Sheeet4:World Bank Group (2022)

The dataset provides World Bank Group estimates for life expectancy across different countries in 2022, focusing only on nations with populations exceeding 50,000 people. It includes comparative data for the years 2014, 2019, and 2022 to analyze trends, particularly noting that 2014 marked a local peak in life expectancy for some leading countries. Data rounding may cause minor inconsistencies (up to 0.01 year) when compared to the raw calculations. The main columns include the country name, life expectancy for the three reference years, and the population size in 2022.

```python
import pandas as pd
sh4=pd.read_excel(r"C:\Users\sh138\OneDrive\Documents\sh4.xlsx",header=1)
sh4.columns= ['Countries and territories', 'All', 'Male', 'Female', 'Sex gap','2014', '2014:2019', '2019','2019:2020', '2020', '2020:2021'
```

First, I read data then I use 'header' to handle the index and the columns.
Secondly, I gave each column its proper name.

```python
sh4.head()
```

	Countries and territories	All	Male	Female	Sex gap	2014	2014:2019	2019	2019:2020	2020	2020:2021	2021	2021:2022	2022	recovery from COVID-19: 2019:2022	
0	Macao SAR, China	85.38	82.75	87.98	5.23	83.88	1.09	84.98	0.21	85.18	0.22	85.40	-0.02	85.38	0.40	NaN
1	Japan	84.00	81.05	87.09	6.04	83.59	0.77	84.36	0.20	84.56	-0.11	84.45	-0.45	84.00	-0.36	NaN
2	Hong Kong SAR, China	83.66	80.70	86.77	6.07	83.94	1.22	85.16	0.34	85.50	0.04	85.53	-1.87	83.66	-1.49	NaN
3	French Polynesia	83.55	81.33	85.98	4.65	82.21	0.78	83.00	-0.78	82.22	-2.73	79.49	4.06	83.55	0.55	NaN
4	Switzerland	83.45	81.60	85.40	3.80	83.20	0.71	83.90	-0.90	83.00	0.75	83.75	-0.30	83.45	-0.45	NaN

```python
sh4.drop(' ', axis=1, inplace=True)
```

We have column unknown completely, I dropped it.

```python
dtype: object
cols_to_convert = ['All', 'Male', 'Female', 'Sex gap', '2014', '2014:2019', '2019', '2019:2020', '2020', '2020:2021', '2021:2022', '2022', 'recovery from COVID-19: 2019:2022']

for col in cols_to_convert:
    sh4[col] = (sh4[col].astype(str)
                .str.replace('-', '-')
                .str.replace(',', '.')
                .str.replace(r'[^\d.-]', '', regex=True))
    sh4[col] = pd.to_numeric(sh4[col], errors='coerce')
```

```python
sh4['All'] = sh4['All'].astype(float)
sh4['Male'] = sh4['Male'].astype(float)
sh4['Female'] = sh4['Female'].astype(float)
sh4['Sex gap'] = sh4['Sex gap'].astype(float)
sh4['2014'] = sh4['2014'].astype(float)
sh4['2014:2019'] = sh4['2014:2019'].astype(float)
sh4['2019'] = sh4['2019'].astype(float)
sh4['2019:2020'] = sh4['2019:2020'].astype(float)
sh4['2020'] = sh4['2020'].astype(float)
sh4['2020:2021'] = sh4['2020:2021'].astype(float)
sh4['2021:2022'] = sh4['2021:2022'].astype(float)
sh4['2022'] = sh4['2022'].astype(float)
sh4['recovery from COVID-19: 2019:2022'] = sh4['recovery from COVID-19: 2019:2022'].astype(float)
```

There was some unknown symbols, I handled them then convert the data types to be appropriate.

```
columns_to_fill = [
    "All", "Male", "Female", "Sex gap",
    "2014", "2014:2019", "2019", "2019:2020",
    "2020", "2020:2021", "2021:2022", "2022",
    "recovery from COVID-19: 2019:2022"
]

for col in columns_to_fill:
    sh4[col] = sh4[col].fillna(sh4[col].mean())
```

Handling nulls values.

Sheet1:UN: Change of life expectancy from 2019 to 2023
 This dataset presents the change in life expectancy across different
countries between 2019 and 2023,based on United Nations estimates,
highlighting trends of increase or decrease during this period.

Sheet2:World Health Organization (2019)

 This dataset, published by the World Health Organization in December 2020,
provides life expectancy at birth for all populations as of 2019. If countries
have equal life expectancy values, they are further sorted by HALE (Healthy
Life Expectancy) for the total population.

Sheet3:CIA World Factbook (2022)

 This dataset, published by the CIA World Factbook in 2022, provides life
expectancy data categorized by total population, male, female, and the gender
gap between males and females.

Sheet5:UN: Estimate of life expectancy for various ages in 2023

 This dataset presents the United Nations' 2023 estimates of life expectancy
at various ages for different countries, with location links directing to
detailed demographic profiles for each country.

Sheet6:OECD(2022)

This dataset provides life expectancy estimates for OECD countries in 2022, with the default sorting based on life expectancy values from 2019.

Name: Salma Hossam Eldeen Abdelbaset

ID:23011284