

```

!pip install wbdata duckdb

import wbdata
import duckdb
import pandas as pd
import datetime
countries = ['AF', 'GB', 'IN', 'KE', 'BR']

indicators = {
    'SP.DYN.LE00.IN': 'life_expectancy',
    'SH.DYN.MORT': 'under5_mortality',
    'SP.ADO.TFRT': 'adolescent_fertility',
    'SE.SEC.ENRR': 'secondary_school_enrollment'
}

df = wbdata.get_dataframe(indicators, country=countries)
df = df.reset_index()

df['year'] = pd.to_datetime(df['date']).dt.year

df = df[(df['year'] >= 2000) & (df['year'] <= 2023)]

```

Requirement already satisfied: wbdata in /opt/anaconda3/lib/python3.12/site-packages (1.0.0)

Requirement already satisfied: duckdb in /opt/anaconda3/lib/python3.12/site-packages (1.3.1)

Requirement already satisfied: appdirs<2.0,>=1.4 in /opt/anaconda3/lib/python3.12/site-packages (1.4.4)

Requirement already satisfied: backoff<3.0.0,>=2.2.1 in /opt/anaconda3/lib/python3.12/site-packages (2.2.1)

Requirement already satisfied: cachetools<6.0.0,>=5.3.2 in /opt/anaconda3/lib/python3.12/site-packages (5.5.2)

Requirement already satisfied: dateparser<2.0.0,>=1.2.0 in /opt/anaconda3/lib/python3.12/site-packages (1.2.0)

Requirement already satisfied: decorator<6.0.0,>=5.1.1 in /opt/anaconda3/lib/python3.12/site-packages (5.1.1)

Requirement already satisfied: requests<3.0,>=2.0 in /opt/anaconda3/lib/python3.12/site-packages (2.32.0)

Requirement already satisfied: shelved-cache<0.4.0,>=0.3.1 in /opt/anaconda3/lib/python3.12/site-packages (0.3.1)

Requirement already satisfied: tabulate<0.9.0,>=0.8.5 in /opt/anaconda3/lib/python3.12/site-packages (0.8.10)

Requirement already satisfied: python-dateutil<3.0.0,>=2.7.0 in /opt/anaconda3/lib/python3.12/site-packages (2.9.0.post0)

Requirement already satisfied: pytz>=2024.2 in /opt/anaconda3/lib/python3.12/site-packages (2024.2)

Requirement already satisfied: regex!=2019.02.19,!=2021.8.27,>=2015.06.24 in /opt/anaconda3/lib/python3.12/site-packages (2024.11.6)

Requirement already satisfied: tzlocal>=0.2 in /opt/anaconda3/lib/python3.12/site-packages (5.2)

Requirement already satisfied: charset-normalizer<4,>=2 in /opt/anaconda3/lib/python3.12/site-packages (3.4.0)

Requirement already satisfied: idna<4,>=2.5 in /opt/anaconda3/lib/python3.12/site-packages (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/anaconda3/lib/python3.12/site-packages (2.2.3)

Requirement already satisfied: certifi>=2017.4.17 in /opt/anaconda3/lib/python3.12/site-packages (2024.12.14)

Requirement already satisfied: six>=1.5 in /opt/anaconda3/lib/python3.12/site-packages (from requests) (1.17.0)

```

duckdb.sql("CREATE OR REPLACE VIEW wdi_raw AS SELECT * FROM df")

df_clean = duckdb.sql("""
SELECT
    country,
    year,
    MAX(life_expectancy) AS life_expectancy,
    MAX(under5_mortality) AS under5_mortality,
    MAX(adolescent_fertility) AS adolescent_fertility,
    MAX(secondary_school_enrollment) AS secondary_school_enrollment
FROM wdi_raw
GROUP BY country, year
ORDER BY country, year
""").df()

df_clean

```

	country	year	life_expectancy	under5_mortality	adolescent_fertility	secondary_school_enrollment
0	Afghanistan	2000	55.005000	131.7	154.310	NaN
1	Afghanistan	2001	55.511000	127.4	152.471	14.471510
2	Afghanistan	2002	56.225000	123.1	150.161	NaN
3	Afghanistan	2003	57.171000	118.7	143.750	14.078050
4	Afghanistan	2004	57.810000	114.2	136.930	19.263140
...
115	United Kingdom	2019	81.369122	4.4	11.064	117.233337
116	United Kingdom	2020	80.331756	4.4	9.701	114.651649
117	United Kingdom	2021	80.647317	4.4	8.578	113.683823
118	United Kingdom	2022	81.011220	4.5	8.379	113.228363
119	United Kingdom	2023	81.238098	4.5	8.359	NaN

```
df_clean.to_csv("../data/population_dynamics_clean.csv", index=False)
```