
        gold           int64
        white          int64
        black          int64
        orange         int64
        mainhue       object
        circles        int64
        crosses        int64
        saltires       int64
        quarters       int64
        sunstars       int64
        crescent       int64
        triangle       int64
        icon           int64
        animate        int64
        text           int64
        topleft       object
        botright      object
        dtype: object
```

```
In [40]: df['mainhue'].head()
```

```
Out[40]: 0    green
         1      red
         2    green
         3     blue
         4     gold
         Name: mainhue, dtype: object
```

```
In [39]: df['religion'].head()
```

```
Out[39]: 0    2
         1    6
         2    2
         3    1
         4    0
         Name: religion, dtype: int64

In [10]: df['mainhue'].value_counts()

Out[10]: red       71
         blue      40
         green     31
         white     22
         gold      19
         black      5
         orange     4
         brown      2
         Name: mainhue, dtype: int64

In [11]: df['mainhue'].mode()

Out[11]: 0    red
         dtype: object

In [12]: df['circles'].value_counts()

Out[12]: 0    165
         1     27
         4      1
         2      1
         Name: circles, dtype: int64

In [25]: df.head()

Out[25]:             name  landmass  zone  area  population  language  religion  bars  \
         0     Afghanistan         5     1   648          16        10         2     0
         1         Albania         3     1    29           3         6         6     0
         2         Algeria         4     1  2388          20         8         2     2
         3  American-Samoa         6     3     0           0         1         1     0
         4         Andorra         3     1     0           0         6         0     3

            stripes  colours    ...     saltires  quarters  sunstars  crescent  \
         0        3        5    ...            0         0         1         0
         1        0        3    ...            0         0         1         0
         2        0        3    ...            0         0         1         1
         3        0        5    ...            0         0         0         0
         4        0        3    ...            0         0         0         0

            triangle  icon  animate text  topleft  botright
```

```
0         0     1        0    0    black     green
1         0     0        1    0      red       red
2         0     0        0    0    green     white
3         1     1        1    0     blue       red
4         0     0        0    0     blue       red

[5 rows x 30 columns]
```

```
In [26]: c = df['circles']
         c.head()
```

```
Out[26]: 0    0
         1    0
         2    0
         3    0
         4    0
         Name: circles, dtype: int64
```

```
In [29]: type(df["circles"])
```

```
Out[29]: pandas.core.series.Series
```

```
In [34]: df['circles'].shape
```

```
Out[34]: (194,)
```

```
In [38]: c_b = df[['circles', 'bars']]
```

```
In [36]: c_b.head()
```

```
Out[36]:    circles  bars
         0        0     0
         1        0     0
         2        0     2
         3        0     0
         4        0     3
```

```
In [42]: df[['circles', 'bars']].shape
```

```
Out[42]: (194, 2)
```

```
In [43]: zero_c = df[df['circles'] == 0]
```

```
In [70]: zero_c['circles'].head()
```

```
Out[70]: 0    0
         1    0
         2    0
         3    0
         4    0
         Name: circles, dtype: int64
```

```
In [57]: zero_c_b = df[['circles', 'bars']]

In [76]: zero_two = zero_c_b[(df['circles'] == 0) & (df['bars'] == 0)]

In [77]: zero_two
```

Out[77]:

	circles	bars
0	0	0
1	0	0
3	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
18	0	0
21	0	0
22	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
..	...	...
155	0	0
156	0	0
158	0	0
159	0	0
160	0	0
162	0	0
163	0	0
164	0	0
165	0	0
166	0	0

```
          167        0     0
          169        0     0
          170        0     0
          171        0     0
          172        0     0
          173        0     0
          175        0     0
          176        0     0
          177        0     0
          180        0     0
          181        0     0
          182        0     0
          183        0     0
          184        0     0
          185        0     0
          187        0     0
          188        0     0
          189        0     0
          190        0     0
          193        0     0

          [131 rows x 2 columns]
```

```
In [78]: df[['bars', 'crosses']].corr()
```

```
Out[78]:              bars   crosses
          bars     1.000000 -0.144432
          crosses -0.144432  1.000000
```

```
In [7]: df[['bars', 'crosses']].corr(method='spearman')
```

```
Out[7]:              bars   crosses
         bars     1.000000 -0.152588
         crosses -0.152588  1.000000
```

```
In [15]: df[['area', 'population']].corr(method='pearson').round(2)
```

```
Out[15]:             area  population
          area        1.00        0.52
          population  0.52        1.00
```