SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

| | |
|---|---|
| Laboratorium Nr 1<br>Data 28.09.2024<br>Temat: "Wprowadzenie do narzędzi i środowiska pracy w analizie danych"<br>Wariant 7 | Tomasz Pietrzyk<br>Informatyka<br>II stopień, niestacjonarne,<br>1semestr, gr.1a |

1. Polecenie: wariant 7 zadania

Zadanie dotyczy pobrania danych z pliku, tworzenia ramki danych, wykonania poszczególnych zadań poniżej na podstawie odpowiedniego zbioru danych:

7. Global Burden of Disease Study 2019 (GBD 2019) Smoking Tobacco Use Prevalence 1990-2019 http://ghdx.healthdata.org/record/ihme-data/gbd-2019-smoking-tobacco-use-prevalence-1990-2019

2. Opis programu opracowanego (kody źródłowe, rzuty ekranu)

In [1]:
```
#Ladowanie biblioteki Pandas
import pandas as pd
```

In [2]:
```
#tworzenie ramki danych ze slownika

data = {'col_1': [3, 2, 1, 0], 'col_2': ['a', 'b', 'c', 'd']}

pd.DataFrame.from_dict(data)
```

Out[2]:

| | col_1 | col_2 |
|---|---|---|
| 0 | 3 | a |
| 1 | 2 | b |
| 2 | 1 | c |
| 3 | 0 | d |

In [3]:
```
#zachowanie ramki danych pobranych z pliku w formacie csv (xlsx)

df = pd.read_csv('IHME_GBD_2019_SMOKING_TOB_1990_2019_NUM_SMOKERS_Y2021M05D27.CSV')
print(df)
```

```
            measure_name  location_id location_name  sex_id sex_name  \
0      Number of Smokers            1        Global       1     Male
1      Number of Smokers            1        Global       2   Female
2      Number of Smokers            1        Global       3     Both
3      Number of Smokers            1        Global       1     Male
4      Number of Smokers            1        Global       2   Female
...                  ...          ...           ...     ...      ...
20965  Number of Smokers          522         Sudan       2   Female
20966  Number of Smokers          522         Sudan       3     Both
20967  Number of Smokers          522         Sudan       1     Male
20968  Number of Smokers          522         Sudan       2   Female
20969  Number of Smokers          522         Sudan       3     Both

       age_group_id age_group_name  year_id           val         upper  \
0                29      15+ years     1990  8.031015e+08  8.096221e+08
1                29      15+ years     1990  1.891488e+08  1.930929e+08
2                29      15+ years     1990  9.922503e+08  1.000161e+09
3                29      15+ years     1991  8.138972e+08  8.200339e+08
4                29      15+ years     1991  1.905375e+08  1.944249e+08
...             ...            ...      ...           ...           ...
20965            29      15+ years     2018  2.435999e+05  3.286166e+05
20966            29      15+ years     2018  2.610672e+06  2.833943e+06
20967            29      15+ years     2019  2.439150e+06  2.656579e+06
20968            29      15+ years     2019  2.500800e+05  3.345384e+05
20969            29      15+ years     2019  2.689230e+06  2.918332e+06

              lower
0      7.959086e+08
1      1.855595e+08
2      9.847880e+08
3      8.069514e+08
4      1.869744e+08
...             ...
20965  1.752508e+05
20966  2.409108e+06
20967  2.236450e+06
20968  1.816686e+05
20969  2.480656e+06

[20970 rows x 11 columns]
```

In [4]:
```
#tworzenie ramki danych z listy list

lists_income = [["Adam", "Kuba", "Robert"],
[4500,   5500,  6500]]

pd.DataFrame(lists_income)
```

Out[4]:

| | 0 | 1 | 2 |
|---|---|---|---|
| 0 | Adam | Kuba | Robert |
| 1 | 4500 | 5500 | 6500 |

In [5]:
```
#transponowanie (wymieniamy kolumny a wierszy)

df1 = pd.DataFrame.transpose(pd.DataFrame(lists_income))
print(df1)
```

```
        0     1
0    Adam  4500
1    Kuba  5500
2  Robert  6500
```

In [6]:
```
#wy´swietli´c pierwsze 10 wierszy ramki danych

df.head(10)
```

Out[6]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 |
| 5 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1991 | 1.004435e+09 | 1.011925e+09 | 9.969811e+08 |
| 6 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1992 | 8.233148e+08 | 8.292228e+08 | 8.167264e+08 |
| 7 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1992 | 1.919026e+08 | 1.957109e+08 | 1.884066e+08 |
| 8 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1992 | 1.015217e+09 | 1.022720e+09 | 1.007847e+09 |
| 9 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1993 | 8.313873e+08 | 8.372931e+08 | 8.249496e+08 |

In [7]:
```
#wy´swietli´c ostatnie 10 wierszy ramki danych

df.tail(10)
```

Out[7]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20960 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2016 | 2.454893e+06 | 2.665441e+06 | 2.267696e+06 |
| 20961 | Number of Smokers | 522 | Sudan | 1 | Male | 29 | 15+ years | 2017 | 2.297622e+06 | 2.490884e+06 | 2.114574e+06 |
| 20962 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2017 | 2.373815e+05 | 3.217514e+05 | 1.729171e+05 |
| 20963 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2017 | 2.535003e+06 | 2.743769e+06 | 2.341329e+06 |
| 20964 | Number of Smokers | 522 | Sudan | 1 | Male | 29 | 15+ years | 2018 | 2.367072e+06 | 2.575100e+06 | 2.173995e+06 |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 29 | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 |

In [8]:
```
#wy´swietli´c informacje, o ramce danych

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20970 entries, 0 to 20969
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   measure_name    20970 non-null  object 
 1   location_id     20970 non-null  int64  
 2   location_name   20970 non-null  object 
 3   sex_id          20970 non-null  int64  
 4   sex_name        20970 non-null  object 
 5   age_group_id    20970 non-null  int64  
 6   age_group_name  20970 non-null  object 
 7   year_id         20970 non-null  int64  
 8   val             20970 non-null  float64
 9   upper           20970 non-null  float64
 10  lower           20970 non-null  float64
dtypes: float64(3), int64(4), object(4)
memory usage: 1.8+ MB
```

In [9]:
```
#wy´swietli´c, ile wierszy i kolumn znajduje sie, w ramce danych

df.shape
```

Out[9]:
```
(20970, 11)
```

In [10]:
```
#wy´swietli´c informacje, statystyczna, o kolumnach liczbowych (warto´sci
#niepowtarzalne, ´srednia, odchylenie standardowe, minimum, kwartyle,
#maksimum)

df.describe()
```

Out[10]:

| | location_id | sex_id | age_group_id | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|
| count | 20970.000000 | 20970.000000 | 20970.0 | 20970.000000 | 2.097000e+04 | 2.097000e+04 | 2.097000e+04 |
| mean | 131.111588 | 2.000000 | 29.0 | 2004.500000 | 1.242807e+07 | 1.269088e+07 | 1.217241e+07 |
| std | 95.055111 | 0.816516 | 0.0 | 8.655648 | 6.489191e+07 | 6.555971e+07 | 6.421446e+07 |
| min | 1.000000 | 1.000000 | 29.0 | 1990.000000 | 6.345717e+01 | 7.868296e+01 | 5.029157e+01 |
| 25% | 61.000000 | 1.000000 | 29.0 | 1997.000000 | 8.201065e+04 | 9.576943e+04 | 6.875439e+04 |
| 50% | 119.000000 | 2.000000 | 29.0 | 2004.500000 | 5.777123e+05 | 6.278332e+05 | 5.329521e+05 |
| 75% | 177.000000 | 3.000000 | 29.0 | 2012.000000 | 2.901197e+06 | 3.070281e+06 | 2.742651e+06 |
| max | 522.000000 | 3.000000 | 29.0 | 2019.000000 | 1.144819e+09 | 1.157286e+09 | 1.131582e+09 |

In [11]:
```
#wy´swietli´c informacje, statystyczna, o kolumnach kategoryzowanych (ile
#unikalnych warto´sci, top - jaka jest najpopularniejsza warto´s´c, freq -
#jak cze,sto najpopularniejsza


df.describe(include = 'all')
```

Out[11]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 20970 | 20970.000000 | 20970 | 20970.000000 | 20970 | 20970.0 | 20970 | 20970.000000 | 2.097000e+04 | 2.097000e+04 | 2.097000e+04 |
| unique | 1 | NaN | 231 | NaN | 3 | NaN | 1 | NaN | NaN | NaN | NaN |
| top | Number of Smokers | NaN | South Asia | NaN | Male | NaN | 15+ years | NaN | NaN | NaN | NaN |
| freq | 20970 | NaN | 180 | NaN | 6990 | NaN | 20970 | NaN | NaN | NaN | NaN |
| mean | NaN | 131.111588 | NaN | 2.000000 | NaN | 29.0 | NaN | 2004.500000 | 1.242807e+07 | 1.269088e+07 | 1.217241e+07 |
| std | NaN | 95.055111 | NaN | 0.816516 | NaN | 0.0 | NaN | 8.655648 | 6.489191e+07 | 6.555971e+07 | 6.421446e+07 |
| min | NaN | 1.000000 | NaN | 1.000000 | NaN | 29.0 | NaN | 1990.000000 | 6.345717e+01 | 7.868296e+01 | 5.029157e+01 |
| 25% | NaN | 61.000000 | NaN | 1.000000 | NaN | 29.0 | NaN | 1997.000000 | 8.201065e+04 | 9.576943e+04 | 6.875439e+04 |
| 50% | NaN | 119.000000 | NaN | 2.000000 | NaN | 29.0 | NaN | 2004.500000 | 5.777123e+05 | 6.278332e+05 | 5.329521e+05 |
| 75% | NaN | 177.000000 | NaN | 3.000000 | NaN | 29.0 | NaN | 2012.000000 | 2.901197e+06 | 3.070281e+06 | 2.742651e+06 |
| max | NaN | 522.000000 | NaN | 3.000000 | NaN | 29.0 | NaN | 2019.000000 | 1.144819e+09 | 1.157286e+09 | 1.131582e+09 |

In [12]:
```
#usuna,´c brakuja,ce warto´sci w ramce danych

df.dropna(inplace=True)
print(df)
```

```
           measure_name  location_id location_name  sex_id sex_name  \
0      Number of Smokers            1        Global       1     Male
1      Number of Smokers            1        Global       2   Female
2      Number of Smokers            1        Global       3     Both
3      Number of Smokers            1        Global       1     Male
4      Number of Smokers            1        Global       2   Female
...                  ...          ...           ...     ...      ...
20965  Number of Smokers          522         Sudan       2   Female
20966  Number of Smokers          522         Sudan       3     Both
20967  Number of Smokers          522         Sudan       1     Male
20968  Number of Smokers          522         Sudan       2   Female
20969  Number of Smokers          522         Sudan       3     Both

       age_group_id age_group_name  year_id          val         upper  \
0                29      15+ years     1990  8.031015e+08  8.096221e+08
1                29      15+ years     1990  1.891488e+08  1.930929e+08
2                29      15+ years     1990  9.922503e+08  1.000161e+09
3                29      15+ years     1991  8.138972e+08  8.200339e+08
4                29      15+ years     1991  1.905375e+08  1.944249e+08
...             ...            ...      ...           ...           ...
20965            29      15+ years     2018  2.435999e+05  3.286166e+05
20966            29      15+ years     2018  2.610672e+06  2.833943e+06
20967            29      15+ years     2019  2.439150e+06  2.656579e+06
20968            29      15+ years     2019  2.500800e+05  3.345384e+05
20969            29      15+ years     2019  2.689230e+06  2.918332e+06

              lower
0      7.959086e+08
1      1.855595e+08
2      9.847880e+08
3      8.069514e+08
4      1.869744e+08
...             ...
20965  1.752508e+05
20966  2.409108e+06
20967  2.236450e+06
20968  1.816686e+05
20969  2.480656e+06

[20970 rows x 11 columns]
```

In [13]:
```
#przedstawi´c wyb´or wierszy i kolumny u´zywaja,c nazw oraz indeks´ow na
#r´o´zne sposoby

df["measure_name"] # zmienic nazwe po swoj zbior
```

Out[13]:
```
0        Number of Smokers
1        Number of Smokers
2        Number of Smokers
3        Number of Smokers
4        Number of Smokers
              ...
20965    Number of Smokers
20966    Number of Smokers
20967    Number of Smokers
20968    Number of Smokers
20969    Number of Smokers
Name: measure_name, Length: 20970, dtype: object
```

In [14]:
```
df.measure_name # zmienic nazwe pod swoj zbior
```

Out[14]:
```
0        Number of Smokers
1        Number of Smokers
2        Number of Smokers
3        Number of Smokers
4        Number of Smokers
              ...
20965    Number of Smokers
20966    Number of Smokers
20967    Number of Smokers
20968    Number of Smokers
20969    Number of Smokers
Name: measure_name, Length: 20970, dtype: object
```

In [15]:
```
df[["measure_name","age_group_name","year_id"]] # wybór kilku kolumn jednocześnie zmienic pod swoj zbior
```

Out[15]:

| | measure_name | age_group_name | year_id |
|---|---|---|---|
| 0 | Number of Smokers | 15+ years | 1990 |
| 1 | Number of Smokers | 15+ years | 1990 |
| 2 | Number of Smokers | 15+ years | 1990 |
| 3 | Number of Smokers | 15+ years | 1991 |
| 4 | Number of Smokers | 15+ years | 1991 |
| ... | ... | ... | ... |
| 20965 | Number of Smokers | 15+ years | 2018 |
| 20966 | Number of Smokers | 15+ years | 2018 |
| 20967 | Number of Smokers | 15+ years | 2019 |
| 20968 | Number of Smokers | 15+ years | 2019 |
| 20969 | Number of Smokers | 15+ years | 2019 |

20970 rows × 3 columns

In [16]:
```
df.loc[:, "location_name":"val"] # wybierz wszystkie wiersze i kolumny od „location" do „val" zmienic pod swoj zbior
```

Out[16]:

| | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val |
|---|---|---|---|---|---|---|---|
| 0 | Global | 1 | Male | 29 | 15+ years | 1990 | 8.031015e+08 |
| 1 | Global | 2 | Female | 29 | 15+ years | 1990 | 1.891488e+08 |
| 2 | Global | 3 | Both | 29 | 15+ years | 1990 | 9.922503e+08 |
| 3 | Global | 1 | Male | 29 | 15+ years | 1991 | 8.138972e+08 |
| 4 | Global | 2 | Female | 29 | 15+ years | 1991 | 1.905375e+08 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Sudan | 2 | Female | 29 | 15+ years | 2018 | 2.435999e+05 |
| 20966 | Sudan | 3 | Both | 29 | 15+ years | 2018 | 2.610672e+06 |
| 20967 | Sudan | 1 | Male | 29 | 15+ years | 2019 | 2.439150e+06 |
| 20968 | Sudan | 2 | Female | 29 | 15+ years | 2019 | 2.500800e+05 |
| 20969 | Sudan | 3 | Both | 29 | 15+ years | 2019 | 2.689230e+06 |

20970 rows × 7 columns

In [17]:
```
df.loc[100:110, "location_name":"val"] #wybierz wiersze od 100-110 i kolumny od locationa_name do val
```

Out[17]:

| | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val |
|---|---|---|---|---|---|---|---|
| 100 | Southeast Asia, East Asia, and Oceania | 2 | Female | 29 | 15+ years | 1993 | 2.843611e+07 |
| 101 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 1993 | 3.984578e+08 |
| 102 | Southeast Asia, East Asia, and Oceania | 1 | Male | 29 | 15+ years | 1994 | 3.744452e+08 |
| 103 | Southeast Asia, East Asia, and Oceania | 2 | Female | 29 | 15+ years | 1994 | 2.908238e+07 |
| 104 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 1994 | 4.035276e+08 |
| 105 | Southeast Asia, East Asia, and Oceania | 1 | Male | 29 | 15+ years | 1995 | 3.779077e+08 |
| 106 | Southeast Asia, East Asia, and Oceania | 2 | Female | 29 | 15+ years | 1995 | 2.964857e+07 |
| 107 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 1995 | 4.075563e+08 |
| 108 | Southeast Asia, East Asia, and Oceania | 1 | Male | 29 | 15+ years | 1996 | 3.803069e+08 |
| 109 | Southeast Asia, East Asia, and Oceania | 2 | Female | 29 | 15+ years | 1996 | 3.015084e+07 |
| 110 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 1996 | 4.104577e+08 |

In [18]:
```
df.iloc[100:110, 0:3] #wybierz wiersze od 100-110 i kolumny od 0-2
```

Out[18]:

| | measure_name | location_id | location_name |
|---|---|---|---|
| 100 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 101 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 102 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 103 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 104 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 105 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 106 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 107 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 108 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |
| 109 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania |

In [19]:
```
#przedstawi´c wyb´or wierszy z ramki danych pod warunkiem odno´snie
#okre´slonej warto´sci kolumny

df[df["sex_name"] == "Both"]
```

Out[19]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 |
| 5 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1991 | 1.004435e+09 | 1.011925e+09 | 9.969811e+08 |
| 8 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1992 | 1.015217e+09 | 1.022720e+09 | 1.007847e+09 |
| 11 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1993 | 1.024669e+09 | 1.031965e+09 | 1.017551e+09 |
| 14 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1994 | 1.032567e+09 | 1.039842e+09 | 1.025631e+09 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20957 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2015 | 2.388216e+06 | 2.587005e+06 | 2.211144e+06 |
| 20960 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2016 | 2.454893e+06 | 2.665441e+06 | 2.267696e+06 |
| 20963 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2017 | 2.535003e+06 | 2.743769e+06 | 2.341329e+06 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 |

6990 rows × 11 columns

In [20]:
```
#przedstawi´c wyb´or wierszy z ramki danych pod warunkiem spelnienia
#kilku warunk´ow jednocze´snie

cardio = df[(df["sex_name"] == "Both")&(df["year_id"] == 2018) & (df["age_group_id"] <= 29)]
cardio
```

Out[20]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 2018 | 1.138485e+09 | 1.150113e+09 | 1.125664e+09 |
| 176 | Number of Smokers | 4 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 2018 | 4.853034e+08 | 4.940899e+08 | 4.759853e+08 |
| 266 | Number of Smokers | 5 | East Asia | 3 | Both | 29 | 15+ years | 2018 | 3.500440e+08 | 3.588876e+08 | 3.412120e+08 |
| 356 | Number of Smokers | 6 | China | 3 | Both | 29 | 15+ years | 2018 | 3.401172e+08 | 3.488947e+08 | 3.312153e+08 |
| 446 | Number of Smokers | 7 | Democratic People's Republic of Korea | 3 | Both | 29 | 15+ years | 2018 | 5.300059e+06 | 5.553985e+06 | 5.024278e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20606 | Number of Smokers | 413 | Tokelau | 3 | Both | 29 | 15+ years | 2018 | 2.612484e+02 | 2.848032e+02 | 2.394825e+02 |
| 20696 | Number of Smokers | 416 | Tuvalu | 3 | Both | 29 | 15+ years | 2018 | 2.793366e+03 | 2.956539e+03 | 2.613007e+03 |
| 20786 | Number of Smokers | 422 | United States Virgin Islands | 3 | Both | 29 | 15+ years | 2018 | 5.633536e+03 | 6.212418e+03 | 5.090184e+03 |
| 20876 | Number of Smokers | 435 | South Sudan | 3 | Both | 29 | 15+ years | 2018 | 5.181444e+05 | 5.736592e+05 | 4.677617e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 |

233 rows × 11 columns

In [21]:
```
# wybra´c wiersze kt´ore zawieraja¸ w kolumnie kategoryzowanej okre´slone slowo

df[df["location_name"].str.contains("States")]
```

Out[21]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1980 | Number of Smokers | 25 | Micronesia (Federated States of) | 1 | Male | 29 | 15+ years | 1990 | 18134.775290 | 19169.248820 | 17155.196930 |
| 1981 | Number of Smokers | 25 | Micronesia (Federated States of) | 2 | Female | 29 | 15+ years | 1990 | 9470.305481 | 11156.303110 | 7825.944174 |
| 1982 | Number of Smokers | 25 | Micronesia (Federated States of) | 3 | Both | 29 | 15+ years | 1990 | 27605.080770 | 29580.226920 | 25829.741340 |
| 1983 | Number of Smokers | 25 | Micronesia (Federated States of) | 1 | Male | 29 | 15+ years | 1991 | 18395.672830 | 19459.617700 | 17385.018410 |
| 1984 | Number of Smokers | 25 | Micronesia (Federated States of) | 2 | Female | 29 | 15+ years | 1991 | 9658.519070 | 11404.994170 | 7961.453848 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20785 | Number of Smokers | 422 | United States Virgin Islands | 2 | Female | 29 | 15+ years | 2018 | 2308.376511 | 2820.434508 | 1871.029388 |
| 20786 | Number of Smokers | 422 | United States Virgin Islands | 3 | Both | 29 | 15+ years | 2018 | 5633.535832 | 6212.418101 | 5090.184376 |
| 20787 | Number of Smokers | 422 | United States Virgin Islands | 1 | Male | 29 | 15+ years | 2019 | 3280.527338 | 3649.862482 | 2939.996840 |
| 20788 | Number of Smokers | 422 | United States Virgin Islands | 2 | Female | 29 | 15+ years | 2019 | 2282.281664 | 2813.914814 | 1831.778372 |
| 20789 | Number of Smokers | 422 | United States Virgin Islands | 3 | Both | 29 | 15+ years | 2019 | 5562.809002 | 6146.429254 | 4990.914042 |

270 rows × 11 columns

In [22]:
```
# wybra´c wiersze kt´ore nie zawieraja¸ w kolumnie kategoryzowanej okre´slone
# slowo

df[df["location_name"].str.contains("States") == False]
```

Out[22]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 29 | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 |

20700 rows × 11 columns

In [23]:
```
#utw´orz kolumne¸ na podstawie istnieja¸cyn

df["Tolerance_range"] = df["upper"] - df["lower"]
df
```

Out[23]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 29 | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 29 | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 29 | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 29 | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 29 | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 29 | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

20970 rows × 12 columns

In [24]:
```
#usuń kolumnę

df = df.drop("age_group_id", axis = 1)
df
```

Out[24]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

20970 rows × 11 columns

In [25]:
```
#zmień nazwę kolumny

df = df.rename(columns = {"sex_name":"sex"})
df
```

Out[25]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

20970 rows × 11 columns

In [26]:
```
df
```

Out[26]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

In [27]:
```
#zachowaj ramke, danych jako plik csv na komputerze

df.to_csv("New_DataFrame.csv")
```

In [28]:
```
df
```

Out[28]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

20970 rows × 11 columns

In [29]:
```
#wy´swietli´c ´srednia (maksymalna,, minimalna,) warto´s´c z jednej kolumny

print(df["year_id"].mean())
print(df["year_id"].max())
print(df["year_id"].min())
```

```
2004.5
2019
1990
```

In [30]:
```
#wy´swietli´c liczbe, wierszy

rows = len(df.axes[0])
rows
```

Out[30]:
```
20970
```

In [31]:
```
#wy´swietli´c warto´sci unikatowe w kolumnie

df['sex'].unique() # wartości unikatowe
```

Out[31]:
```
array(['Male', 'Female', 'Both'], dtype=object)
```

In [32]:
```
#wy´swietli´c liczby rekord´ow odpowiadaja,cych do warto´sci

df['sex'].value_counts() # Liczba rekordów pasujących do unikalnych wartości
```

Out[32]:
```
sex
Male      6990
Female    6990
Both      6990
Name: count, dtype: int64
```

In [33]:
```
#sortowanie wierszy ramki danych wedlug warto´sci okre´slonej kolumny
#(maleja,co, rosna,co)

df.sort_values(['val'], ascending = True) # sortowanie rosnąco
```

Out[33]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20572 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2007 | 6.345717e+01 | 7.868296e+01 | 5.029157e+01 | 2.839139e+01 |
| 20575 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2008 | 6.373069e+01 | 7.947261e+01 | 5.046282e+01 | 2.900979e+01 |
| 20569 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2006 | 6.493000e+01 | 8.077211e+01 | 5.130362e+01 | 2.946849e+01 |
| 20578 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2009 | 6.519800e+01 | 8.157621e+01 | 5.118533e+01 | 3.039089e+01 |
| 20581 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2010 | 6.727737e+01 | 8.381787e+01 | 5.291036e+01 | 3.090751e+01 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 71 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2013 | 1.132095e+09 | 1.140134e+09 | 1.123371e+09 | 1.676275e+07 |
| 68 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2012 | 1.132823e+09 | 1.140135e+09 | 1.124584e+09 | 1.555096e+07 |
| 83 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2017 | 1.133641e+09 | 1.144223e+09 | 1.121587e+09 | 2.263597e+07 |
| 86 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2018 | 1.138485e+09 | 1.150113e+09 | 1.125664e+09 | 2.444870e+07 |
| 89 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2019 | 1.144819e+09 | 1.157286e+09 | 1.131582e+09 | 2.570451e+07 |

20970 rows × 11 columns

In [34]:
```
df.sort_values(['val'], ascending = False) # sortowanie malejąco
```

Out[34]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 89 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2019 | 1.144819e+09 | 1.157286e+09 | 1.131582e+09 | 2.570451e+07 |
| 86 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2018 | 1.138485e+09 | 1.150113e+09 | 1.125664e+09 | 2.444870e+07 |
| 83 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2017 | 1.133641e+09 | 1.144223e+09 | 1.121587e+09 | 2.263597e+07 |
| 68 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2012 | 1.132823e+09 | 1.140135e+09 | 1.124584e+09 | 1.555098e+07 |
| 71 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2013 | 1.132095e+09 | 1.140134e+09 | 1.123371e+09 | 1.676275e+07 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20581 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2010 | 6.727737e+01 | 8.381787e+01 | 5.291036e+01 | 3.090751e+01 |
| 20578 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2009 | 6.519800e+01 | 8.157621e+01 | 5.118533e+01 | 3.039089e+01 |
| 20569 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2005 | 6.493000e+01 | 8.077211e+01 | 5.130362e+01 | 2.946849e+01 |
| 20575 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2008 | 6.373069e+01 | 7.947261e+01 | 5.046282e+01 | 2.900979e+01 |
| 20572 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2007 | 6.345717e+01 | 7.868296e+01 | 5.029157e+01 | 2.839139e+01 |

20970 rows × 11 columns

In [35]:
```
df.sort_values(['val'], ascending = True).head(10) # 10 Największych
```

Out[35]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20572 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2007 | 63.457166 | 78.682957 | 50.291565 | 28.391392 |
| 20575 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2008 | 63.730688 | 79.472611 | 50.462818 | 29.009794 |
| 20569 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2005 | 64.929999 | 80.772113 | 51.303621 | 29.468492 |
| 20578 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2009 | 65.197998 | 81.576212 | 51.185326 | 30.390887 |
| 20581 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2010 | 67.277366 | 83.817868 | 52.910362 | 30.907506 |
| 20566 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2005 | 68.832209 | 86.262158 | 54.075288 | 32.186870 |
| 20584 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2011 | 68.910776 | 85.482803 | 54.302381 | 31.180422 |
| 20587 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2012 | 69.935769 | 86.894159 | 55.018949 | 31.875210 |
| 20590 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2013 | 70.365558 | 88.023339 | 54.985971 | 33.037368 |
| 20593 | Number of Smokers | 413 | Tokelau | 2 | Female | 15+ years | 2014 | 70.913538 | 88.504042 | 55.650310 | 32.853732 |

In [36]:
```
df.sort_values(['val'], ascending = False).head(10) # 10 Najmniejszych
```

Out[36]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 89 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2019 | 1.144819e+09 | 1.157286e+09 | 1.131582e+09 | 25704510.0 |
| 86 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2018 | 1.138485e+09 | 1.150113e+09 | 1.125664e+09 | 24448697.0 |
| 83 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2017 | 1.133641e+09 | 1.144223e+09 | 1.121587e+09 | 22635972.0 |
| 68 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2012 | 1.132823e+09 | 1.140135e+09 | 1.124584e+09 | 15550978.0 |
| 71 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2013 | 1.132095e+09 | 1.140134e+09 | 1.123371e+09 | 16762752.0 |
| 80 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2016 | 1.131316e+09 | 1.141395e+09 | 1.120251e+09 | 21143874.0 |
| 65 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2011 | 1.131118e+09 | 1.137976e+09 | 1.123491e+09 | 14484589.0 |
| 74 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2014 | 1.131104e+09 | 1.139929e+09 | 1.121633e+09 | 18295997.0 |
| 77 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2015 | 1.130417e+09 | 1.139826e+09 | 1.120117e+09 | 19709146.0 |
| 62 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 2010 | 1.127518e+09 | 1.134409e+09 | 1.120340e+09 | 14069177.0 |

In [37]:
```
#wyświetlić wierszy dla 10 najwiekszych wartości określonej kolumny
#pod warunkiem określonych wartości innej kolumny

df[(df['location_name'].isin(['Sudan','Poland','Hungary']))].nlargest(10,'val')
```

Out[37]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4322 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1990 | 11704480.59 | 12012219.84 | 1.138875e+07 | 623474.230 |
| 4325 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1991 | 11576469.57 | 11885703.34 | 1.125896e+07 | 626738.790 |
| 4328 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1992 | 11445059.00 | 11758271.45 | 1.112616e+07 | 632108.160 |
| 4331 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1993 | 11299362.83 | 11629281.89 | 1.097590e+07 | 653384.360 |
| 4334 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1994 | 11144876.31 | 11470659.42 | 1.080890e+07 | 661758.370 |
| 4337 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1995 | 10979602.98 | 11324575.92 | 1.062839e+07 | 696184.240 |
| 4340 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1996 | 10805578.46 | 11158004.92 | 1.045887e+07 | 699130.530 |
| 4343 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1997 | 10632641.54 | 10983286.74 | 1.027121e+07 | 712075.690 |
| 4346 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1998 | 10474421.99 | 10830307.69 | 1.012083e+07 | 709474.650 |
| 4349 | Number of Smokers | 51 | Poland | 3 | Both | 15+ years | 1999 | 10314177.42 | 10684277.63 | 9.960286e+06 | 723991.242 |

In [38]:
```
#grupowanie wierszy wedlug warto´sci kolumny kategoryzowanej, potem
#- u´srednienie warto´sci wszystkich kolumn w grupie - MultiIndex

df_new = df.groupby(['location_name','sex']).agg({'val': 'mean',
                                                   'upper': 'mean',
                                                   'lower': 'mean'})
df_new
```

Out[38]:

| | | val | upper | lower |
|---|---|---|---|---|
| location_name | sex | | | |
| Afghanistan | Both | 1.076844e+06 | 1.184427e+06 | 9.776876e+05 |
| | Female | 1.408633e+05 | 1.867379e+05 | 1.060589e+05 |
| | Male | 9.359803e+05 | 1.037830e+06 | 8.447279e+05 |
| Albania | Both | 6.016696e+05 | 6.302436e+05 | 5.752316e+05 |
| | Female | 1.060032e+05 | 1.248055e+05 | 8.917709e+04 |
| ... | ... | ... | ... | ... |
| Zambia | Female | 2.285319e+05 | 2.766568e+05 | 1.879562e+05 |
| | Male | 7.708210e+05 | 8.156664e+05 | 7.266267e+05 |
| Zimbabwe | Both | 1.075152e+06 | 1.132936e+06 | 1.018202e+06 |
| | Female | 1.164704e+05 | 1.442346e+05 | 9.511072e+04 |
| | Male | 9.586813e+05 | 1.010215e+06 | 9.072602e+05 |

693 rows × 3 columns

In [39]:
```
#grupowanie wierszy wedlug warto´sci kolumny kategoryzowanej, potem
#- u´srednienie warto´sci dla pewnych kolumn, liczba warto´sci i mediana
#dla pozostalych kolumn w grupach

df_new = df.groupby(['location_name','sex']).agg({
                        'val': 'mean',
                        'upper': ['median','count'],
                        'sex_id' : ['median','count']})
df_new
```

Out[39]:

| | | val | upper | | sex_id | |
|---|---|---|---|---|---|---|
| | | mean | median | count | median | count |
| location_name | sex | | | | | |
| Afghanistan | Both | 1.076844e+06 | 1.051483e+06 | 30 | 3.0 | 30 |
| | Female | 1.408633e+05 | 1.583590e+05 | 30 | 2.0 | 30 |
| | Male | 9.359803e+05 | 9.289502e+05 | 30 | 1.0 | 30 |
| Albania | Both | 6.016696e+05 | 5.955954e+05 | 30 | 3.0 | 30 |
| | Female | 1.060032e+05 | 1.133635e+05 | 30 | 2.0 | 30 |
| ... | ... | ... | ... | ... | ... | ... |
| Zambia | Female | 2.285319e+05 | 2.926822e+05 | 30 | 2.0 | 30 |
| | Male | 7.708210e+05 | 7.427456e+05 | 30 | 1.0 | 30 |
| Zimbabwe | Both | 1.075152e+06 | 1.065280e+06 | 30 | 3.0 | 30 |
| | Female | 1.164704e+05 | 1.521148e+05 | 30 | 2.0 | 30 |
| | Male | 9.586813e+05 | 9.324646e+05 | 30 | 1.0 | 30 |

693 rows × 5 columns

In [40]:
```
#wy´swietli´c nazwy kolumn indeksu zlo˙zonego

df_new.columns
```

Out[40]:
```
MultiIndex([(   'val',    'mean'),
            ( 'upper',  'median'),
            ( 'upper',   'count'),
            ('sex_id',  'median'),
            ('sex_id',   'count')],
           )
```

In [41]:
```
#sortowa´c kolumne˛ indeksu zlo˙zonego

df_new['upper']['median'].sort_values(ascending = True)
```

Out[41]:
```
location_name                          sex
Tokelau                                Female    9.568048e+01
Niue                                   Female    1.109018e+02
                                       Male      1.823211e+02
Tokelau                                Male      1.978756e+02
Niue                                   Both      2.831822e+02
                                                     ...
East Asia                              Both      3.135026e+08
Southeast Asia, East Asia, and Oceania Male      3.981023e+08
                                       Both      4.314121e+08
Global                                 Male      8.892060e+08
                                       Both      1.097092e+09
Name: median, Length: 693, dtype: float64
```

In [42]:
```
#stworzyć tabelę przystawną (pivot table) na podstawie ramki danych

df_pivot = df.pivot_table(values='val', index='location_name', columns='sex', aggfunc='mean',
                          margins=False, dropna=True, fill_value=None) # tabela podsumowująca
df_pivot
```

Out[42]:

| sex | Both | Female | Male |
|---|---|---|---|
| location_name | | | |
| Afghanistan | 1.076844e+06 | 1.408633e+05 | 9.359803e+05 |
| Albania | 6.016696e+05 | 1.060032e+05 | 4.956664e+05 |
| Algeria | 3.873312e+06 | 2.192907e+05 | 3.654021e+06 |
| American Samoa | 1.225455e+04 | 4.320555e+03 | 7.933993e+03 |
| Andean Latin America | 3.566739e+06 | 9.573733e+05 | 2.609366e+06 |
| ... | ... | ... | ... |
| Western Europe | 9.714860e+07 | 4.233137e+07 | 5.481723e+07 |
| Western Sub-Saharan Africa | 1.377671e+07 | 1.868140e+06 | 1.190857e+07 |
| Yemen | 2.233432e+06 | 4.807306e+05 | 1.752701e+06 |
| Zambia | 9.993529e+05 | 2.285319e+05 | 7.708210e+05 |
| Zimbabwe | 1.075152e+06 | 1.164704e+05 | 9.586813e+05 |

231 rows × 3 columns

In [43]:
```
#wyświetlić indeksy i kolumny tabeli przystawnej

print(df_pivot.index)
print(df_pivot.columns)
```

```
Index(['Afghanistan', 'Albania', 'Algeria', 'American Samoa',
       'Andean Latin America', 'Andorra', 'Angola', 'Antigua and Barbuda',
       'Argentina', 'Armenia',
       ...
       'Uruguay', 'Uzbekistan', 'Vanuatu',
       'Venezuela (Bolivarian Republic of)', 'Viet Nam', 'Western Europe',
       'Western Sub-Saharan Africa', 'Yemen', 'Zambia', 'Zimbabwe'],
      dtype='object', name='location_name', length=231)
Index(['Both', 'Female', 'Male'], dtype='object', name='sex')
```

In [44]:
```
#utworz indeks zlozony tabeli przystawnej i wyświetl go

df_pivot = df.pivot_table(values='val', index=['location_name','year_id'], columns='sex', aggfunc='mean',
                          margins=False, dropna=True, fill_value=None)
df_pivot
```

Out[44]:

| | sex | Both | Female | Male |
|---|---|---|---|---|
| location_name | year_id | | | |
| Afghanistan | 1990 | 1.952669e+05 | 33006.33880 | 1.622605e+05 |
| | 1991 | 2.296694e+05 | 37383.30729 | 1.922861e+05 |
| | 1992 | 2.895567e+05 | 44365.25615 | 2.451914e+05 |
| | 1993 | 3.435076e+05 | 50238.47594 | 2.932691e+05 |
| | 1994 | 3.806563e+05 | 54130.87718 | 3.265254e+05 |
| ... | ... | ... | ... | ... |
| Zimbabwe | 2015 | 1.327739e+06 | 141442.90390 | 1.186297e+06 |
| | 2016 | 1.361949e+06 | 141928.43040 | 1.220020e+06 |
| | 2017 | 1.396007e+06 | 141905.15710 | 1.254102e+06 |
| | 2018 | 1.430277e+06 | 141451.47270 | 1.288825e+06 |
| | 2019 | 1.465099e+06 | 140772.04550 | 1.324327e+06 |

6930 rows × 3 columns

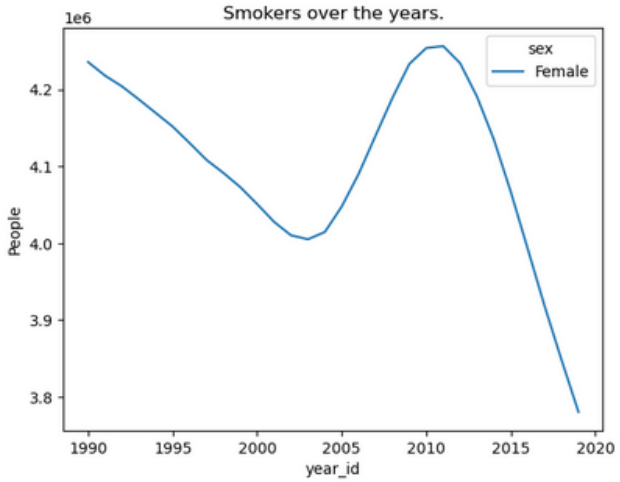
In [45]:
```
#zaimportuj modul pyplot z biblioteki matplotlib
import matplotlib.pyplot as plt
```

In [46]:
```
#wy´swietli´c wykres na podstawie tabeli przystawnej

df[(df['location_name'] == 'Poland') & (df['sex'] == 'Female')].pivot_table(values='val', index='year_id', columns='sex', aggfunc='mean',
                                        fill_value=None, margins=False, dropna=True).plot(kind = 'line')

plt.ylabel('People') # etykieta osi y
plt.title('Smokers over the years.') # tytuł wykresu
```

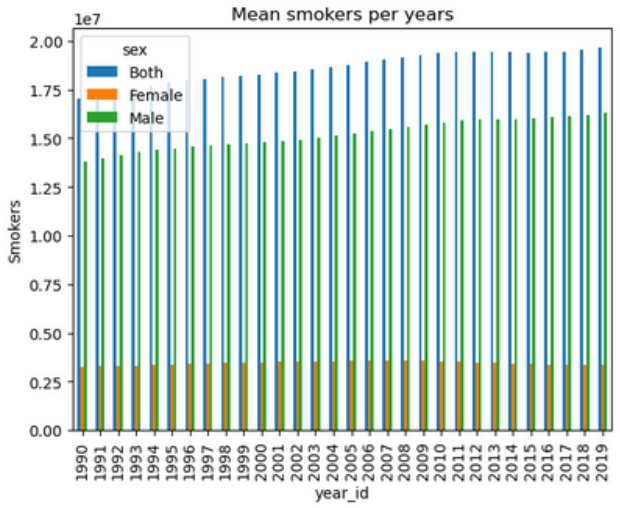
Out[46]:
```
Text(0.5, 1.0, 'Smokers over the years.')
```




In [47]:
```
#narysowa´c histogram na podstawie warto´sci kolumny

df_bar = df[(df['sex'].isin(['Male','Female','Both']))].pivot_table(values='val',
                index='year_id', columns='sex', aggfunc='mean',
                        fill_value=None, margins=False, dropna=True)
df_bar.plot(kind = 'bar')
plt.ylabel('Smokers')
plt.title('Mean smokers per years')
```

Out[47]:
```
Text(0.5, 1.0, 'Mean smokers per years')
```



In [48]:
```
#przedstawić sposoby łączenia ramek danych za pomocą metod merge i
#concat

df2 = pd.read_csv('IHME_GBD_2019_SMOKING_TOB_1990_2019_CIG_PC_Y2021M05D27.CSV')
df2
```

Out[48]:

| | measure_name | location_id | location_name | sex_id | sex_name | age_group_id | age_group_name | year_id | val | upper | lower |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Cigarette-Equivalents Per Capita | 1 | Global | 3 | Both | 29 | 15+ years | 1990 | 1484.256502 | 1531.563739 | 1436.151878 |
| 1 | Cigarette-Equivalents Per Capita | 1 | Global | 3 | Both | 29 | 15+ years | 2019 | 1113.754663 | 1161.263946 | 1069.765828 |
| 2 | Cigarette-Equivalents Per Capita | 4 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 1990 | 1827.374739 | 1959.359086 | 1692.900863 |
| 3 | Cigarette-Equivalents Per Capita | 4 | Southeast Asia, East Asia, and Oceania | 3 | Both | 29 | 15+ years | 2019 | 1778.846098 | 1927.560165 | 1640.645875 |
| 4 | Cigarette-Equivalents Per Capita | 5 | East Asia | 3 | Both | 29 | 15+ years | 1990 | 2089.743405 | 2267.199999 | 1908.301510 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 461 | Cigarette-Equivalents Per Capita | 422 | United States Virgin Islands | 3 | Both | 29 | 15+ years | 2019 | 648.023999 | 821.503370 | 497.645622 |
| 462 | Cigarette-Equivalents Per Capita | 435 | South Sudan | 3 | Both | 29 | 15+ years | 1990 | 337.087376 | 431.023296 | 256.275436 |
| 463 | Cigarette-Equivalents Per Capita | 435 | South Sudan | 3 | Both | 29 | 15+ years | 2019 | 317.318526 | 410.888023 | 241.957923 |
| 464 | Cigarette-Equivalents Per Capita | 522 | Sudan | 3 | Both | 29 | 15+ years | 1990 | 170.613365 | 200.058577 | 142.371142 |
| 465 | Cigarette-Equivalents Per Capita | 522 | Sudan | 3 | Both | 29 | 15+ years | 2019 | 227.024745 | 279.919615 | 180.773129 |

466 rows × 11 columns

In [49]:
```
df
```

Out[49]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

20970 rows × 11 columns

In [50]:
```
df2.rename(columns = {'val': 'val_Cigarette-Equivalents Per Capita', 'upper': 'upper_Cigarette-Equivalents Per Capita', 'lower':'deaths_lc
df.drop('sex_id', axis = 1)
```

Out[50]:

| | measure_name | location_id | location_name | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 |
| 1 | Number of Smokers | 1 | Global | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20965 | Number of Smokers | 522 | Sudan | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

In [51]:
```
df_all = pd.merge(df, df2, on = ['location_id', 'location_name', 'age_group_name', 'year_id'], how = 'inner')
```

In [52]:
```
df_all
```

Out[52]:

| | measure_name_x | location_id | location_name | sex_id_x | sex | age_group_name | year_id | val | upper | lower | Tolerance_range | measure_name_y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1990 | 8.031015e+08 | 8.096221e+08 | 7.959086e+08 | 1.371347e+07 | Cigarette-Equivalents Per Capita |
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 | Cigarette-Equivalents Per Capita |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 | Cigarette-Equivalents Per Capita |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 2019 | 9.505177e+08 | 9.615655e+08 | 9.392571e+08 | 2.230844e+07 | Cigarette-Equivalents Per Capita |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 2019 | 1.943009e+08 | 2.004529e+08 | 1.885878e+08 | 1.186517e+07 | Cigarette-Equivalents Per Capita |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1393 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 1990 | 1.295362e+05 | 1.719868e+05 | 9.532772e+04 | 7.665907e+04 | Cigarette-Equivalents Per Capita |
| 1394 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 1990 | 1.340050e+06 | 1.481698e+06 | 1.204444e+06 | 2.772542e+05 | Cigarette-Equivalents Per Capita |
| 1395 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 | Cigarette-Equivalents Per Capita |
| 1396 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 | Cigarette-Equivalents Per Capita |
| 1397 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 | Cigarette-Equivalents Per Capita |

1398 rows × 18 columns

In [53]:
```
df_all_1 = df.iloc[1:15,:]
df_all_1
```

Out[53]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7533418.8 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 15373214.2 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 13082478.1 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7450504.2 |
| 5 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1991 | 1.004435e+09 | 1.011925e+09 | 9.969811e+08 | 14943782.9 |
| 6 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1992 | 8.233148e+08 | 8.292228e+08 | 8.167264e+08 | 12496456.0 |
| 7 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1992 | 1.919026e+08 | 1.957109e+08 | 1.884066e+08 | 7304269.8 |
| 8 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1992 | 1.015217e+09 | 1.022720e+09 | 1.007847e+09 | 14873159.0 |
| 9 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1993 | 8.313873e+08 | 8.372931e+08 | 8.249496e+08 | 12343464.8 |
| 10 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1993 | 1.932818e+08 | 1.970626e+08 | 1.898392e+08 | 7223414.6 |
| 11 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1993 | 1.024669e+09 | 1.031965e+09 | 1.017551e+09 | 14413782.0 |
| 12 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1994 | 8.378204e+08 | 8.437233e+08 | 8.316340e+08 | 12089304.4 |
| 13 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1994 | 1.947462e+08 | 1.985205e+08 | 1.913568e+08 | 7163636.7 |
| 14 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1994 | 1.032567e+09 | 1.039842e+09 | 1.025631e+09 | 14211884.0 |

In [54]:
```
df_all_2 = df.iloc[-15::]
df_all_2
```

Out[54]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20955 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2015 | 2.159385e+06 | 2.329364e+06 | 1.990166e+06 | 339198.3440 |
| 20956 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2015 | 2.288306e+05 | 3.056884e+05 | 1.694027e+05 | 136285.7259 |
| 20957 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2015 | 2.388216e+06 | 2.587005e+06 | 2.211144e+06 | 375861.2900 |
| 20958 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2016 | 2.223791e+06 | 2.402490e+06 | 2.048482e+06 | 354008.3570 |
| 20959 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2016 | 2.311014e+05 | 3.101747e+05 | 1.700120e+05 | 140162.7220 |
| 20960 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2016 | 2.454893e+06 | 2.665441e+06 | 2.267696e+06 | 397744.9040 |
| 20961 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2017 | 2.297622e+06 | 2.490884e+06 | 2.114574e+06 | 376309.7400 |
| 20962 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2017 | 2.373815e+05 | 3.217514e+05 | 1.729171e+05 | 148834.2947 |
| 20963 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2017 | 2.535003e+06 | 2.743769e+06 | 2.341329e+06 | 402439.6570 |
| 20964 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2018 | 2.367072e+06 | 2.575100e+06 | 2.173995e+06 | 401105.5250 |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 153365.8312 |

In [55]:
```
df_all_new = pd.concat([df_all_1, df_all_2], axis = 0) # połącz ramki danych: jeśli axis = 0, to po wierszach, jeśli
# axis = 1, potem według kolumn
print(df_all_new.shape)
df_all_new
```

```
(29, 11)
```

Out[55]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 |
| 5 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1991 | 1.004435e+09 | 1.011925e+09 | 9.969811e+08 | 1.494378e+07 |
| 6 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1992 | 8.233148e+08 | 8.292228e+08 | 8.167264e+08 | 1.249646e+07 |
| 7 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1992 | 1.919026e+08 | 1.957109e+08 | 1.884066e+08 | 7.304270e+06 |
| 8 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1992 | 1.015217e+09 | 1.022720e+09 | 1.007847e+09 | 1.487316e+07 |
| 9 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1993 | 8.313873e+08 | 8.372931e+08 | 8.249496e+08 | 1.234346e+07 |
| 10 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1993 | 1.932818e+08 | 1.970626e+08 | 1.898392e+08 | 7.223415e+06 |
| 11 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1993 | 1.024669e+09 | 1.031965e+09 | 1.017551e+09 | 1.441378e+07 |
| 12 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1994 | 8.378204e+08 | 8.437233e+08 | 8.316340e+08 | 1.208930e+07 |
| 13 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1994 | 1.947462e+08 | 1.985205e+08 | 1.913568e+08 | 7.163637e+06 |
| 14 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1994 | 1.032567e+09 | 1.039842e+09 | 1.025631e+09 | 1.421188e+07 |
| 20955 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2015 | 2.159385e+06 | 2.329364e+06 | 1.990166e+06 | 3.391983e+05 |
| 20956 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2015 | 2.288306e+05 | 3.056884e+05 | 1.694027e+05 | 1.362857e+05 |
| 20957 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2015 | 2.388216e+06 | 2.587005e+06 | 2.211144e+06 | 3.758613e+05 |
| 20958 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2016 | 2.223791e+06 | 2.402490e+06 | 2.048482e+06 | 3.540084e+05 |
| 20959 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2016 | 2.311014e+05 | 3.101747e+05 | 1.700120e+05 | 1.401627e+05 |
| 20960 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2016 | 2.454893e+06 | 2.665441e+06 | 2.267696e+06 | 3.977449e+05 |
| 20961 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2017 | 2.297622e+06 | 2.490884e+06 | 2.114574e+06 | 3.763097e+05 |
| 20962 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2017 | 2.373815e+05 | 3.217514e+05 | 1.729171e+05 | 1.488343e+05 |
| 20963 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2017 | 2.535003e+06 | 2.743769e+06 | 2.341329e+06 | 4.024397e+05 |
| 20964 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2018 | 2.367072e+06 | 2.575100e+06 | 2.173995e+06 | 4.011055e+05 |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 |

In [56]:
```
#pokaza´c dodawanie nowych kolumn za pomoca¸ operacji matematy-
#cznych

df_all_new["smokers"] = df_all_new["val"] + df_all_new["upper"] + df_all_new["lower"]
df_all_new["%ValFromUpper"] = df_all_new["val"] / df_all_new["upper"]*100
df_all_new["%ValFromLower"] = df_all_new["val"] / df_all_new["lower"]*100
```

In [57]:
```
df_all_new
```

Out[57]:

| | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range | smokers | %Val |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 | 5.678012e+08 | |
| 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 | 2.977200e+09 | |
| 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 | 2.440883e+09 | |
| 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 | 5.719369e+08 | |
| 5 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1991 | 1.004435e+09 | 1.011925e+09 | 9.969811e+08 | 1.494378e+07 | 3.013341e+09 | |
| 6 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1992 | 8.233148e+08 | 8.292228e+08 | 8.167264e+08 | 1.249646e+07 | 2.469264e+09 | |
| 7 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1992 | 1.919026e+08 | 1.957109e+08 | 1.884066e+08 | 7.304270e+06 | 5.760201e+08 | |
| 8 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1992 | 1.015217e+09 | 1.022720e+09 | 1.007847e+09 | 1.487316e+07 | 3.045784e+09 | |
| 9 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1993 | 8.313873e+08 | 8.372931e+08 | 8.249496e+08 | 1.234346e+07 | 2.493630e+09 | |
| 10 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1993 | 1.932818e+08 | 1.970626e+08 | 1.898392e+08 | 7.223415e+06 | 5.801836e+08 | |
| 11 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1993 | 1.024669e+09 | 1.031965e+09 | 1.017551e+09 | 1.441378e+07 | 3.074184e+09 | |
| 12 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1994 | 8.378204e+08 | 8.437233e+08 | 8.316340e+08 | 1.208930e+07 | 2.513178e+09 | |
| 13 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1994 | 1.947462e+08 | 1.985205e+08 | 1.913568e+08 | 7.163637e+06 | 5.846234e+08 | |
| 14 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1994 | 1.032567e+09 | 1.039842e+09 | 1.025631e+09 | 1.421188e+07 | 3.098040e+09 | |
| 20955 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2015 | 2.159385e+06 | 2.329364e+06 | 1.990166e+06 | 3.391983e+05 | 6.478916e+06 | |
| 20956 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2015 | 2.288306e+05 | 3.056884e+05 | 1.694027e+05 | 1.362857e+05 | 7.039218e+05 | |
| 20957 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2015 | 2.388216e+06 | 2.587005e+06 | 2.211144e+06 | 3.758613e+05 | 7.186365e+06 | |
| 20958 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2016 | 2.223791e+06 | 2.402490e+06 | 2.048482e+06 | 3.540084e+05 | 6.674764e+06 | |
| 20959 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2016 | 2.311014e+05 | 3.101747e+05 | 1.700120e+05 | 1.401627e+05 | 7.112880e+05 | |
| 20960 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2016 | 2.454893e+06 | 2.665441e+06 | 2.267696e+06 | 3.977449e+05 | 7.388030e+06 | |
| 20961 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2017 | 2.297622e+06 | 2.490884e+06 | 2.114574e+06 | 3.763097e+05 | 6.903079e+06 | |
| 20962 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2017 | 2.373815e+05 | 3.217514e+05 | 1.729171e+05 | 1.488343e+05 | 7.320499e+05 | |
| 20963 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2017 | 2.535003e+06 | 2.743769e+06 | 2.341329e+06 | 4.024397e+05 | 7.620102e+06 | |
| 20964 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2018 | 2.367072e+06 | 2.575100e+06 | 2.173995e+06 | 4.011055e+05 | 7.116167e+06 | |
| 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 | 7.474673e+05 | |
| 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 | 7.853723e+06 | |
| 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 | 7.332178e+06 | |
| 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 | 7.662870e+05 | |
| 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 | 8.088217e+06 | |

In [58]:
```
#przedstawić na przykładzie dodawanie nowych kolumn z pomocą funkcji
#Lambda
year_id = [2017,2018,2019]
df_all_new = df_all_new.reset_index()
df_all_new['Is2017-2019'] = df_all['year_id'].apply(lambda x: True if x in year_id else False )
df_all_new
```

Out[58]:

| | index | measure_name | location_id | location_name | sex_id | sex | age_group_name | year_id | val | upper | lower | Tolerance_range | smokers | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1990 | 1.891488e+08 | 1.930929e+08 | 1.855595e+08 | 7.533419e+06 | 5.678012e+08 | |
| 1 | 2 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1990 | 9.922503e+08 | 1.000161e+09 | 9.847880e+08 | 1.537321e+07 | 2.977200e+09 | |
| 2 | 3 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1991 | 8.138972e+08 | 8.200339e+08 | 8.069514e+08 | 1.308248e+07 | 2.440883e+09 | |
| 3 | 4 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1991 | 1.905375e+08 | 1.944249e+08 | 1.869744e+08 | 7.450504e+06 | 5.719369e+08 | |
| 4 | 5 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1991 | 1.004435e+09 | 1.011925e+09 | 9.969811e+08 | 1.494378e+07 | 3.013341e+09 | |
| 5 | 6 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1992 | 8.233148e+08 | 8.292228e+08 | 8.167264e+08 | 1.249646e+07 | 2.469264e+09 | |
| 6 | 7 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1992 | 1.919026e+08 | 1.957109e+08 | 1.884066e+08 | 7.304270e+06 | 5.760201e+08 | |
| 7 | 8 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1992 | 1.015217e+09 | 1.022720e+09 | 1.007847e+09 | 1.487316e+07 | 3.045784e+09 | |
| 8 | 9 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1993 | 8.313873e+08 | 8.372931e+08 | 8.249496e+08 | 1.234346e+07 | 2.493630e+09 | |
| 9 | 10 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1993 | 1.932818e+08 | 1.970626e+08 | 1.896392e+08 | 7.223415e+06 | 5.801836e+08 | |
| 10 | 11 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1993 | 1.024669e+09 | 1.031965e+09 | 1.017551e+09 | 1.441378e+07 | 3.074184e+09 | |
| 11 | 12 | Number of Smokers | 1 | Global | 1 | Male | 15+ years | 1994 | 8.378204e+08 | 8.437233e+08 | 8.316340e+08 | 1.208930e+07 | 2.513178e+09 | |
| 12 | 13 | Number of Smokers | 1 | Global | 2 | Female | 15+ years | 1994 | 1.947462e+08 | 1.985205e+08 | 1.913568e+08 | 7.163637e+06 | 5.846234e+08 | |
| 13 | 14 | Number of Smokers | 1 | Global | 3 | Both | 15+ years | 1994 | 1.032567e+09 | 1.039842e+09 | 1.025631e+09 | 1.421188e+07 | 3.098040e+09 | |
| 14 | 20955 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2015 | 2.159385e+06 | 2.329364e+06 | 1.990166e+06 | 3.391983e+05 | 6.478916e+06 | |
| 15 | 20956 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2015 | 2.288306e+05 | 3.056884e+05 | 1.694027e+05 | 1.362857e+05 | 7.039218e+05 | |
| 16 | 20957 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2015 | 2.388216e+06 | 2.587005e+06 | 2.211144e+06 | 3.758613e+05 | 7.186365e+06 | |
| 17 | 20958 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2016 | 2.223791e+06 | 2.402490e+06 | 2.048482e+06 | 3.540084e+05 | 6.674764e+06 | |
| 18 | 20959 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2016 | 2.311014e+05 | 3.101747e+05 | 1.700120e+05 | 1.401627e+05 | 7.112880e+05 | |
| 19 | 20960 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2016 | 2.454893e+06 | 2.665441e+06 | 2.267696e+06 | 3.977449e+05 | 7.388030e+06 | |
| 20 | 20961 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2017 | 2.297622e+06 | 2.490884e+06 | 2.114574e+06 | 3.763097e+05 | 6.903079e+06 | |
| 21 | 20962 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2017 | 2.373815e+05 | 3.217514e+05 | 1.729171e+05 | 1.488343e+05 | 7.320499e+05 | |
| 22 | 20963 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2017 | 2.535003e+06 | 2.743769e+06 | 2.341329e+06 | 4.024397e+05 | 7.620102e+06 | |
| 23 | 20964 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2018 | 2.367072e+06 | 2.575100e+06 | 2.173995e+06 | 4.011055e+05 | 7.116167e+06 | |
| 24 | 20965 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2018 | 2.435999e+05 | 3.286166e+05 | 1.752508e+05 | 1.533658e+05 | 7.474673e+05 | |
| 25 | 20966 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2018 | 2.610672e+06 | 2.833943e+06 | 2.409108e+06 | 4.248345e+05 | 7.853723e+06 | |
| 26 | 20967 | Number of Smokers | 522 | Sudan | 1 | Male | 15+ years | 2019 | 2.439150e+06 | 2.656579e+06 | 2.236450e+06 | 4.201295e+05 | 7.332178e+06 | |
| 27 | 20968 | Number of Smokers | 522 | Sudan | 2 | Female | 15+ years | 2019 | 2.500800e+05 | 3.345384e+05 | 1.816686e+05 | 1.528698e+05 | 7.662870e+05 | |
| 28 | 20969 | Number of Smokers | 522 | Sudan | 3 | Both | 15+ years | 2019 | 2.689230e+06 | 2.918332e+06 | 2.480656e+06 | 4.376758e+05 | 8.088217e+06 | |

In [59]:
```
#przedstawić możliwości pracy z dużymi plikami przy użyciu argumentu
#chunksize

for chunk_df in pd.read_csv('IHME_GBD_2019_SMOKING_TOB_1990_2019_CIG_PC_Y2021M05D27.CSV',
                            chunksize = 50000):
    print("CHUNK DF")
    print(chunk_df.head())
```

```
CHUNK DF
                       measure_name  location_id  \
0  Cigarette-Equivalents Per Capita            1
1  Cigarette-Equivalents Per Capita            1
2  Cigarette-Equivalents Per Capita            4
3  Cigarette-Equivalents Per Capita            4
4  Cigarette-Equivalents Per Capita            5

                         location_name  sex_id sex_name  age_group_id  \
0                               Global       3     Both            29
1                               Global       3     Both            29
2  Southeast Asia, East Asia, and Oceania       3     Both            29
3  Southeast Asia, East Asia, and Oceania       3     Both            29
4                           East Asia       3     Both            29

  age_group_name  year_id          val        upper        lower
0      15+ years     1990  1484.256502  1531.563739  1436.151878
1      15+ years     2019  1113.754663  1161.263946  1069.765828
2      15+ years     1990  1827.374739  1959.359086  1692.900863
3      15+ years     2019  1778.846098  1927.560165  1640.645875
4      15+ years     1990  2089.743485  2267.199999  1908.301510
```

In [60]:
```
new_df = pd.DataFrame() # pusta ramka danych
for chunk_df in pd.read_csv('IHME_GBD_2019_SMOKING_TOB_1990_2019_CIG_PC_Y2021M05D27.CSV',
                            chunksize = 50000):
    result = chunk_df.groupby(['location_name', 'year_id']).agg({'val': 'mean',
                                                                 'upper': 'max'})
    new_df = pd.concat([new_df,result])

new_df
```

Out[60]:

| | | val | upper |
|---|---|---|---|
| location_name | year_id | | |
| Afghanistan | 1990 | 274.126957 | 320.558021 |
| | 2019 | 444.334632 | 546.171500 |
| Albania | 1990 | 1894.040861 | 2224.731864 |
| | 2019 | 1941.384044 | 2305.846372 |
| Algeria | 1990 | 1259.079364 | 1381.657971 |
| ... | ... | ... | ... |
| Yemen | 2019 | 1391.887788 | 1712.648491 |
| Zambia | 1990 | 308.165288 | 343.536927 |
| | 2019 | 296.250416 | 366.554416 |
| Zimbabwe | 1990 | 931.803728 | 1130.361142 |
| | 2019 | 898.367226 | 1132.376771 |

462 rows × 2 columns

In [ ]:

## 3. Wnioski

Biblioteka Pandas w szybki sposób pozwala uzyskać odpowiednie rezultaty analizy dużych zbiorów danych. W połączeniu ze środowiskiem *Jupyter Notebook*, użytkownik w bardzo szybki sposób jest w stanie przeprowadzić kilka operacji na zbiorze danych i w każdym z kroków sprawdzić czy wyniki bieżne są do oczekiwanych. Przy użyciu biblioteki matplotlib.pyplot dane można przedstawić na wielu wariantach wykresów.