Untitled1

May 20, 2022

[89]:
```
import pandas as pd
import numpy as np
from sklearn.preprocessing import OneHotEncoder
```

[90]:
```
data = pd.read_csv('cogs118aDATA.csv')
```

[91]:
```
data.shape
```

[91]: (8254, 36)

[92]:
```
data.head()
```

[92]:

	Entity	Code	Year	Number of executions (Amnesty International)	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Neoplasms - Sex: Both - Age: All Ages (Number)	Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	2007	15	2933.0	15925.0	481.0
1	Afghanistan	AFG	2008	17	2731.0	16148.0	462.0
2	Afghanistan	AFG	2009	0	2460.0	16383.0	448.0
3	Afghanistan	AFG	2011	2	2327.0	17094.0	
4	Afghanistan	AFG	2012	14	2254.0	17522.0	

```
3                                                 448.0
4                                                 445.0

   Deaths - Malaria - Sex: Both - Age: All Ages (Number)  \
0                                                 393.0
1                                                 255.0
2                                                 239.0
3                                                 390.0
4                                                  94.0

   Deaths - Drowning - Sex: Both - Age: All Ages (Number)  \
0                                                2127.0
1                                                1973.0
2                                                1852.0
3                                                1775.0
4                                                1716.0

   Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)  …  \
0                                                3657.0                     …
1                                                3785.0                     …
2                                                3874.0                     …
3                                                4170.0                     …
4                                                4245.0                     …

   Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)  \
0                                                2439.0
1                                                2231.0
2                                                1998.0
3                                                1805.0
4                                                1667.0

   Terrorism (deaths)  \
0              1199.0
1              1092.0
2              1065.0
3              1525.0
4              3521.0

   Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)  \
0                                               53962.0
1                                               54051.0
2                                               53964.0
3                                               54347.0
4                                               54868.0

   Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)  \
0                                                4490.0
```

```
1                                                4534.0
2                                                4597.0
3                                                4785.0
4                                                4846.0

   Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number)  \
0                                                7222.0
1                                                7143.0
2                                                7045.0
3                                                6916.0
4                                                6878.0

   Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All
Ages (Number)  \
0                                                3346.0
1                                                3316.0
2                                                3291.0
3                                                3318.0
4                                                3353.0

   Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number)  \
0                                                6458.0
1                                                6408.0
2                                                6358.0
3                                                6370.0
4                                                6398.0

   Deaths - Acute hepatitis - Sex: Both - Age: All Ages (Number)  \
0                                                3437.0
1                                                3005.0
2                                                2663.0
3                                                2365.0
4                                                2264.0

   Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All Ages
(Number)  \
0                                                1402.0
1                                                1424.0
2                                                1449.0
3                                                1508.0
4                                                1544.0

   Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number)
0                                                 450.0
1                                                 455.0
2                                                 460.0
3                                                 473.0
```

```
4                                                  482.0

[5 rows x 36 columns]
```

[93]:
```
data.describe()
```

[93]:
```
              Year  Deaths - Meningitis - Sex: Both - Age: All Ages (Number)  \
count  8254.000000                                        8010.000000
mean   2004.448025                                       12909.701124
std       8.642230                                       41799.388071
min    1990.000000                                           0.000000
25%    1997.000000                                          29.000000
50%    2004.000000                                         294.000000
75%    2012.000000                                        3187.750000
max    2019.000000                                      432524.000000

       Deaths - Neoplasms - Sex: Both - Age: All Ages (Number)  \
count                                       8.010000e+03
mean                                        2.983985e+05
std                                         8.643901e+05
min                                         1.000000e+00
25%                                         1.934250e+03
50%                                         1.033850e+04
75%                                         9.186925e+04
max                                         1.007964e+07

       Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages
(Number)  \
count                                        8010.000000
mean                                         4444.838077
std                                         12111.913749
min                                             0.000000
25%                                            35.000000
50%                                           244.000000
75%                                          1470.750000
max                                        129705.000000

       Deaths - Malaria - Sex: Both - Age: All Ages (Number)  \
count                                        8010.000000
mean                                        31812.044569
std                                        123035.872293
min                                             0.000000
25%                                             0.000000
50%                                             1.000000
75%                                          2462.000000
max                                        961129.000000
```

	Deaths - Drowning - Sex: Both - Age: All Ages (Number)	\
count	8010.000000	
mean	12532.637953	
std	40095.990735	
min	0.000000	
25%	58.000000	
50%	393.500000	
75%	3017.750000	
max	460665.000000	

	Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)	\
count	8010.000000	
mean	15315.848315	
std	42888.544878	
min	0.000000	
25%	76.250000	
50%	494.000000	
75%	4372.500000	
max	463129.000000	

	Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number)	\
count	8.010000e+03	
mean	4.725143e+04	
std	1.744798e+05	
min	0.000000e+00	
25%	2.600000e+01	
50%	4.200000e+02	
75%	9.484500e+03	
max	1.844490e+06	

	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	\
count	8010.000000	
mean	3469.958926	
std	11186.514866	
min	0.000000	
25%	7.000000	
50%	57.000000	
75%	518.750000	
max	128083.000000	

	Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number)	…	\
count	8.010000e+03	…	
mean	5.605527e+04	…	
std	1.837876e+05	…	
min	0.000000e+00	…	
25%	6.200000e+01	…	
50%	9.560000e+02	…	

```
75%                                          1.037775e+04           …
max                                          1.808478e+06           …

       Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)
\
count                                         8010.000000
mean                                         14441.384519
std                                          47987.721059
min                                              0.000000
25%                                             10.000000
50%                                            233.500000
75%                                           4245.000000
max                                         656314.000000

       Terrorism (deaths)  \
count         2891.000000
mean           349.235905
std           1917.143788
min              0.000000
25%              0.000000
50%              5.000000
75%             60.000000
max          44490.000000

       Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)  \
count                                        8.010000e+03
mean                                         5.672777e+05
std                                          1.606918e+06
min                                          4.000000e+00
25%                                          4.348500e+03
50%                                          2.326550e+04
75%                                          1.663318e+05
max                                          1.856251e+07

       Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)  \
count                                        8.010000e+03
mean                                         3.614545e+04
std                                          1.028788e+05
min                                          0.000000e+00
25%                                          2.810000e+02
50%                                          1.651000e+03
75%                                          1.192175e+04
max                                          1.427232e+06

       Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages
(Number)  \
count                                        8.010000e+03
```

```
mean                                                1.315012e+05
std                                                 4.174924e+05
min                                                 1.000000e+00
25%                                                 5.262500e+02
50%                                                 2.960500e+03
75%                                                 2.815650e+04
max                                                 3.974315e+06

       Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age:
All Ages (Number)  \
count                                               8.010000e+03
mean                                                4.668634e+04
std                                                 1.282383e+05
min                                                 0.000000e+00
25%                                                 3.040000e+02
50%                                                 2.134000e+03
75%                                                 1.680225e+04
max                                                 1.472012e+06

       Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number)  \
count                                               8.010000e+03
mean                                                8.261491e+04
std                                                 2.253554e+05
min                                                 0.000000e+00
25%                                                 5.990000e+02
50%                                                 4.032500e+03
75%                                                 2.838875e+04
max                                                 2.557689e+06

       Deaths - Acute hepatitis - Sex: Both - Age: All Ages (Number)  \
count                                           8010.000000
mean                                            4586.226592
std                                            16692.425941
min                                                0.000000
25%                                                3.000000
50%                                               47.000000
75%                                              453.750000
max                                           166405.000000

       Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All
Ages (Number)  \
count                                               8.010000e+03
mean                                                3.923395e+04
std                                                 1.179772e+05
min                                                 0.000000e+00
25%                                                 2.010000e+02
50%                                                 1.337000e+03
```

```
75%                                                1.186775e+04
max                                                1.623276e+06

       Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number)
count                                                8010.000000
mean                                                 9367.016979
std                                                 27358.717966
min                                                     0.000000
25%                                                    55.000000
50%                                                   331.000000
75%                                                  2954.000000
max                                                362907.000000

[8 rows x 33 columns]
```

[94]:
```
data_nan_ratio = {k:data[k].isna().sum()/data.shape[0] for k in list(data.
↪columns)}
```

[95]:
```
data_nan_ratio
```

[95]:
```
{'Entity': 0.0,
 'Code': 0.24812212260722075,
 'Year': 0.0,
 'Number of executions (Amnesty International)': 0.9676520474921251,
 'Deaths - Meningitis - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Neoplasms - Sex: Both - Age: All Ages (Number)': 0.02956142476375091,
 'Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Malaria - Sex: Both - Age: All Ages (Number)': 0.02956142476375091,
 'Deaths - Drowning - Sex: Both - Age: All Ages (Number)': 0.02956142476375091,
 'Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number)': 0.02956142476375091,
 'Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Road injuries - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Maternal disorders - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Neonatal disorders - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)':
```

```
0.02956142476375091,
 'Deaths - Exposure to forces of nature - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Diarrheal diseases - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Environmental heat and cold exposure - Sex: Both - Age: All Ages
(Number)': 0.02956142476375091,
 'Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Self-harm - Sex: Both - Age: All Ages (Number)': 0.02956142476375091,
 'Deaths - Conflict and terrorism - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Diabetes mellitus - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Poisonings - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Terrorism (deaths)': 0.6497455779016235,
 'Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All
Ages (Number)': 0.02956142476375091,
 'Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 'Deaths - Acute hepatitis - Sex: Both - Age: All Ages (Number)':
0.02956142476375091,
 "Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All Ages
(Number)": 0.02956142476375091,
 "Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number)":
0.02956142476375091}
```

[96]:
```
words = []
for i in list(data):
    if i.find('-') == -1:
        words.append(i)
    else:
        label = i.split(' - ')[1].strip()
        words.append(label)
```

[97]:
```
words
```

[97]:
```
['Entity',
 'Code',
 'Year',
 'Number of executions (Amnesty International)',
 'Meningitis',
 'Neoplasms',
 'Fire, heat, and hot substances',
 'Malaria',
 'Drowning',
 'Interpersonal violence',
 'HIV/AIDS',
 'Drug use disorders',
 'Tuberculosis',
 'Road injuries',
 'Maternal disorders',
 'Lower respiratory infections',
 'Neonatal disorders',
 'Alcohol use disorders',
 'Exposure to forces of nature',
 'Diarrheal diseases',
 'Environmental heat and cold exposure',
 'Nutritional deficiencies',
 'Self-harm',
 'Conflict and terrorism',
 'Diabetes mellitus',
 'Poisonings',
 'Protein-energy malnutrition',
 'Terrorism (deaths)',
 'Cardiovascular diseases',
 'Chronic kidney disease',
 'Chronic respiratory diseases',
 'Cirrhosis and other chronic liver diseases',
 'Digestive diseases',
 'Acute hepatitis',
 "Alzheimer's disease and other dementias",
 "Parkinson's disease"]
```

[98]:
```
dictio = {k:v for k, v in zip(data.columns, words)}

data = data.rename(columns = dictio)
```

[99]:
```
data.columns
data.head()
```

[99]:

	Entity	Code	Year	Number of executions (Amnesty International)
0	Afghanistan	AFG	2007	15
1	Afghanistan	AFG	2008	17

```
2  Afghanistan  AFG  2009                                              0
3  Afghanistan  AFG  2011                                              2
4  Afghanistan  AFG  2012                                             14

   Meningitis  Neoplasms  Fire, heat, and hot substances  Malaria  Drowning  \
0      2933.0    15925.0                           481.0    393.0    2127.0
1      2731.0    16148.0                           462.0    255.0    1973.0
2      2460.0    16383.0                           448.0    239.0    1852.0
3      2327.0    17094.0                           448.0    390.0    1775.0
4      2254.0    17522.0                           445.0     94.0    1716.0

   Interpersonal violence  …  Protein-energy malnutrition  \
0                  3657.0  …                       2439.0
1                  3785.0  …                       2231.0
2                  3874.0  …                       1998.0
3                  4170.0  …                       1805.0
4                  4245.0  …                       1667.0

   Terrorism (deaths)  Cardiovascular diseases  Chronic kidney disease  \
0              1199.0                  53962.0                  4490.0
1              1092.0                  54051.0                  4534.0
2              1065.0                  53964.0                  4597.0
3              1525.0                  54347.0                  4785.0
4              3521.0                  54868.0                  4846.0

   Chronic respiratory diseases  Cirrhosis and other chronic liver diseases  \
0                        7222.0                                       3346.0
1                        7143.0                                       3316.0
2                        7045.0                                       3291.0
3                        6916.0                                       3318.0
4                        6878.0                                       3353.0

   Digestive diseases  Acute hepatitis  \
0              6458.0           3437.0
1              6408.0           3005.0
2              6358.0           2663.0
3              6370.0           2365.0
4              6398.0           2264.0

   Alzheimer's disease and other dementias  Parkinson's disease
0                                   1402.0                450.0
1                                   1424.0                455.0
2                                   1449.0                460.0
3                                   1508.0                473.0
4                                   1544.0                482.0

[5 rows x 36 columns]
```

[100]:
```
#Noticed that Terrorism and Executions had extreme amounts of missing na values
data = data.drop(columns = ['Number of executions (Amnesty International)',␣
 ↪'Terrorism (deaths)', 'Code'])
```

[101]:
```
data.columns
```

[101]:
```
Index(['Entity', 'Year', 'Meningitis', 'Neoplasms',
       'Fire, heat, and hot substances', 'Malaria', 'Drowning',
       'Interpersonal violence', 'HIV/AIDS', 'Drug use disorders',
       'Tuberculosis', 'Road injuries', 'Maternal disorders',
       'Lower respiratory infections', 'Neonatal disorders',
       'Alcohol use disorders', 'Exposure to forces of nature',
       'Diarrheal diseases', 'Environmental heat and cold exposure',
       'Nutritional deficiencies', 'Self-harm', 'Conflict and terrorism',
       'Diabetes mellitus', 'Poisonings', 'Protein-energy malnutrition',
       'Cardiovascular diseases', 'Chronic kidney disease',
       'Chronic respiratory diseases',
       'Cirrhosis and other chronic liver diseases', 'Digestive diseases',
       'Acute hepatitis', 'Alzheimer's disease and other dementias',
       'Parkinson's disease'],
      dtype='object')
```

[102]:
```
data.isna().sum()
```

[102]:
```
Entity                                      0
Year                                        0
Meningitis                                244
Neoplasms                                 244
Fire, heat, and hot substances            244
Malaria                                   244
Drowning                                  244
Interpersonal violence                    244
HIV/AIDS                                  244
Drug use disorders                        244
Tuberculosis                              244
Road injuries                             244
Maternal disorders                        244
Lower respiratory infections              244
Neonatal disorders                        244
Alcohol use disorders                     244
Exposure to forces of nature              244
Diarrheal diseases                        244
Environmental heat and cold exposure      244
Nutritional deficiencies                  244
Self-harm                                 244
Conflict and terrorism                    244
Diabetes mellitus                         244
```

```
Poisonings                                      244
Protein-energy malnutrition                     244
Cardiovascular diseases                         244
Chronic kidney disease                          244
Chronic respiratory diseases                    244
Cirrhosis and other chronic liver diseases      244
Digestive diseases                              244
Acute hepatitis                                 244
Alzheimer's disease and other dementias         244
Parkinson's disease                             244
dtype: int64
```

[103]:
```
df = data.dropna()
```

[104]:
```
df.shape
```

[104]:
```
(8010, 33)
```

[105]:
```
df.describe()
```

[105]:
```
              Year     Meningitis     Neoplasms  \
count  8010.000000    8010.000000  8.010000e+03
mean   2004.500000   12909.701124  2.983985e+05
std       8.655982   41799.388071  8.643901e+05
min    1990.000000       0.000000  1.000000e+00
25%    1997.000000      29.000000  1.934250e+03
50%    2004.500000     294.000000  1.033850e+04
75%    2012.000000    3187.750000  9.186925e+04
max    2019.000000  432524.000000  1.007964e+07

       Fire, heat, and hot substances        Malaria       Drowning  \
count                     8010.000000    8010.000000    8010.000000
mean                      4444.838077   31812.044569   12532.637953
std                      12111.913749  123035.872293   40095.990735
min                          0.000000       0.000000       0.000000
25%                         35.000000       0.000000      58.000000
50%                        244.000000       1.000000     393.500000
75%                       1470.750000    2462.000000    3017.750000
max                     129705.000000  961129.000000  460665.000000

       Interpersonal violence      HIV/AIDS  Drug use disorders  Tuberculosis  \
count             8010.000000  8.010000e+03         8010.000000  8.010000e+03
mean             15315.848315  4.725143e+04         3469.958926  5.605527e+04
std              42888.544878  1.744798e+05        11186.514866  1.837876e+05
min                  0.000000  0.000000e+00            0.000000  0.000000e+00
25%                 76.250000  2.600000e+01            7.000000  6.200000e+01
50%                494.000000  4.200000e+02           57.000000  9.560000e+02
```

```
75%              4372.500000  9.484500e+03          518.750000  1.037775e+04
max            463129.000000  1.844490e+06       128083.000000  1.808478e+06

       …    Poisonings  Protein-energy malnutrition  \
count  …   8010.000000                   8010.000000
mean   …   3189.111111                  14441.384519
std    …   9180.094933                  47987.721059
min    …      0.000000                      0.000000
25%    …     13.000000                     10.000000
50%    …    125.000000                    233.500000
75%    …    797.750000                   4245.000000
max    …  92101.000000                 656314.000000

       Cardiovascular diseases  Chronic kidney disease  \
count             8.010000e+03            8.010000e+03
mean              5.672777e+05            3.614545e+04
std               1.606918e+06            1.028788e+05
min               4.000000e+00            0.000000e+00
25%               4.348500e+03            2.810000e+02
50%               2.326550e+04            1.651000e+03
75%               1.663318e+05            1.192175e+04
max               1.856251e+07            1.427232e+06

       Chronic respiratory diseases  \
count                  8.010000e+03
mean                   1.315012e+05
std                    4.174924e+05
min                    1.000000e+00
25%                    5.262500e+02
50%                    2.960500e+03
75%                    2.815650e+04
max                    3.974315e+06

       Cirrhosis and other chronic liver diseases  Digestive diseases  \
count                                8.010000e+03        8.010000e+03
mean                                 4.668634e+04        8.261491e+04
std                                  1.282383e+05        2.253554e+05
min                                  0.000000e+00        0.000000e+00
25%                                  3.040000e+02        5.990000e+02
50%                                  2.134000e+03        4.032500e+03
75%                                  1.680225e+04        2.838875e+04
max                                  1.472012e+06        2.557689e+06

       Acute hepatitis  Alzheimer's disease and other dementias  \
count      8010.000000                              8.010000e+03
mean       4586.226592                              3.923395e+04
std       16692.425941                              1.179772e+05
```

```
min           0.000000                              0.000000e+00
25%           3.000000                              2.010000e+02
50%          47.000000                              1.337000e+03
75%         453.750000                              1.186775e+04
max      166405.000000                              1.623276e+06

       Parkinson's disease
count          8010.000000
mean           9367.016979
std           27358.717966
min               0.000000
25%              55.000000
50%             331.000000
75%            2954.000000
max          362907.000000

[8 rows x 32 columns]
```

[109]:
```
to_encode = set(df['Entity'].values)
```

[110]:
```
#Preparation for very rough kmeans

enc = OneHotEncoder(handle_unknown = 'ignore')
to_encode = list(enumerate(to_encode))
to_encode
```

[110]:
```
[(0, 'Cook Islands'),
 (1, 'Central Asia'),
 (2, 'Grenada'),
 (3, 'Switzerland'),
 (4, 'Turkmenistan'),
 (5, 'Laos'),
 (6, 'Africa'),
 (7, 'New Zealand'),
 (8, 'Malawi'),
 (9, 'Oman'),
 (10, 'Denmark'),
 (11, 'Spain'),
 (12, 'Qatar'),
 (13, 'North Korea'),
 (14, 'European Region'),
 (15, 'Middle East & North Africa'),
 (16, 'Italy'),
 (17, 'Greenland'),
 (18, 'Tanzania'),
 (19, 'Singapore'),
 (20, 'Zimbabwe'),
```

```
(21, 'Bolivia'),
(22, 'Mozambique'),
(23, 'Western Pacific Region'),
(24, 'Palau'),
(25, 'Panama'),
(26, 'Commonwealth Low Income'),
(27, 'Ethiopia'),
(28, 'Malaysia'),
(29, 'Oceania'),
(30, 'Moldova'),
(31, 'Haiti'),
(32, 'Georgia'),
(33, 'Guinea-Bissau'),
(34, 'United Arab Emirates'),
(35, 'United States Virgin Islands'),
(36, 'Brunei'),
(37, 'American Samoa'),
(38, 'Ireland'),
(39, 'Mali'),
(40, 'North Macedonia'),
(41, 'Scotland'),
(42, 'Western sub-Saharan Africa'),
(43, 'Australasia'),
(44, 'Niue'),
(45, 'Belarus'),
(46, 'Kazakhstan'),
(47, 'North Africa and Middle East'),
(48, 'Luxembourg'),
(49, 'Burkina Faso'),
(50, 'Seychelles'),
(51, 'Eastern sub-Saharan Africa'),
(52, 'Fiji'),
(53, 'South Africa'),
(54, 'Iraq'),
(55, 'Somalia'),
(56, 'Cape Verde'),
(57, 'Southeast Asia'),
(58, 'Kyrgyzstan'),
(59, 'Liberia'),
(60, 'World Bank Low Income'),
(61, 'Comoros'),
(62, 'Mexico'),
(63, 'Papua New Guinea'),
(64, 'Japan'),
(65, 'Venezuela'),
(66, 'Western Europe'),
(67, 'Gabon'),
```

```
(68, 'Saint Kitts and Nevis'),
(69, 'Gambia'),
(70, 'Europe'),
(71, 'High-income North America'),
(72, 'Jamaica'),
(73, 'United Kingdom'),
(74, 'Palestine'),
(75, 'Nordic Region'),
(76, 'United States'),
(77, 'Romania'),
(78, 'Malta'),
(79, 'Philippines'),
(80, 'Yemen'),
(81, 'Sweden'),
(82, 'World Bank Lower Middle Income'),
(83, 'Guatemala'),
(84, 'African Region'),
(85, 'Belgium'),
(86, 'Andean Latin America'),
(87, 'Jordan'),
(88, 'Costa Rica'),
(89, 'Afghanistan'),
(90, 'Slovakia'),
(91, 'South-East Asia Region'),
(92, 'Guyana'),
(93, 'Vietnam'),
(94, 'Morocco'),
(95, 'Ghana'),
(96, 'Burundi'),
(97, 'Hungary'),
(98, 'Saint Vincent and the Grenadines'),
(99, 'Vanuatu'),
(100, 'Mongolia'),
(101, 'Benin'),
(102, 'Cameroon'),
(103, 'Barbados'),
(104, 'Latvia'),
(105, 'Southeast Asia, East Asia, and Oceania'),
(106, 'Cuba'),
(107, 'Equatorial Guinea'),
(108, 'Turkey'),
(109, 'Chad'),
(110, 'Lebanon'),
(111, 'Central sub-Saharan Africa'),
(112, 'High-income'),
(113, 'Lithuania'),
(114, 'Montenegro'),
```

```
(115, 'Tropical Latin America'),
(116, 'Germany'),
(117, 'Algeria'),
(118, 'Peru'),
(119, 'Cyprus'),
(120, 'Eritrea'),
(121, 'Libya'),
(122, 'Nicaragua'),
(123, 'Low-middle SDI'),
(124, 'Uruguay'),
(125, 'Togo'),
(126, 'Cambodia'),
(127, 'Mauritius'),
(128, 'Tokelau'),
(129, 'Saudi Arabia'),
(130, 'Monaco'),
(131, 'Austria'),
(132, 'African Union'),
(133, 'Bahamas'),
(134, 'Central Europe'),
(135, 'Rwanda'),
(136, 'Norway'),
(137, 'Greece'),
(138, 'Suriname'),
(139, 'South Korea'),
(140, 'China'),
(141, 'Bulgaria'),
(142, 'Puerto Rico'),
(143, 'Northern Ireland'),
(144, 'Estonia'),
(145, 'Israel'),
(146, 'Namibia'),
(147, 'Canada'),
(148, 'Europe & Central Asia - World Bank region'),
(149, 'Central Latin America'),
(150, 'Sierra Leone'),
(151, 'Syria'),
(152, 'Bermuda'),
(153, 'Uzbekistan'),
(154, 'Solomon Islands'),
(155, "Cote d'Ivoire"),
(156, 'Andorra'),
(157, 'Honduras'),
(158, 'Madagascar'),
(159, 'Middle SDI'),
(160, 'High SDI'),
(161, 'North America'),
```

```
(162, 'Wales'),
(163, 'Commonwealth'),
(164, 'Myanmar'),
(165, 'Eswatini'),
(166, 'Mauritania'),
(167, 'Niger'),
(168, 'Poland'),
(169, 'India'),
(170, 'Micronesia (country)'),
(171, 'Latin America & Caribbean - World Bank region'),
(172, 'Nauru'),
(173, 'Tonga'),
(174, 'Dominica'),
(175, 'Guam'),
(176, 'Commonwealth Middle Income'),
(177, 'Congo'),
(178, 'France'),
(179, 'Nigeria'),
(180, 'Trinidad and Tobago'),
(181, 'Bahrain'),
(182, 'Belize'),
(183, 'Tuvalu'),
(184, 'Chile'),
(185, 'Antigua and Barbuda'),
(186, 'World'),
(187, 'Democratic Republic of Congo'),
(188, 'Taiwan'),
(189, 'Croatia'),
(190, 'Region of the Americas'),
(191, 'Djibouti'),
(192, 'Guinea'),
(193, 'Pakistan'),
(194, 'Bosnia and Herzegovina'),
(195, 'Samoa'),
(196, 'Serbia'),
(197, 'European Union'),
(198, 'Argentina'),
(199, 'Sri Lanka'),
(200, 'Kiribati'),
(201, 'Lesotho'),
(202, 'Finland'),
(203, 'G20'),
(204, 'Kenya'),
(205, 'America'),
(206, 'Eastern Mediterranean Region'),
(207, 'Marshall Islands'),
(208, 'Tajikistan'),
```

```
(209, 'Tunisia'),
(210, 'Brazil'),
(211, 'Low SDI'),
(212, 'Zambia'),
(213, 'Southern sub-Saharan Africa'),
(214, 'Asia'),
(215, 'Azerbaijan'),
(216, 'Timor'),
(217, 'East Asia & Pacific - World Bank region'),
(218, 'Saint Lucia'),
(219, 'South Sudan'),
(220, 'Bangladesh'),
(221, 'Colombia'),
(222, 'Southern Latin America'),
(223, 'Armenia'),
(224, 'Angola'),
(225, 'Eastern Europe'),
(226, 'South Asia - World Bank region'),
(227, 'England'),
(228, 'Portugal'),
(229, 'El Salvador'),
(230, 'Dominican Republic'),
(231, 'Bhutan'),
(232, 'Albania'),
(233, 'Caribbean'),
(234, 'Commonwealth High Income'),
(235, 'Sudan'),
(236, 'Thailand'),
(237, 'Indonesia'),
(238, 'Slovenia'),
(239, 'Russia'),
(240, 'Botswana'),
(241, 'Maldives'),
(242, 'Czechia'),
(243, 'Central Europe, Eastern Europe, and Central Asia'),
(244, 'Australia'),
(245, 'Paraguay'),
(246, 'Iceland'),
(247, 'Senegal'),
(248, 'High-income Asia Pacific'),
(249, 'Egypt'),
(250, 'Northern Mariana Islands'),
(251, 'San Marino'),
(252, 'World Bank Upper Middle Income'),
(253, 'OECD Countries'),
(254, 'Uganda'),
(255, 'High-middle SDI'),
```

```
 (256, 'Ukraine'),
 (257, 'Netherlands'),
 (258, 'Nepal'),
 (259, 'Kuwait'),
 (260, 'Central African Republic'),
 (261, 'East Asia'),
 (262, 'World Bank High Income'),
 (263, 'Ecuador'),
 (264, 'Sub-Saharan Africa - World Bank region'),
 (265, 'Iran'),
 (266, 'Sao Tome and Principe')]
```

[149]:
```
enc.fit(to_encode)
```

[149]:
```
OneHotEncoder(handle_unknown='ignore')
```

[112]:
```
arr = df.to_numpy()
```

[167]:
```
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler, LabelEncoder
```

[168]:
```
le = LabelEncoder()
```

[173]:
```
Y = le.fit_transform(arr[:,0])
```

[177]:
```
X = arr[:,1:]
Y
```

[177]:
```
array([  0,   0,   0, …, 266, 266, 266])
```

[178]:
```
from sklearn.svm import SVC
X
```

[178]:
```
array([[2007, 2933.0, 15925.0, …, 3437.0, 1402.0, 450.0],
       [2008, 2731.0, 16148.0, …, 3005.0, 1424.0, 455.0],
       [2009, 2460.0, 16383.0, …, 2663.0, 1449.0, 460.0],
       …,
       [2017, 1460.0, 11744.0, …, 144.0, 781.0, 223.0],
       [2018, 1450.0, 12038.0, …, 139.0, 795.0, 227.0],
       [2019, 1450.0, 12353.0, …, 136.0, 812.0, 232.0]], dtype=object)
```

[179]:
```
clf = make_pipeline(StandardScaler(), SVC(gamma = 'auto'))
clf.fit(X, Y)
```

[179]:
```
Pipeline(steps=[('standardscaler', StandardScaler()),
                ('svc', SVC(gamma='auto'))])
```

[182]:
```
le.inverse_transform(clf.predict([X[0]]))
```

[182]:
```
array(['Sudan'], dtype=object)
```

[ ]:
```

```