A1_GeineMA

November 21, 2021

:

[ ]:
```
surname = "   "  #

alp = "                    "
w = [1, 4, 21, 25, 34,  6, 44, 26, 13, 44, 38, 26, 4, 43,  4, 49, 46,
       17, 42, 29,  4,  9, 36, 34, 31, 22,  15, 30,  4, 19, 28, 28, 33]

d = dict(zip(alp, w))
variant =  sum([d[el] for el in surname.lower()]) % 40 + 1

print("    № 1,    5 -     : ", variant % 5 + 1)
```

```
    № 1,    5 -     :  1
```

1 1

(Corruption Perceptions Index, CPI) 1.

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

2. DataFrame

[ ]:
```
FILE = "data/CPI2020.xlsx"
df = pd.read_excel(FILE, header=2, sheet_name=1, index_col=0)
df.head()
```

[ ]:
```
            ISO3 Region  CPI score 2020  Rank 2020  Sources 2020  \
Country
Denmark      DNK  WE/EU              88          1             8
New Zealand  NZL     AP              88          1             8
Finland      FIN  WE/EU              85          3             8
Singapore    SGP     AP              85          3             9
Sweden       SWE  WE/EU              85          3             8

             Standard error 2020  CPI score 2019  Rank 2019  Sources 2019  \
Country
```

```
Denmark                 1.775809              87          1             8
New Zealand             1.479342              87          1             8
Finland                 1.748594              86          3             8
Singapore               1.203239              85          4             9
Sweden                  1.303953              85          4             8

             Standard error 2019  …  Standard error 2015  CPI score 2014  \
Country                           …
Denmark                 2.542474  …                  2.16            92.0
New Zealand             2.286490  …                  2.32            91.0
Finland                 2.924511  …                  1.77            89.0
Singapore               2.048400  …                  2.02            84.0
Sweden                  1.977693  …                  1.71            87.0

             Sources 2014  Standard error 2014  CPI Score 2013  Sources 2013  \
Country
Denmark               7.0                 2.04            91.0           7.0
New Zealand           7.0                 2.28            91.0           7.0
Finland               7.0                 2.05            89.0           7.0
Singapore             8.0                 1.75            86.0           9.0
Sweden                7.0                 3.41            89.0           7.0

             Standard error 2013  CPI Score 2012  Sources 2012  \
Country
Denmark                      2.2            90.0           7.0
New Zealand                  2.3            90.0           7.0
Finland                      1.7            90.0           7.0
Singapore                    2.3            87.0           9.0
Sweden                       2.3            88.0           7.0

             Standard error 2012
Country
Denmark                      2.0
New Zealand                  2.2
Finland                      3.0
Singapore                    2.1
Sweden                       1.9

[5 rows x 33 columns]
```

3. DataFrame

[ ]:
```
df = df.sort_index(ascending=False)
df.head()
```

[ ]:
```
          ISO3 Region  CPI score 2020  Rank 2020  Sources 2020  \
Country
```

Zimbabwe	ZWE	SSA	24	157	9
Zambia	ZMB	SSA	33	117	9
Yemen	YEM	MENA	15	176	7
Vietnam	VNM	AP	36	104	8
Venezuela	VEN	AME	15	176	8

Country	Standard error 2020	CPI score 2019	Rank 2019	Sources 2019
Zimbabwe	1.352586	24	158	9
Zambia	0.956605	34	113	9
Yemen	1.333599	15	177	7
Vietnam	1.808218	37	96	8
Venezuela	0.914430	16	173	8

Country	Standard error 2019	…	Standard error 2015	CPI score 2014
Zimbabwe	2.307745	…	4.54	21.0
Zambia	1.973667	…	2.55	38.0
Yemen	1.885994	…	2.66	19.0
Vietnam	2.841134	…	2.56	31.0
Venezuela	2.048858	…	2.64	19.0

Country	Sources 2014	Standard error 2014	CPI Score 2013	Sources 2013
Zimbabwe	8.0	4.18	21.0	8.0
Zambia	8.0	2.27	38.0	8.0
Yemen	6.0	2.37	18.0	6.0
Vietnam	8.0	2.55	31.0	8.0
Venezuela	7.0	1.69	20.0	7.0

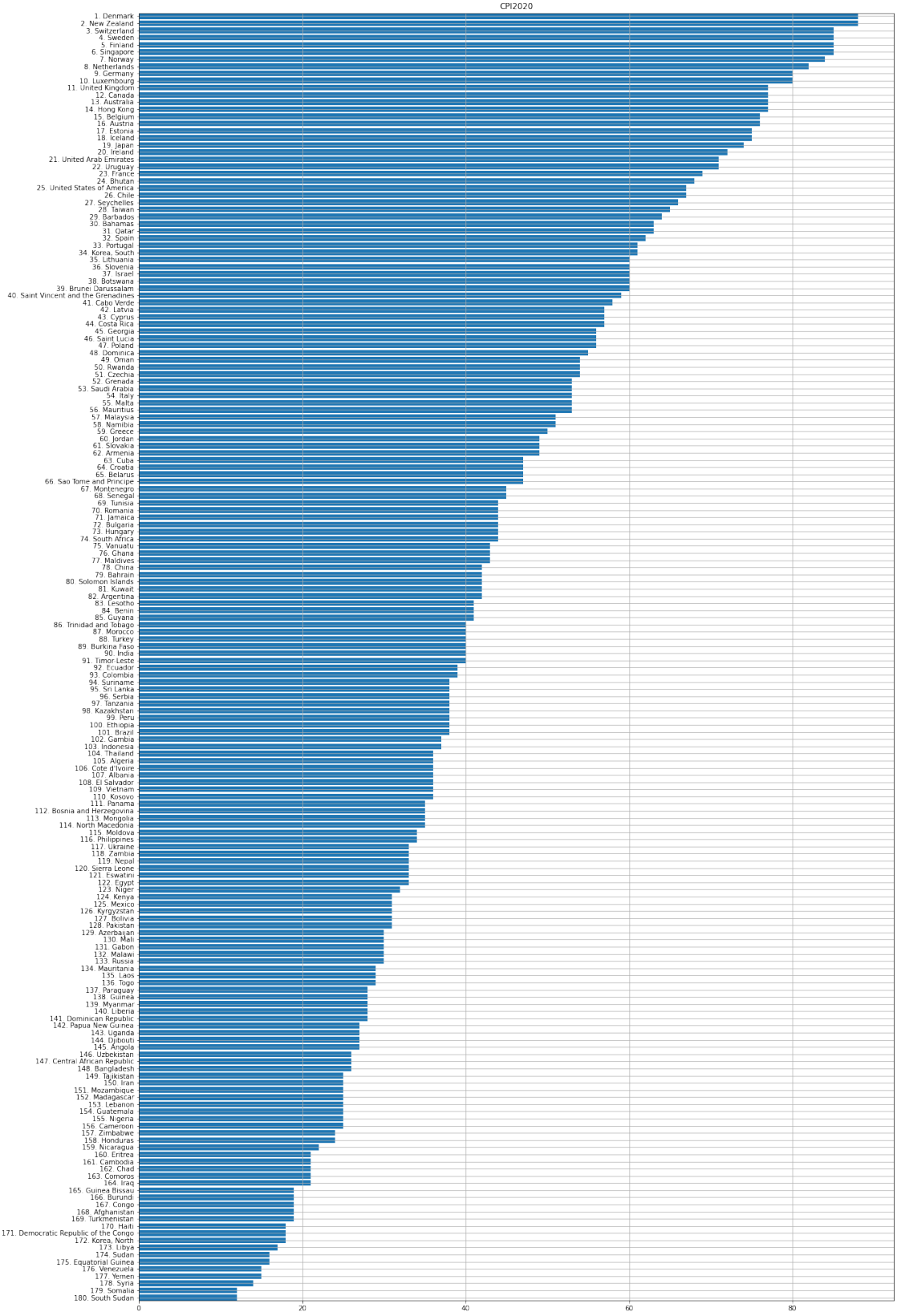
Country	Standard error 2013	CPI Score 2012	Sources 2012
Zimbabwe	4.3	20.0	8.0
Zambia	2.1	37.0	8.0
Yemen	2.6	23.0	6.0
Vietnam	2.6	31.0	8.0
Venezuela	2.3	19.0	7.0

Country	Standard error 2012
Zimbabwe	4.3
Zambia	3.1
Yemen	2.2
Vietnam	2.5
Venezuela	2.1

[5 rows x 33 columns]

4. CPI 2020

[ ]:
```
tmp = df.sort_values(by=['CPI score 2020'])
ys = tmp.index.to_series()
ys = ys.reset_index(drop=True)
for index, row in ys.iteritems():
    ys[index] = "{}. {}".format(180-index, row)
plt.figure("CPI2020", figsize=[20,35])
plt.subplot(1,1,1)
plt.title("CPI2020")
plt.barh(ys, tmp['CPI score 2020'])
plt.margins(y=0)
plt.grid(True)
```


CPI2020
1. Denmark
2. New Zealand
3. Switzerland
4. Sweden
5. Finland
6. Singapore
7. Norway
8. Netherlands
9. Germany
10. Luxembourg
11. United Kingdom
12. Canada
13. Australia
14. Hong Kong
15. Belgium
16. Austria
17. Estonia
18. Iceland
19. Japan
20. Ireland
21. United Arab Emirates
22. Uruguay
23. France
24. Bhutan
25. United States of America
26. Chile
27. Seychelles
28. Taiwan
29. Barbados
30. Bahamas
31. Qatar
32. Spain
33. Portugal
34. Korea, South
35. Lithuania
36. Slovenia
37. Israel
38. Botswana
39. Brunei Darussalam
40. Saint Vincent and the Grenadines
41. Cabo Verde
42. Latvia
43. Cyprus
44. Costa Rica
45. Georgia
46. Saint Lucia
47. Poland
48. Dominica
49. Oman
50. Rwanda
51. Czechia
52. Grenada
53. Saudi Arabia
54. Italy
55. Malta
56. Mauritius
57. Malaysia
58. Namibia
59. Greece
60. Jordan
61. Slovakia
62. Armenia
63. Cuba
64. Croatia
65. Belarus
66. Sao Tome and Principe
67. Montenegro
68. Senegal
69. Tunisia
70. Romania
71. Jamaica
72. Bulgaria
73. Hungary
74. South Africa
75. Vanuatu
76. Ghana
77. Maldives
78. China
79. Bahrain
80. Solomon Islands
81. Kuwait
82. Argentina
83. Lesotho
84. Benin
85. Guyana
86. Trinidad and Tobago
87. Morocco
88. Turkey
89. Burkina Faso
90. India
91. Timor-Leste
92. Ecuador
93. Colombia
94. Suriname
95. Sri Lanka
96. Serbia
97. Tanzania
98. Kazakhstan
99. Peru
100. Ethiopia
101. Brazil
102. Gambia
103. Indonesia
104. Thailand
105. Algeria
106. Cote d'Ivoire
107. Albania
108. El Salvador
109. Vietnam
110. Kosovo
111. Panama
112. Bosnia and Herzegovina
113. Mongolia
114. North Macedonia
115. Moldova
116. Philippines
117. Ukraine
118. Zambia
119. Nepal
120. Sierra Leone
121. Eswatini
122. Egypt
123. Niger
124. Kenya
125. Mexico
126. Kyrgyzstan
127. Bolivia
128. Pakistan
129. Azerbaijan
130. Mali
131. Gabon
132. Malawi
133. Russia
134. Mauritania
135. Laos
136. Togo
137. Paraguay
138. Guinea
139. Myanmar
140. Liberia
141. Dominican Republic
142. Papua New Guinea
143. Uganda
144. Djibouti
145. Angola
146. Uzbekistan
147. Central African Republic
148. Bangladesh
149. Tajikistan
150. Iran
151. Mozambique
152. Madagascar
153. Lebanon
154. Guatemala
155. Nigeria
156. Cameroon
157. Zimbabwe
158. Honduras
159. Nicaragua
160. Eritrea
161. Cambodia
162. Chad
163. Comoros
164. Iraq
165. Guinea Bissau
166. Burundi
167. Congo
168. Afghanistan
169. Turkmenistan
170. Haiti
171. Democratic Republic of the Congo
172. Korea, North
173. Libya
174. Sudan
175. Equatorial Guinea
176. Venezuela
177. Yemen
178. Syria
179. Somalia
180. South Sudan
0
20
40
60
80

5. DataFrame :
   - 1. Asia Pacific
6. DataFrame'a

[ ]:
```
ts = df[df.Region == 'AP']
ts.head()
```

[ ]:

Country	ISO3	Region	CPI score 2020	Rank 2020	Sources 2020	\
Vietnam	VNM	AP	36	104	8	
Vanuatu	VUT	AP	43	75	3	
Timor-Leste	TLS	AP	40	86	4	
Thailand	THA	AP	36	104	9	
Taiwan	TWN	AP	65	28	8	

Country	Standard error 2020	CPI score 2019	Rank 2019	Sources 2019	\
Vietnam	1.808218	37	96	8	
Vanuatu	3.587864	46	64	3	
Timor-Leste	2.707735	38	93	4	
Thailand	1.307877	36	101	9	
Taiwan	2.059654	65	28	8	

Country	Standard error 2019	…	Standard error 2015	CPI score 2014	\
		…			
Vietnam	2.841134	…	2.56	31.0	
Vanuatu	6.376219	…	NaN	NaN	
Timor-Leste	4.495380	…	5.17	28.0	
Thailand	2.448041	…	2.12	38.0	
Taiwan	3.034662	…	3.78	61.0	

Country	Sources 2014	Standard error 2014	CPI Score 2013	Sources 2013	\
Vietnam	8.0	2.55	31.0	8.0	
Vanuatu	NaN	NaN	NaN	NaN	
Timor-Leste	3.0	5.18	30.0	3.0	
Thailand	8.0	1.60	35.0	8.0	
Taiwan	7.0	4.17	61.0	7.0	

Country	Standard error 2013	CPI Score 2012	Sources 2012	\
Vietnam	2.6	31.0	8.0	
Vanuatu	NaN	NaN	NaN	
Timor-Leste	3.2	33.0	3.0	
Thailand	1.2	37.0	8.0	
Taiwan	4.3	61.0	7.0	

```
             Standard error 2012
Country
Vietnam                      2.5
Vanuatu                      NaN
Timor-Leste                  5.6
Thailand                     1.6
Taiwan                       3.9

[5 rows x 33 columns]
```

7. CPI 2012-2020 .

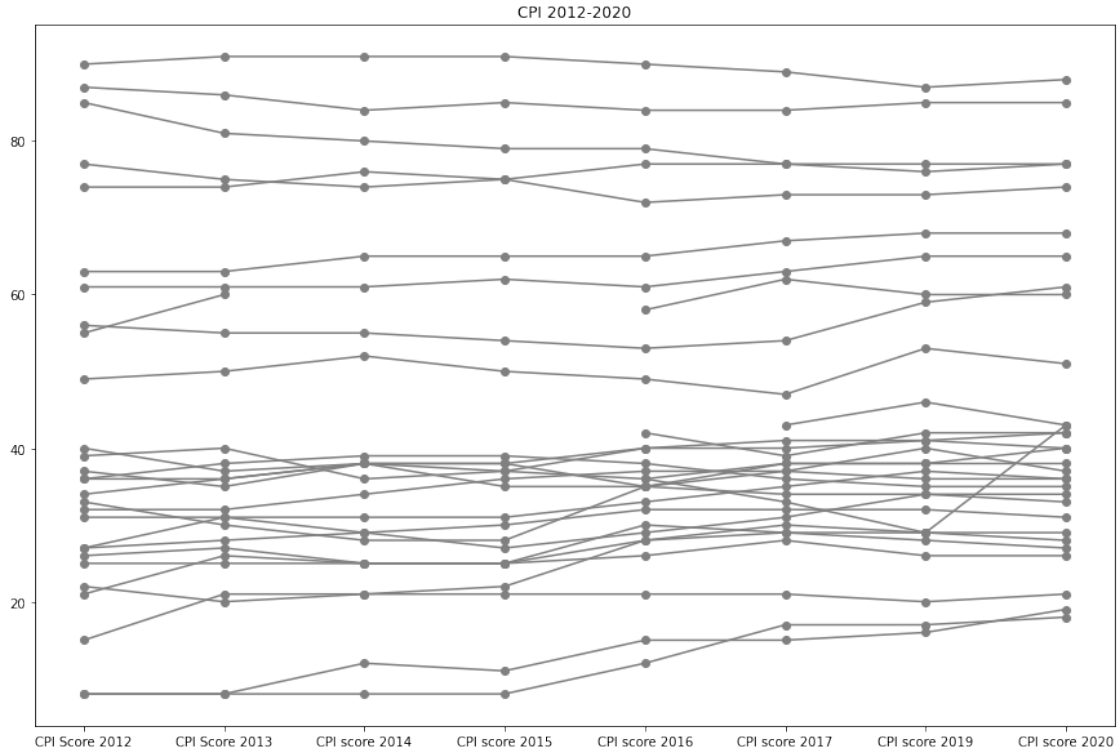
[ ]:
```
cols = ['CPI Score 2012', 'CPI Score 2013', 'CPI score 2014', 'CPI score 2015',␣
↪'CPI score 2016', 'CPI score 2017', 'CPI score 2019', 'CPI score 2020']
ts = ts[cols]
ts.head()
```

[ ]:
```
             CPI Score 2012  CPI Score 2013  CPI score 2014  CPI score 2015  \
Country
Vietnam                31.0            31.0            31.0            31.0
Vanuatu                 NaN             NaN             NaN             NaN
Timor-Leste            33.0            30.0            28.0            28.0
Thailand               37.0            35.0            38.0            38.0
Taiwan                 61.0            61.0            61.0            62.0

             CPI score 2016  CPI score 2017  CPI score 2019  CPI score 2020
Country
Vietnam                33.0              35              37              36
Vanuatu                 NaN              43              46              43
Timor-Leste            35.0              38              38              40
Thailand               35.0              37              36              36
Taiwan                 61.0              63              65              65
```

[ ]:
```
ts.transpose().plot(figsize=[15, 10], style='o-', title='CPI 2012-2020',␣
↪color='Grey', legend=False)
```

[ ]:
```
<AxesSubplot:title={'center':'CPI 2012-2020'}>
```



8. CPI 2020

[ ]:
```
max_cpi = ts.loc[ts['CPI score 2020'].idxmax()]
print('Max CPI 2020:')
print(max_cpi)
min_cpi = ts.loc[ts['CPI score 2020'].idxmin()]
print('Min CPI 2020:')
print(min_cpi)
```

```
Max CPI 2020:
CPI Score 2012    90.0
CPI Score 2013    91.0
CPI score 2014    91.0
CPI score 2015    91.0
CPI score 2016    90.0
CPI score 2017    89.0
CPI score 2019    87.0
CPI score 2020    88.0
Name: New Zealand, dtype: float64
Min CPI 2020:
CPI Score 2012     8.0
CPI Score 2013     8.0
CPI score 2014     8.0
CPI score 2015     8.0
```

```
CPI score 2016    12.0
CPI score 2017    17.0
CPI score 2019    17.0
CPI score 2020    18.0
Name: Korea, North, dtype: float64
```

9. 2012 2020

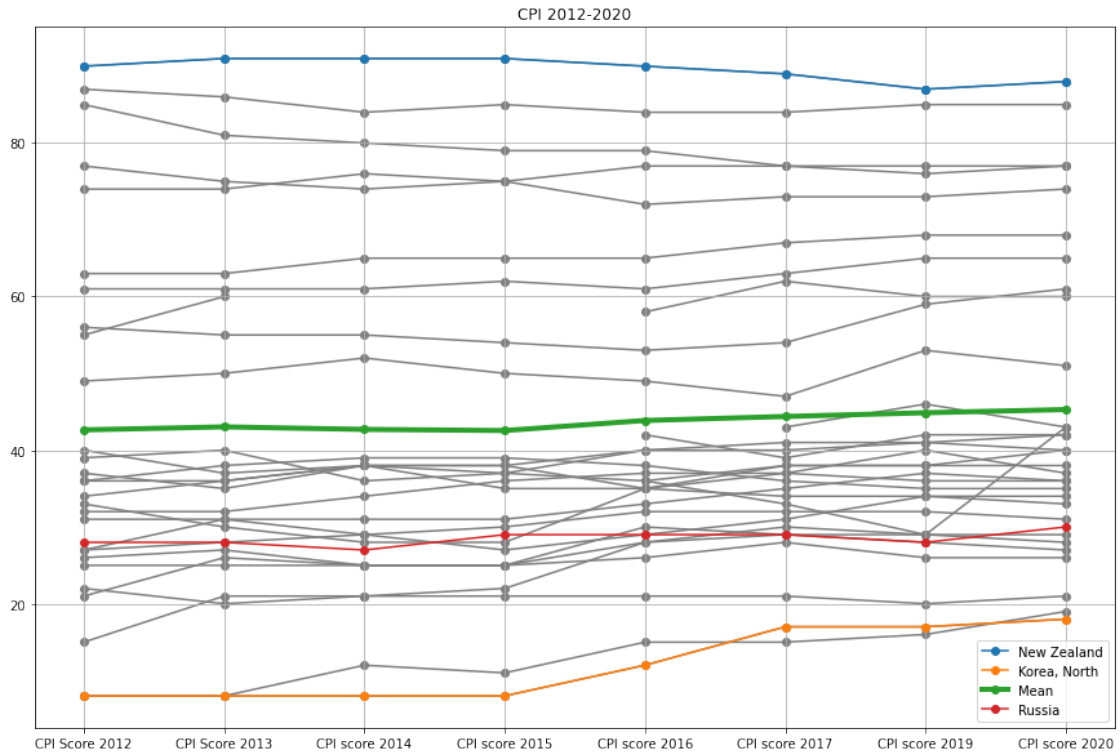
[ ]:
```
mean = ts.mean()
print('Mean AP CPI:')
print(mean)
```

```
Mean AP CPI:
CPI Score 2012    42.642857
CPI Score 2013    43.035714
CPI score 2014    42.703704
CPI score 2015    42.555556
CPI score 2016    43.866667
CPI score 2017    44.387097
CPI score 2019    44.870968
CPI score 2020    45.290323
dtype: float64
```

10. CPI 2012-2020 CPI 2020 ,

[ ]:
```
ax = max_cpi.plot(style='o-', legend=True)
min_cpi.plot(style='o-', legend=True, ax=ax)
mean.plot(style='o-', label='Mean', legend=True, linewidth=4, ax=ax)
df.loc['Russia'][cols].plot(style='o-', legend=True, grid=True, ax=ax)
ts.transpose().plot(title='CPI 2012-2020', style='o-', color='Grey',␣
↪figsize=[15, 10], legend=False, ax=ax, zorder=0, grid=True)
```

[ ]:
```
<AxesSubplot:title={'center':'CPI 2012-2020'}>
```



2 2

1. numpy, pandas folium

[ ]:
```
import folium
from folium.plugins import HeatMap
import pandas as pd
import numpy as np
from IPython.display import display
```

2. “data/data-54518-2021-10-18.xlsx” DataFrame.

[ ]:
```
FILE = 'data/data-54518-2021-10-18.xlsx'
df = pd.read_excel(FILE, header=0)
df.head()
```

```
/home/mike_geine/anaconda3/lib/python3.8/site-
packages/openpyxl/styles/stylesheet.py:221: UserWarning: Workbook contains no
default style, apply openpyxl's default
  warn("Workbook contains no default style, apply openpyxl's default")
```

[ ]:

	ShortName	FullName	LicensingAndAccreditation	EducationalServices	ReorganizationStatus	IDEKIS	Number	InstitutionsAddresses	INN	KPP	…	global_id	PublicPhone	Email	WebSite	EducationPrograms
0	« »	…	LicenseAvailability: \nLicenseSeries:77 01\nL…	NaN	13072	NaN	More then 1 level	7734077088	773401001	…	658727748	PublicPhone:(499) 197-51-23\n\n	Email:horoshevo@edu.mos.ru\n\n	dtim.mskobr.ru	|
1	« »	…	LicenseAvailability: \nLicenseSeries:77 01\nL…	NaN	13258	NaN	More then 1 level	7725663400	770901001	…	658727903	PublicPhone:(495) 678-55-62\n\n	Email:labp@edu.mos.ru\n\n	lab-putesh.mskobr.ru	#
2	« »	…	LicenseAvailability: \nLicenseSeries:77 01\nL…	NaN	11708	NaN	More then 1 level	7737057354	772401001	…	658727904	PublicPhone:(495) 391-23-33\n\n	Email:tsaritsino@edu.mos.ru\n\n	collegetsaritsyno.mskobr.ru	|
3	« « »	…	LicenseAvailability: \nLicenseSeries:77 01\nL…	NaN	13093	NaN	More then 1 level	7718191202	771801001	…	658727905	PublicPhone:(499) 161-15-06\n\n	Email:preobr@edu.mos.ru\n\n	dtdimvouo.mskobr.ru	|
4	№ 35	…	LicenseAvailability: \nLicenseSeries:77 01\nL…	NaN	13326	35.0	More then 1 level	7737014368	772401001	…	658727972	PublicPhone:(495) 327-50-91\n\n	Email:spo-35@edu.mos.ru\n\n	35finance.mskobr.ru	|

```
4                  #            …

                                                OrgType NumberofStudentsInOO  \
0                                                      NaN
1                                                                        NaN
2                                                  NaN
3                                                    NaN
4                                                  NaN

  TheContingentOfPreschoolersStudyingOO  \
0                                   NaN
1                                   NaN
2                                   NaN
3                                   NaN
4                                   NaN

                                        geodata_center geoarea
0  {"coordinates":[37.4796073910091,55.7838269797…      NaN
1  {"coordinates":[37.6765457430991,55.7396604854…      NaN
2  {"coordinates":[37.7158448802973,55.6218414311…      NaN
3  {"coordinates":[37.7330417729531,55.8026265640…      NaN
4  {"coordinates":[37.6678226722998,55.6049547310…      NaN

[5 rows x 25 columns]
```

3. - (LegalOrganization)

```
[ ]: df['LegalOrganization'].unique()
```

```
[ ]: array([' ', ' ', ' '], dtype=object)
```

4. 10

```
[ ]: df[df.LegalOrganization==' '].head(10)
```

```
[ ]:                                    ShortName  \
16                                «      »
17
25                                  № 24
36                                  № 11
43                          № 8  . . .
47                  «              »
89                                 № 548
153                               «    № 1518»
260                                  «   16»
351         № 1306 - «             »

                                        FullName  \
```

```
16                                …
17                                …
25                                …
36                                …
43                                …
47                                …
89                                …
153                               …
260                               …
351                               …

                             LicensingAndAccreditation  \
16   LicenseAvailability: \nLicenseSeries:77 01\nL…
17   LicenseAvailability: \nLicenseSeries:77 01\nL…
25   LicenseAvailability: \nLicenseSeries:77 01\nL…
36   LicenseAvailability: \nLicenseSeries:    …
43   LicenseAvailability: \nLicenseSeries:    …
47   LicenseAvailability: \nLicenseSeries:    …
89   LicenseAvailability: \nLicenseSeries:77 01\nL…
153  LicenseAvailability: \nLicenseSeries:77 01\nL…
260  LicenseAvailability: \nLicenseSeries:77 01\nL…
351  LicenseAvailability: \nLicenseSeries:77 01\nL…

                                  EducationalServices ReorganizationStatus  \
16                                                NaN
17                                                NaN
25                                                NaN
36                                                NaN
43                                                NaN
47                                                NaN
89   FullTimeEdu: \nPartTimeEdu: \nFullPartTimeEd…
153  FullTimeEdu: \nPartTimeEdu: \nFullPartTimeEd…
260  FullTimeEdu: \nPartTimeEdu: \nFullPartTimeEd…
351  FullTimeEdu: \nPartTimeEdu: \nFullPartTimeEd…

     IDEKIS  Number InstitutionsAddresses         INN        KPP  …  \
16    13175     NaN     More then 1 level  7734570529  773401001  …
17    13070     NaN     More then 1 level  7725618950  770401001  …
25    13096    24.0     More then 1 level  7719286023  771901001  …
36    12887    11.0     More then 1 level  7743085712  774301001  …
43    12017     8.0     More then 1 level  7714805691  771401001  …
47    14494     NaN     More then 1 level  7714239823  771301001  …
89    13866   548.0     More then 1 level  7737027335  772401001  …
153   13863  1518.0     More then 1 level  7717666051  771701001  …
260   13761     NaN     More then 1 level  7717025259  771701001  …
351   13930  1306.0     More then 1 level  7729413009  772901001  …
```

```
      global_id                      PublicPhone  \
16    908561038  PublicPhone:(499) 198-09-92\n\n
17    911869478  PublicPhone:(499) 242-27-82\n\n
25   1056398232  PublicPhone:(499) 461-52-19\n\n
36   1056398741  PublicPhone:(495) 456-44-01\n\n
43   1092464718  PublicPhone:(495) 640-60-58\n\n
47   1137248095  PublicPhone:(499) 976-59-89\n\n
89   1139420809  PublicPhone:(495) 398-83-32\n\n
153  1139420873  PublicPhone:(495) 687-66-66\n\n
260  1139420982  PublicPhone:(495) 682-62-34\n\n
351  1139421076  PublicPhone:(495) 932-99-58\n\n

                             Email              WebSite  \
16   Email:domisolka@edu.mos.ru\n\n       cdtt.mskobr.ru
17         Email:cpm@edu.mos.ru\n\n      cpm.dogm.mos.ru
25      Email:spo-24@edu.mos.ru\n\n  kollege24.mskobr.ru
36      Email:spo-11@edu.mos.ru\n\n       kp11.mskobr.ru
43       Email:spo-8@edu.mos.ru\n\n       pk-8.mskobr.ru
47      Email:mcrkpo@edu.mos.ru\n\n            mcrkpo.ru
89         Email:548@edu.mos.ru\n\n     cou548.mskobr.ru
153       Email:1518@edu.mos.ru\n\n    gum1518.mskobr.ru
260         Email:16@edu.mos.ru\n\n   schisv16.mskobr.ru
351       Email:1306@edu.mos.ru\n\n   gymg1306.mskobr.ru

                                    EducationPrograms  \
16                  |           …
17                  |           …
25                 #            …
36                 #            …
43                   |          …
47                   |          …
89              |               # …
153             |               # …
260             |               # …
351             |               # …

                                                OrgType  \
16
17                             …
25
36
43
47                             …
89
153
260
351
```

```
                        NumberofStudentsInOO  \
16                                       NaN
17                                       NaN
25                                       NaN
36                                       NaN
43                                       NaN
47                                       NaN
89   PupilsQuantity:5359\nDate:08.09.2021\n\n
153  PupilsQuantity:1521\nDate:08.09.2021\n\n
260  PupilsQuantity:1059\nDate:08.09.2021\n\n
351   PupilsQuantity:785\nDate:08.09.2021\n\n

       TheContingentOfPreschoolersStudyingOO  \
16                                       NaN
17                                       NaN
25                                       NaN
36                                       NaN
43                                       NaN
47                                       NaN
89   PupilsQuantity:420\nDate:08.09.2021\n\n
153  PupilsQuantity:300\nDate:08.09.2021\n\n
260  PupilsQuantity:169\nDate:08.09.2021\n\n
351  PupilsQuantity:390\nDate:08.09.2021\n\n

                                     geodata_center geoarea
16   {"coordinates":[37.5004519489548,55.7781030172…      NaN
17   {"coordinates":[37.576293734056,55.71734826701…      NaN
25   {"coordinates":[37.7989790586712,55.7478148000…      NaN
36   {"coordinates":[37.520325559775,55.84481608033…      NaN
43   {"coordinates":[37.5684471299924,55.7963416512…      NaN
47   {"coordinates":[37.5725054520134,55.7604996906…      NaN
89   {"coordinates":[37.7024293991346,55.6235702947…      NaN
153  {"coordinates":[37.6351078985414,55.8036940335…      NaN
260  {"coordinates":[37.664089701977,55.83402925532…      NaN
351  {"coordinates":[37.5064575807255,55.6987504361…      NaN

[10 rows x 25 columns]
```

5. , geojson “data/mo.geojson”

[ ]:
```
MOSCOW_ZONES_PATH = "data/mo.geojson"
```

[ ]:
```
def embed_map(m):
    from IPython.display import IFrame
    m.save('index.html')
    return IFrame('index.html', width='100%', height='750px')
```

[ ]:
```
m = folium.Map()
style_function = lambda x: {
    "color" : "orange",
    "weight": 1
}

folium.GeoJson(MOSCOW_ZONES_PATH, name="geojson",␣
 ↪style_function=style_function).add_to(m)
m.fit_bounds(m.get_bounds())
display(m)
```

```
<folium.folium.Map at 0x7f9d98dd8d00>
```

6. , geodata_center.
, - .

[ ]:
```
import re
import json

def extract_coordinates(geo_data):
    """Parse geoData value."""
    try:
        geo_data_ = re.sub("(\w+)=(\w+), (\w+)=", r'"\1": "\2", "\3": ',␣
 ↪geo_data)
        return json.loads(geo_data_)["coordinates"]
    except:
        pass
```

[ ]:
```
for indx, row in df.iterrows():
    try:
        coords = extract_coordinates(row.geodata_center)
        folium.Circle(
            radius=5,
            location=[coords[1],coords[0]],
            popup=row["ShortName"]+'; '+row['LegalAddress']+'; '+row['WebSite'],
            color="red",
            fill=True,
            fill_opacity=1.0
        ).add_to(m)
    except:
        pass
display(m)
```

```
<folium.folium.Map at 0x7f9d98dd8d00>
```