

First Assessment

April 15, 2024

```
[1]: import pandas as pd
```

```
[3]: df = pd.read_csv("https://github.com/HamoyeHQ/
↳HDSC-Introduction-to-Python-for-machine-learning/files/7768140/
↳FoodBalanceSheets_E_Africa_NOFLAG.csv", encoding="latin-1")
```

```
[4]: df.describe(include="all")
```

```
[4]:
```

	Area Code	Area	Item Code	Item \
count	60943.000000	60943	60943.000000	60943
unique	NaN	49	NaN	119
top	NaN	Kenya	NaN	Milk - Excluding Butter
freq	NaN	1560	NaN	1262
mean	134.265576	NaN	2687.176706	NaN
std	72.605709	NaN	146.055739	NaN
min	4.000000	NaN	2501.000000	NaN
25%	74.000000	NaN	2562.000000	NaN
50%	136.000000	NaN	2630.000000	NaN
75%	195.000000	NaN	2775.000000	NaN
max	276.000000	NaN	2961.000000	NaN

	Element Code	Element	Unit	Y2014 \
count	60943.000000	60943	60943	59354.000000
unique	NaN	18	5	NaN
top	NaN	Domestic supply quantity	1000 tonnes	NaN
freq	NaN	5295	40933	NaN
mean	3814.856456	NaN	NaN	134.196282
std	2212.007033	NaN	NaN	1567.663696
min	511.000000	NaN	NaN	-1796.000000
25%	684.000000	NaN	NaN	0.000000
50%	5142.000000	NaN	NaN	0.090000
75%	5511.000000	NaN	NaN	8.340000
max	5911.000000	NaN	NaN	176405.000000

	Y2015	Y2016	Y2017	Y2018
count	59395.000000	59408.000000	59437.000000	59507.000000
unique	NaN	NaN	NaN	NaN
top	NaN	NaN	NaN	NaN

freq	NaN	NaN	NaN	NaN
mean	135.235966	136.555222	140.917765	143.758381
std	1603.403984	1640.007194	1671.862359	1710.782658
min	-3161.000000	-3225.000000	-1582.000000	-3396.000000
25%	0.000000	0.000000	0.000000	0.000000
50%	0.080000	0.080000	0.100000	0.070000
75%	8.460000	8.430000	9.000000	9.000000
max	181137.000000	185960.000000	190873.000000	195875.000000

```
[5]: # question 1
df[["Y2017", "Area"]]
```

```
[5]:      Y2017      Area
0    41389.00  Algeria
1         0.00  Algeria
2    3341.00  Algeria
3     92.82  Algeria
4     80.19  Algeria
...
60938    19.00  Zimbabwe
60939     1.33  Zimbabwe
60940     1.00  Zimbabwe
60941     0.04  Zimbabwe
60942     0.02  Zimbabwe

[60943 rows x 2 columns]
```

```
[6]: df.groupby("Area")["Y2017"].sum()
```

```
[6]: Area
Algeria      325644.27
Angola       229159.57
Benin        124771.22
Botswana      22101.30
Burkina Faso  101855.07
Cabo Verde    14650.74
Cameroon     232030.43
Central African Republic  29937.00
Chad          71594.68
Comoros         59.84
Congo         41181.68
Côte d'Ivoire  224599.01
Djibouti      22729.91
Egypt         866379.92
Eswatini       54343.33
Ethiopia      448683.76
Ethiopia PDR      0.00
```

Gabon	27979.64
Gambia	23154.18
Ghana	337599.06
Guinea	98138.87
Guinea-Bissau	19102.77
Kenya	264660.66
Lesotho	21267.96
Liberia	29342.20
Madagascar	131197.73
Malawi	181098.71
Mali	149928.33
Mauritania	156665.46
Mauritius	51114.83
Morocco	388495.36
Mozambique	161407.98
Namibia	29874.89
Niger	126707.58
Nigeria	1483268.23
Rwanda	73663.69
Sao Tome and Principe	12662.63
Senegal	95681.15
Seychelles	442.34
Sierra Leone	55311.33
South Africa	517590.54
Sudan	239931.92
Sudan (former)	0.00
Togo	49841.88
Tunisia	124167.20
Uganda	213950.38
United Republic of Tanzania	322616.85
Zambia	103223.77
Zimbabwe	75919.34

Name: Y2017, dtype: float64

```
[8]: grouped_values= df.groupby("Area")["Y2017"].sum().reset_index()
```

```
[11]: grouped_values.sort_values("Y2017", ascending=True)
```

```
[11]:
```

	Area	Y2017
34	Nigeria	1483268.23
13	Egypt	866379.92
40	South Africa	517590.54
15	Ethiopia	448683.76
30	Morocco	388495.36
19	Ghana	337599.06
0	Algeria	325644.27
46	United Republic of Tanzania	322616.85

22	Kenya	264660.66
41	Sudan	239931.92
6	Cameroon	232030.43
1	Angola	229159.57
11	Côte d'Ivoire	224599.01
45	Uganda	213950.38
26	Malawi	181098.71
31	Mozambique	161407.98
28	Mauritania	156665.46
27	Mali	149928.33
25	Madagascar	131197.73
33	Niger	126707.58
2	Benin	124771.22
44	Tunisia	124167.20
47	Zambia	103223.77
4	Burkina Faso	101855.07
20	Guinea	98138.87
37	Senegal	95681.15
48	Zimbabwe	75919.34
35	Rwanda	73663.69
8	Chad	71594.68
39	Sierra Leone	55311.33
14	Eswatini	54343.33
29	Mauritius	51114.83
43	Togo	49841.88
10	Congo	41181.68
7	Central African Republic	29937.00
32	Namibia	29874.89
24	Liberia	29342.20
17	Gabon	27979.64
18	Gambia	23154.18
12	Djibouti	22729.91
3	Botswana	22101.30
23	Lesotho	21267.96
21	Guinea-Bissau	19102.77
5	Cabo Verde	14650.74
36	Sao Tome and Principe	12662.63
38	Seychelles	442.34
9	Comoros	59.84
42	Sudan (former)	0.00
16	Ethiopia PDR	0.00

```
[13]: # question 2
grouped_elements= df.groupby("Element")
```

```
[15]: grouped_elements.describe(include="all")
```

[15]:

Element	Area Code				mean
	count	unique	top	freq	
Domestic supply quantity	5295.0	NaN	NaN	NaN	133.898584
Export Quantity	4403.0	NaN	NaN	NaN	136.250284
Fat supply quantity (g/capita/day)	5023.0	NaN	NaN	NaN	134.391798
Feed	1319.0	NaN	NaN	NaN	133.640637
Food	4941.0	NaN	NaN	NaN	134.243068
Food supply (kcal/capita/day)	5014.0	NaN	NaN	NaN	134.434982
Food supply quantity (kg/capita/yr)	4905.0	NaN	NaN	NaN	134.239755
Import Quantity	5139.0	NaN	NaN	NaN	133.912045
Losses	2009.0	NaN	NaN	NaN	136.561971
Other uses (non-food)	1732.0	NaN	NaN	NaN	136.001155
Processing	2010.0	NaN	NaN	NaN	135.293532
Production	3881.0	NaN	NaN	NaN	135.725586
Protein supply quantity (g/capita/day)	5023.0	NaN	NaN	NaN	134.391798
Residuals	4655.0	NaN	NaN	NaN	133.892803
Seed	762.0	NaN	NaN	NaN	135.875328
Stock Variation	4232.0	NaN	NaN	NaN	135.058129
Total Population - Both sexes	45.0	NaN	NaN	NaN	133.777778
Tourist consumption	555.0	NaN	NaN	NaN	90.645045

Element	std	min	25%	50%	75%
Domestic supply quantity	72.429582	4.0	74.0	136.0	195.0
Export Quantity	72.474261	4.0	75.0	137.0	197.0
Fat supply quantity (g/capita/day)	72.408886	4.0	74.0	136.0	195.0
Feed	71.452934	4.0	74.0	133.0	197.0
Food	72.466059	4.0	74.0	136.0	195.0
Food supply (kcal/capita/day)	72.358700	4.0	74.0	136.0	195.0
Food supply quantity (kg/capita/yr)	72.455564	4.0	74.0	136.0	195.0
Import Quantity	72.365119	4.0	74.0	136.0	195.0
Losses	72.785591	4.0	74.0	137.0	197.0
Other uses (non-food)	73.106619	4.0	74.0	137.0	202.0
Processing	74.949822	4.0	59.0	143.0	195.0
Production	72.972281	4.0	74.0	137.0	197.0
Protein supply quantity (g/capita/day)	72.408886	4.0	74.0	136.0	195.0
Residuals	72.429026	4.0	74.0	136.0	195.0
Seed	73.266647	4.0	74.0	137.0	202.0
Stock Variation	72.874997	4.0	74.0	137.0	197.0
Total Population - Both sexes	73.112581	4.0	74.0	136.0	195.0
Tourist consumption	59.925421	7.0	35.0	59.0	137.0

...	Y2018	
...	unique top freq	mean
Element	...	

Domestic supply quantity	...	NaN	NaN	NaN	411.185712
Export Quantity	...	NaN	NaN	NaN	49.806583
Fat supply quantity (g/capita/day)	...	NaN	NaN	NaN	2.057086
Feed	...	NaN	NaN	NaN	190.603820
Food	...	NaN	NaN	NaN	265.385972
Food supply (kcal/capita/day)	...	NaN	NaN	NaN	91.620246
Food supply quantity (kg/capita/yr)	...	NaN	NaN	NaN	10.058817
Import Quantity	...	NaN	NaN	NaN	58.240059
Losses	...	NaN	NaN	NaN	81.624502
Other uses (non-food)	...	NaN	NaN	NaN	55.166749
Processing	...	NaN	NaN	NaN	153.983525
Production	...	NaN	NaN	NaN	539.540533
Protein supply quantity (g/capita/day)	...	NaN	NaN	NaN	2.373357
Residuals	...	NaN	NaN	NaN	7.552860
Seed	...	NaN	NaN	NaN	33.328681
Stock Variation	...	NaN	NaN	NaN	4.879751
Total Population - Both sexes	...	NaN	NaN	NaN	25346.777778
Tourist consumption	...	NaN	NaN	NaN	0.162162

		std	min	25%
Element				
Domestic supply quantity	2544.664930	-184.0	0.0000	
Export Quantity	842.633449	0.0	0.0000	
Fat supply quantity (g/capita/day)	7.799864	0.0	0.0000	
Feed	1382.718629	0.0	0.0000	
Food	1407.952979	0.0	0.0000	
Food supply (kcal/capita/day)	357.469150	0.0	0.0000	
Food supply quantity (kg/capita/yr)	29.555214	0.0	0.0100	
Import Quantity	467.670555	0.0	0.0000	
Losses	402.263215	0.0	0.0000	
Other uses (non-food)	497.694277	0.0	0.0000	
Processing	1022.278276	0.0	0.0000	
Production	3019.990719	0.0	0.1625	
Protein supply quantity (g/capita/day)	8.933189	0.0	0.0000	
Residuals	363.873325	-598.0	0.0000	
Seed	105.787366	0.0	1.0000	
Stock Variation	132.428693	-3396.0	0.0000	
Total Population - Both sexes	35337.267187	211.0	4403.0000	
Tourist consumption	0.617736	0.0	0.0000	

		50%	75%	max
Element				
Domestic supply quantity	7.00	82.0000	117038.00	
Export Quantity	0.00	3.0000	34641.00	
Fat supply quantity (g/capita/day)	0.03	0.4900	94.94	

Feed	0.20	24.0000	31567.00
Food	5.38	63.8800	54077.00
Food supply (kcal/capita/day)	3.00	26.0000	3450.00
Food supply quantity (kg/capita/yr)	0.60	5.6900	443.56
Import Quantity	0.48	8.0000	18036.00
Losses	4.00	31.0000	10355.00
Other uses (non-food)	0.00	2.0000	17221.00
Processing	0.00	5.0000	22609.00
Production	16.00	153.7500	118172.00
Protein supply quantity (g/capita/day)	0.03	0.6475	101.08
Residuals	0.00	0.0000	17323.00
Seed	4.00	22.7500	1474.00
Stock Variation	0.00	0.0000	2226.00
Total Population - Both sexes	15478.00	29767.0000	195875.00
Tourist consumption	0.00	0.0000	6.00

[18 rows x 121 columns]

```
[24]: grouped_elements=df.groupby("Element")["Y2017"].sum()
```

```
[25]: grouped_elements.loc["Stock Variation"]
```

```
[25]: 54316.91
```

```
[27]: grouped_elements=df.groupby("Element").agg({
      'Y2014' : 'sum',
      'Y2015' : 'sum',
      'Y2016' : 'sum',
      'Y2017' : 'sum',
    })
```

```
[28]: grouped_elements.loc["Stock Variation"]
```

```
[28]: Y2014    58749.83
      Y2015    34910.99
      Y2016    33140.12
      Y2017    54316.91
      Name: Stock Variation, dtype: float64
```

```
[150]: #question 3
      grouped_areas= df.groupby([df["Area"]=="Madagascar", df["Element"]=="Protein_
      ↪supply quantity (g/capita/day)"])[ "Y2015"].sum()
```

```
[167]: grouped_areas.iloc[0:]
```

```
[167]: Area  Element
      False  False    7894004.43
           True    11660.90
```

```

True    False    126501.85
      True      173.05
Name: Y2015, dtype: float64

```

```

[168]: #question 4
grouped_element_codes = df[["Element Code", "Y2014","Y2015","Y2016","Y2017"]]

```

```

[169]: grouped_element_codes.corr()

```

```

[169]:
      Element Code  Y2014  Y2015  Y2016  Y2017
Element Code      1.000000 0.024457 0.023889 0.023444 0.024254
Y2014              0.024457 1.000000 0.994647 0.996081 0.995230
Y2015              0.023889 0.994647 1.000000 0.995739 0.988048
Y2016              0.023444 0.996081 0.995739 1.000000 0.992785
Y2017              0.024254 0.995230 0.988048 0.992785 1.000000

```

```

[172]: # question 6
wine_group = df.groupby(df["Item"] == "Wine").agg({
    "Y2015" : "sum",
    "Y2018" : "sum"})

```

```

[173]: wine_group.loc[0:]

```

```

[173]:
      Y2015  Y2018
Item
False  8028088.42  8550590.64
True    4251.81   4039.32

```

```

[182]: # question 9
grouped_elements= df.groupby(df["Element"]=="Processing")["Y2017"].sum()
grouped_elements.loc[0:]

```

```

[182]: Element
False    8082893.19
True      292836.00
Name: Y2017, dtype: float64

```

```

[183]: #question 12
df[["Y2017", "Area"]]

```

```

[183]:
      Y2017  Area
0    41389.00  Algeria
1         0.00  Algeria
2    3341.00  Algeria
3      92.82  Algeria
4     80.19  Algeria
...
60938    19.00  Zimbabwe

```



```
60939      1.33  Zimbabwe
60940      1.00  Zimbabwe
60941      0.04  Zimbabwe
60942      0.02  Zimbabwe
```

```
[60943 rows x 2 columns]
```

```
[190]: question12 = df.groupby("Area")["Y2017"].sum().reset_index()
```

```
[192]: question12.sort_values("Y2017", ascending=False)
```

```
[192]:
```

	Area	Y2017
34	Nigeria	1483268.23
13	Egypt	866379.92
40	South Africa	517590.54
15	Ethiopia	448683.76
30	Morocco	388495.36
19	Ghana	337599.06
0	Algeria	325644.27
46	United Republic of Tanzania	322616.85
22	Kenya	264660.66
41	Sudan	239931.92
6	Cameroon	232030.43
1	Angola	229159.57
11	Côte d'Ivoire	224599.01
45	Uganda	213950.38
26	Malawi	181098.71
31	Mozambique	161407.98
28	Mauritania	156665.46
27	Mali	149928.33
25	Madagascar	131197.73
33	Niger	126707.58
2	Benin	124771.22
44	Tunisia	124167.20
47	Zambia	103223.77
4	Burkina Faso	101855.07
20	Guinea	98138.87
37	Senegal	95681.15
48	Zimbabwe	75919.34
35	Rwanda	73663.69
8	Chad	71594.68
39	Sierra Leone	55311.33
14	Eswatini	54343.33
29	Mauritius	51114.83
43	Togo	49841.88
10	Congo	41181.68
7	Central African Republic	29937.00

32	Namibia	29874.89
24	Liberia	29342.20
17	Gabon	27979.64
18	Gambia	23154.18
12	Djibouti	22729.91
3	Botswana	22101.30
23	Lesotho	21267.96
21	Guinea-Bissau	19102.77
5	Cabo Verde	14650.74
36	Sao Tome and Principe	12662.63
38	Seychelles	442.34
9	Comoros	59.84
42	Sudan (former)	0.00
16	Ethiopia PDR	0.00

```
[201]: # question 18
nan_in_2014 = df["Y2014"].isnull().sum()
print(f"Number of NaN 2014 = {nan_in_2014}")
all_in_2014 = df["Y2014"].count()
percentage_of_nan = (nan_in_2014 / all_in_2014) * 100
print(f"Percentage of Nan in 2014 = {percentage_of_nan}")
```

Number of NaN 2014 = 1589

Percentage of Nan in 2014 = 2.6771573946153584

```
[ ]:
```