“Data-Driven Insights on Global Sustainability: Emissions, Water, Land and Biodiversity”

Course: Data Warehouse & Business Intelligence

Study Program: Business Consulting Master

Group: 01

Subhashri Ravichandran & Team

Prof. Dr. Monika Frey-Luxemburger

Agenda

1. Introduction
2. Business Case
3. Tools
4. Quality Assurance
5. Design & Implementation
6. Conclusions and Lessons Learned

1. Introduction

Scope

- Develop a DW & BI system to centralize and analyze critical global sustainability data—specifically on emissions, water resources, and biodiversity.
- This platform will uncover patterns and insights to support informed decision-making on environmental challenges.

Objectives

- **Identify Global Trends & Gaps**
  (Analyze trends in CO_2 emissions, water access, land and biodiversity threats.)
- **Empower Stakeholders with Insights**
  (Present findings in user-friendly dashboards and reports, allowing government agencies, NGOs, and policymakers to prioritize high-impact initiatives and track progress effectively.)

GANTT CHART

PROJECT TITLE	Data-Driven Insights on Global Sustainability: Emissions,Water, and Biodiversity.	**COMPANY NAME**	SKiMS
PROJECT MANAGER	Ina Shtëmbari	**DATE**	02/10/24

PHASE ONE: WEEK 1, WEEK 2, WEEK 3, WEEK 4 — PHASE TWO: WEEK 5, WEEK 6, WEEK 7, WEEK 8 — PHASE THREE: WEEK 9, WEEK 10, WEEK 11, WEEK 12 — PHASE FOUR: WEEK 13, WEEK 14, WEEK 15 (days M T W R F)

WBS NUMBER	TASK TITLE	START DATE	DUE DATE	DURATION HOURS	% OF TASK COMPLETE
1	**Planning and Conception**			**69**	
1.1	Project Definiton & Business Case	14/10/24	19/10/24	17	100%
1.2	Project Planning & Role Assignment	20/10/24	26/10/24	32	100%
1.3	Risk management	26/10/24	02/11/24	8	100%
1.4	Initial Presentation + Preparation	28/10/24	30/10/24	12	100%
2	**Data Preparation and Modeling**			**155**	
2.1	Tool Selection and Evaluation	28/10/24	02/11/24	12	100%
2.2	KPIs Definition	29/10/24	09/11/24	28	100%
2.3	ETL Process	04/11/24	17/11/24	36	100%
2.4	Data Warehouse Design	04/11/24	17/11/24	36	100%
2.5	Schema Normalization & Refinement	04/11/24	17/11/24	14	100%
2.6	Excel Prototype	18/11/24	28/11/24	17	100%
2.7	Intermediate Presentation + Preparation	27/11/24	04/12/24	12	80%
3	**BI Report Development**			**148**	
3.1	BI Dashboard Development	18/11/24	18/12/24	50	50%
3.2	Testing	25/11/24	07/12/24	58	50%
3.3	Predictive Analysis	04/12/24	20/12/24	40	0%
4	**Project Analysis and Delivery**			**144**	
4.1	Data Analysis and Insight Generation	18/11/24	15/01/25	69	68%
4.2	Quality Assurance	14/10/24	15/01/25	54	75%
4.3	Final Presentation + Preparation	08/01/25	15/01/25	21	0%

2. Business Case

Problem Statement:

Environmental data is often fragmented and dispersed, making it difficult for stakeholders to extract actionable insights for sustainability.

Motivation:

- Address urgent issues like emissions, water & sanitation, threatened species and land use.
- Provide a centralized platform for data-driven decisions.

Challenges:

- Dispersed data sources
- Limited tools for actionable insights

Solution:

Develop a centralized Data Warehouse (DW) with BI tools to:

- Consolidate data for easy access.
- Enable actionable insights for policies, resource allocation, and strategies.

2. Business Case
Stakeholders Benefiting

STAKEHOLDER	DESCRIPTION	KEY BENEFITS
Government Agencies	Entities responsible for policy creation and environmental regulations at local, regional, or national levels.	• Comprehensive data to develop targeted environmental policies. • Monitor progress in sustainability initiatives.
NGOs	Non-profits working to address environmental challenges, resource allocation, and sustainability efforts.	• Prioritize resources for high-impact areas • Measure effectiveness of programs using clear insights.
Researchers	Academic and independent professionals studying environmental issues to provide actionable insights.	• Centralized access to reliable data for academic and field studies. • Identify trends in emissions, water, and biodiversity.
Environmental Strategists	Consultants and planners developing sustainability strategies for businesses and governments.	• Data-driven strategies for corporate/governmental sustainability goals. • Visual tools for better decision-making.

2. Business Case
Dataset Overview

Dataset Selection Process

Initial Research Phase
(Explored multiple open data sources such as: UN Data, Maveen Analytics, Kaggle, and others.)

Criteria Development:
- ***Data Relevance***
  (global sustainability issues)
- ***Coverage***
  (broad geographic)
- ***Reliability***
  (data sets from credible sources)

Dataset Evaluation
(How well each dataset would integrate our DW for developing BI insights.)

Final Decision:
Environment Dataset (UNData)

2. Business Case
Dataset Overview

Dataset Link & Summary: Data from the United Nations Statistical Yearbook 2023, providing a reliable foundation for environmental analysis. https://data.un.org/

1. **Emissions:** Data on CO_2 emissions by country, revealing trends and regional insights.
2. **Water & Sanitation:** Access levels across regions, identifying disparities.
3. **Threatened Species:** Information on endangered species by region, indicating biodiversity risks.
4. **Land Use:** Land distribution metrics, highlighting environmental impact.

3. Tools

Tool Overview

TOOLS

Data Warehouse (DW):

MySQL for querying and managing data within the DW, facilitating structured analysis.

Business Intelligence (BI):

Power BI for creating interactive visualizations and dashboards to communicate insights effectively.

CONNECTORS

Excel (CSV Files):

Data sources stored in Excel; used as early analysis tool to validate and refine data before creating dashboards.

Python:

Python used for ensuring the functionality and accuracy of Power BI dashboards.

3. Tools
Evaluation and Benefits

TOOLS

KEY CONTRIBUTIONS

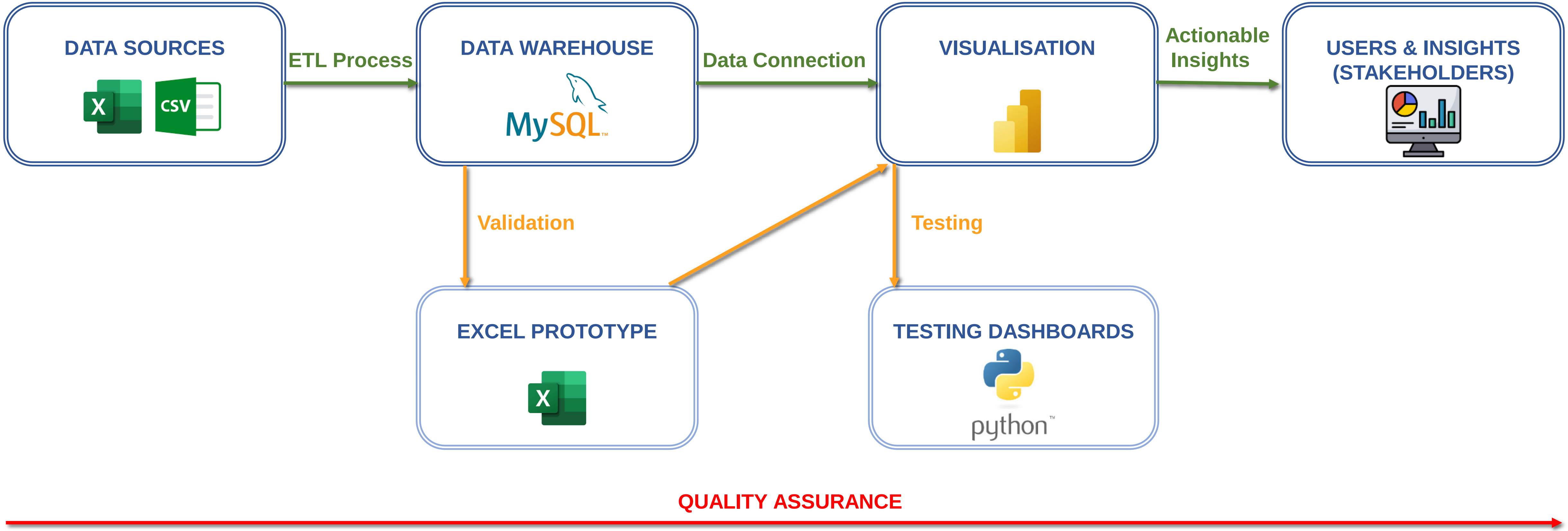
MySQL	Power BI	Excel / CSV	python
Scalability Handles large datasets efficiently. **Performance** Centralizes and normalizes data for accuracy analysis. **Fast** Enables quick querying for insights.	**Ease of Visualization** Transforms complex data into clear, actionable dashboards and reports. **User-Friendly** Easy for stakeholders to explore data. **Integration** Connects seamlessly with MySQL.	**Early Validation** Tests KPIs and data before dashboard creation. **User Accessibility** Familiar tool for stakeholders to explore data. **Cost-Effective** Minimizes errors early in the process, saving time and resources.	**Automation** Tests and validates Power BI dashboards for accuracy and consistency. **Advanced Analytics** Allows deeper insights and anomaly detection beyond Power BI's capabilities. **Integration** Enhances the overall workflow accuracy.

OVERALL IMPACT

- **Efficiency:** Streamlined processes reduce manual errors.
- **Scalability:** Tools adapt to large datasets and future needs.
- **Accuracy:** Early validation and automation ensure reliable results.
- **Actionable Insights:** Stakeholders benefit from intuitive and interactive outputs.

3. Tools

Workflow for Data Integration

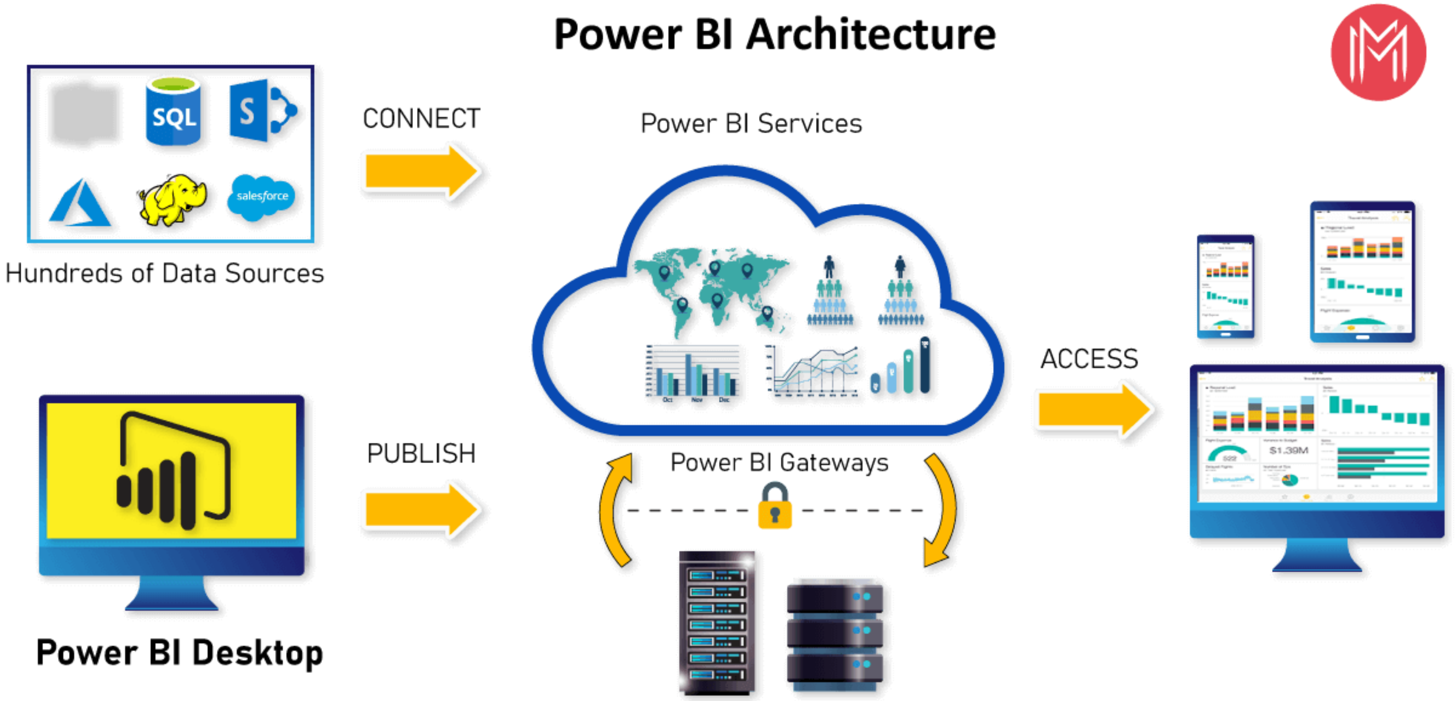


Power BI Key Components:

1. Data Sources
2. Power BI Desktop
3. Power BI Service
4. Power BI Mobile Apps
5. Power BI Gateway
6. Power BI Embedded

How Does Power BI Architecture Work?

1. Connect data sources
2. Transform and clean data
3. Design reports and visuals
4. Publish to Power BI Service
5. Share Dashboards and Reports
6. Usage on mobile
7. Embed into Apps



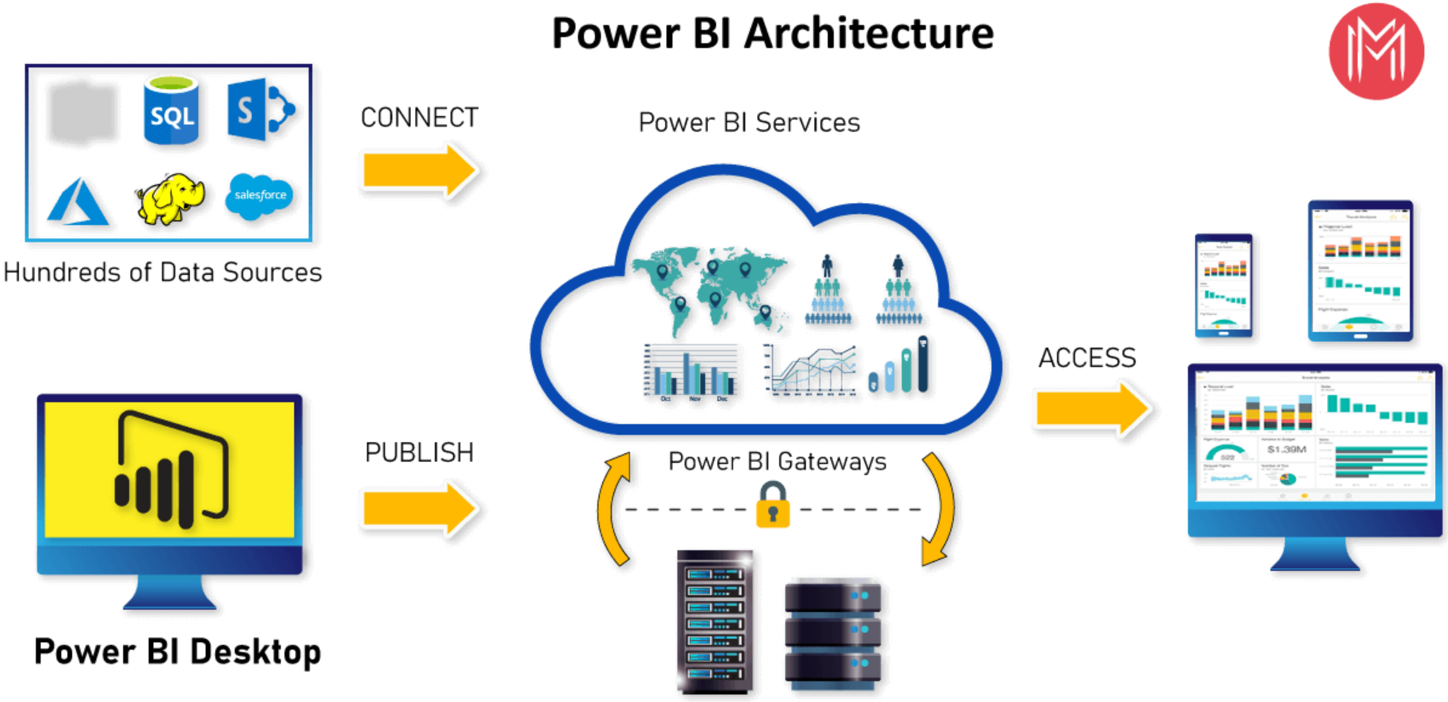
3. Tools
BIA Tool Architecture

Benefits of Using Power BI:

- Self-Service analytics
- Unified view of data
- Interactive dashboards
- Mobile analytics
- Real-time analytics
- Collaboration
- Embedded analytics

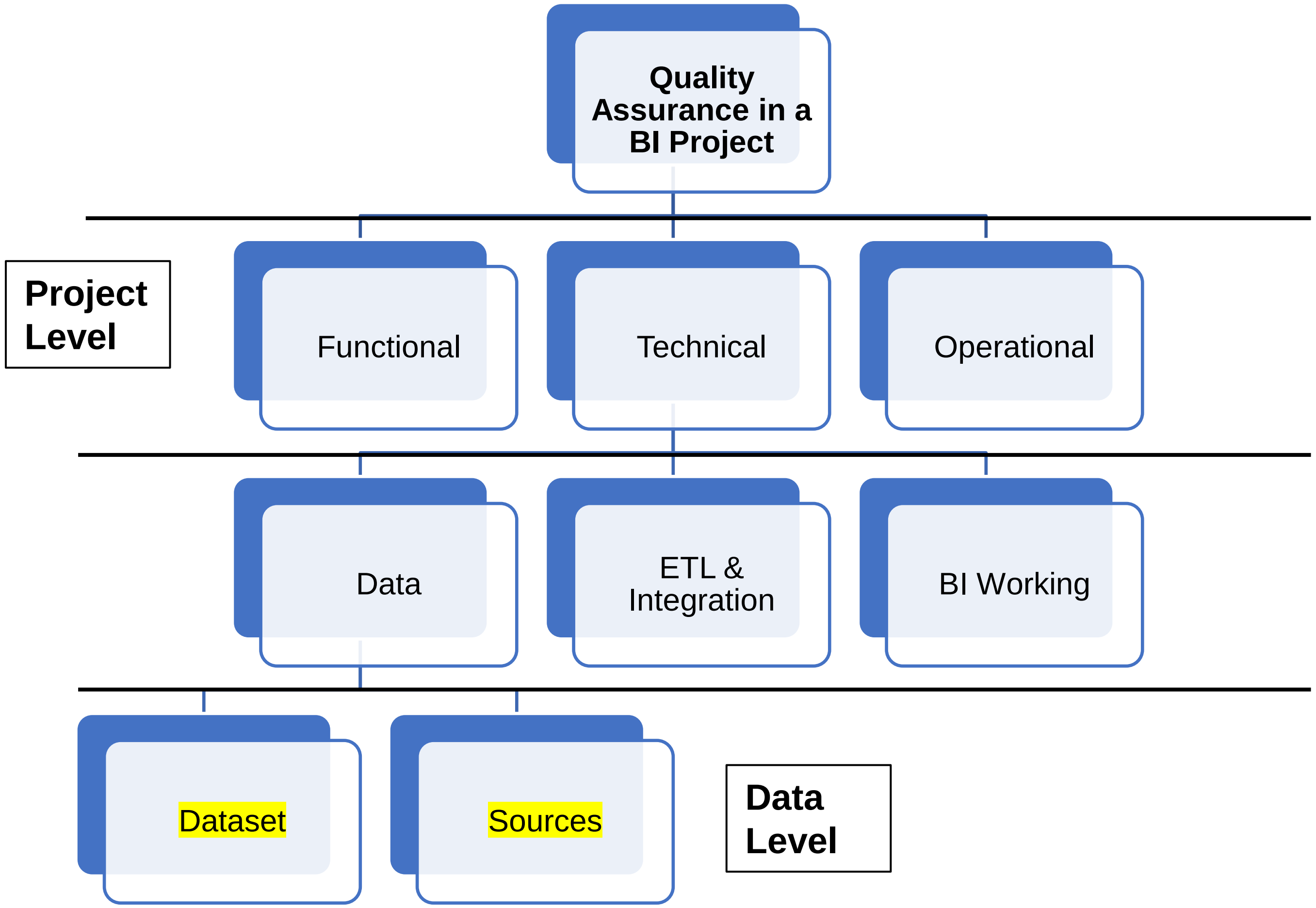
Use Cases Where Power BI Delivers Value:

- Sales analytics
- Marketing analytics
- Operations analytics
- Financial planning
- Remote monitoring
- Customer Analytics
- Fields workforce analytics
- Healthcare analytica



4. Quality Assurance

Overview



Risk Management:

All levels entail risks that can derail a project from meeting its objectives.

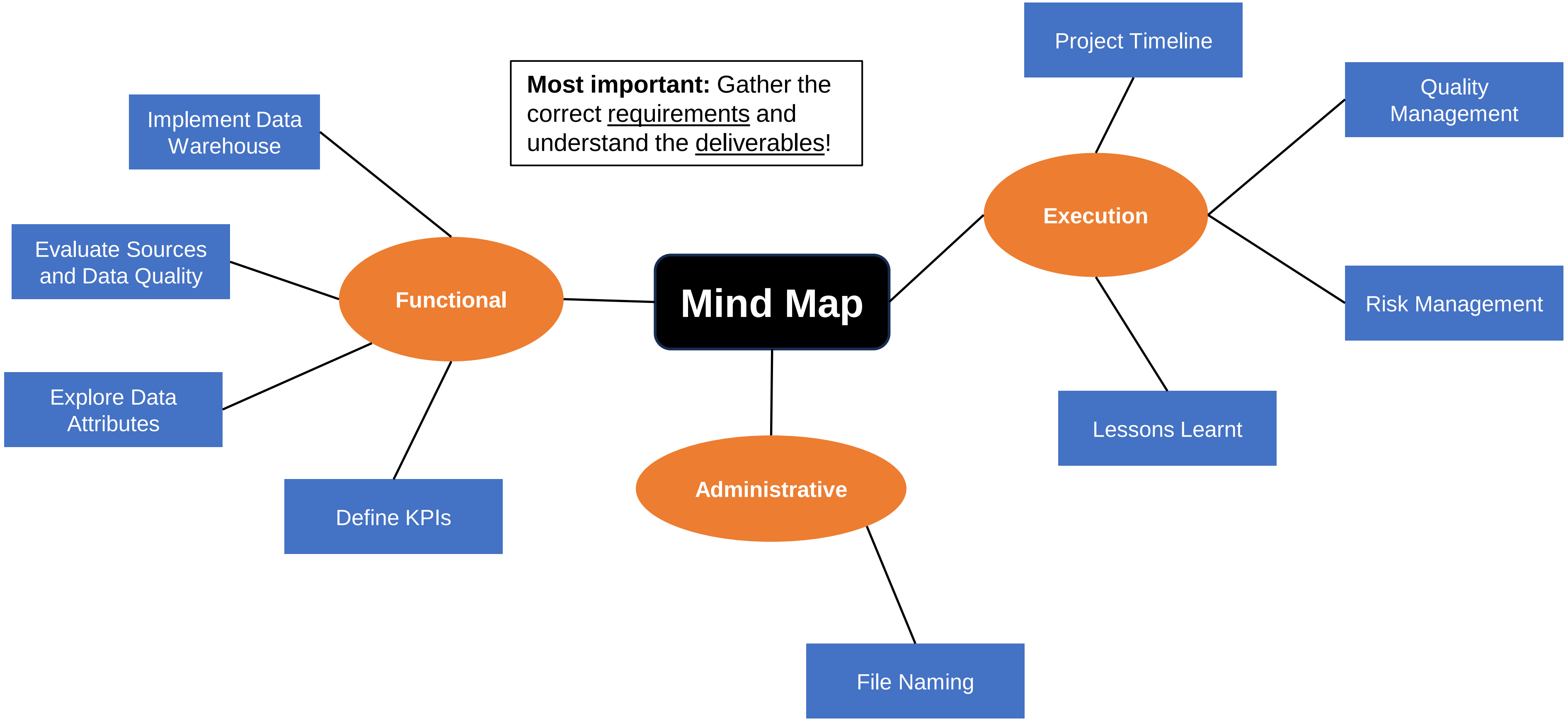
Our strategy:

- Four-eyes principle and backup resource for every task.
- Assuring quality during all steps and at all levels

4. Quality Assurance

Project Quality

Most important: Gather the correct requirements and understand the deliverables!



SKiMS

Source: https://data.un.org/Default.aspx

4. Quality Assurance

Data Quality

HOCHSCHULE FURTWANGEN UNIVERSITY HFU

Datasets

Co2.csv

	A	B	C	D	E	F	G
1	Region/Country/Area	Land	Year	Series	Value	Footnotes	Source
2	8	Albania	1975	Emissions	4,524		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
3	8	Albania	1985	Emissions	7,145		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
4	8	Albania	2005	Emissions	3,980		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
5	8	Albania	2010	Emissions	4,074		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
6	8	Albania	2015	Emissions	3,975		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
7	8	Albania	2018	Emissions	4,525		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a

Land.csv

	A	B	C	D	E	F	G
1	Region/Country/Area	Land	Year	Series	Value	Footnotes	Source
2	1	Total, all c	2005	Land area	########	Estimate.	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.
3	1	Total, all c	2005	Arable lan	13,65,372	Estimate.	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.
4	1	Total, all c	2005	Permanen	1,44,872	Estimate.	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.
5	1	Total, all c	2005	Forest cov	41,32,183	Estimate.	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.
6	1	Total, all c	2005	Arable lan	10.5	Estimate.	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.
7	1	Total, all c	2005	Permanen	1.1	Estimate.	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.

Species.csv

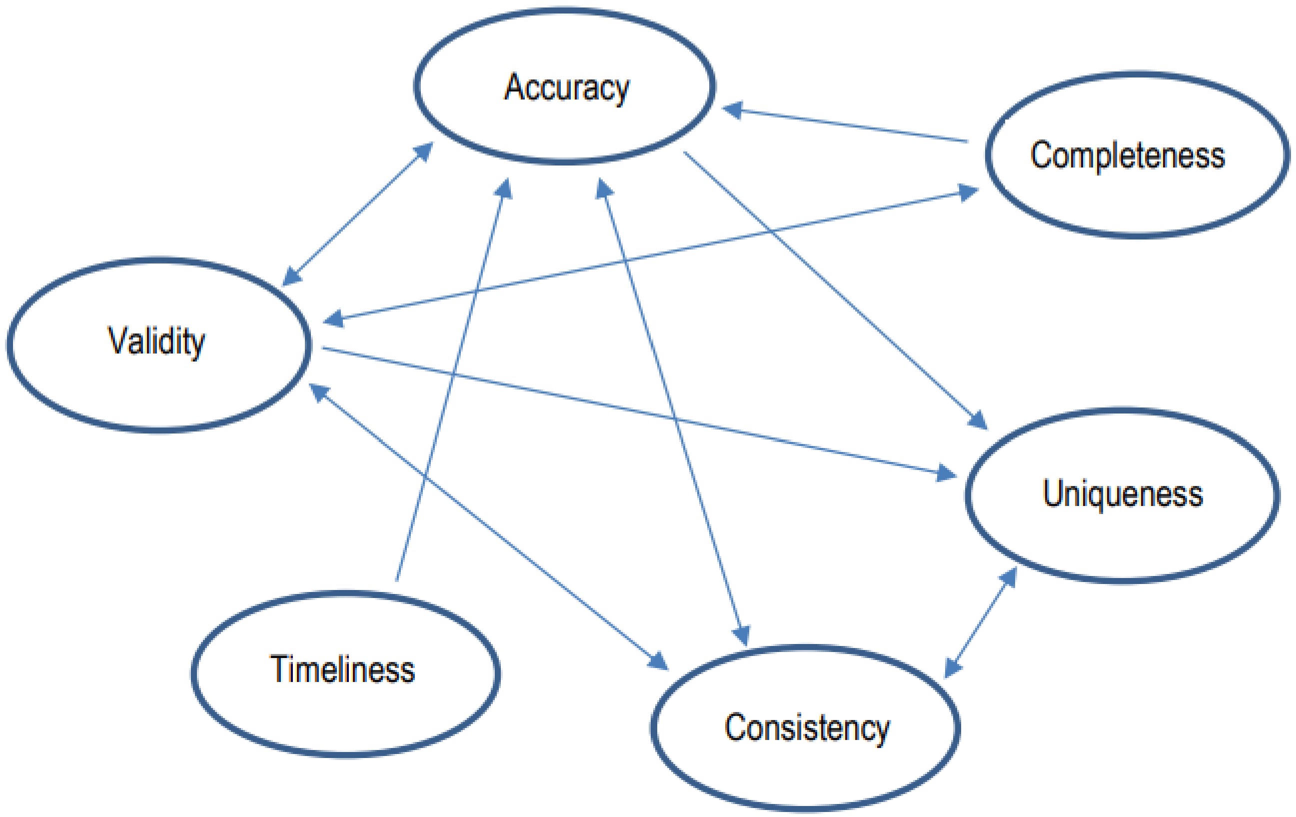
	A	B	C	D	E	F	G
1	Region/Country/Area	Land	Year	Series	Value	Footnotes	Source
2	4	Afghanistan	2004	Threatene	31		World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication,
3	4	Afghanistan	2010	Threatene	31		World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication,
4	4	Afghanistan	2015	Threatene	31		World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication,
5	4	Afghanistan	2019	Threatene	33		World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication,
6	4	Afghanistan	2020	Threatene	33		World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication,
7	4	Afghanistan	2021	Threatene	38		World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication,

Water & Sanitation.cs

	A	B	C	D	E	F	G
1	Region/Country/Area		Year	Series	Value	Footnotes	Source
2	1	Total, all countries or areas	2010	Safely mar	79.7		World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, th
3	1	Total, all countries or areas	2010	Safely mar	51.3		World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, th
4	1	Total, all countries or areas	2010	Safely mar	65.9		World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, th
5	1	Total, all countries or areas	2010	Safely mar	41.8		World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, th
6	1	Total, all countries or areas	2010	Safely mar	28.4		World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, th
7	1	Total, all countries or areas	2010	Safely mar	42.2		World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, th

4. Quality Assurance

Data Quality : Dimensions



Source: https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/#UN-NQAF-Manual

4. Quality Assurance

Data Quality : Approach

Progress so far:

1. No duplicate records found.
2. No inaccuracies found (so far).
3. Formatting has been standardized.
4. Irrelevant data have been identified. Yet to decide if we are to keep or drop.
5. Decided not to evaluate outliers (if any).

Tools used for data cleaning:

1. Microsoft Excel
2. Structured Query Language in the MySQL database management system.

Pending problems

Description

(A) Issues within individual tables

(B) Assembling a global table

1. Missing years (rows) for one or more regions and/or one or more studies (series).
2. Missing regions (rows) for one or more studies (series).

Source (used as guide for data cleaning): Everything About Data Exploration and Cleaning | by Krishna yogi | Medium

4. Quality Assurance

Data Quality : Remediation

*for pending problems

Yet to be decided

4. Quality Assurance

Source Quality

Hierarchy

Source: https://data.un.org/Default.aspx (scroll down and click on environment).

- **Primary source**
  1. United Nations, Department of Economic and Social Affairs, Statistics Division (last accessed October).

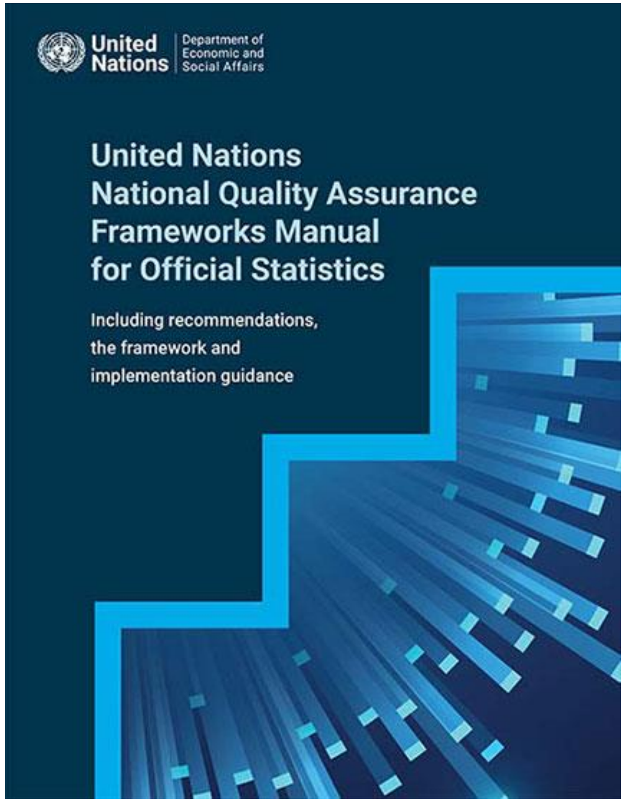
- **Secondary sources**
  2. Food and Agriculture Organization of the United Nations (FAO), Rome (last accessed December 2022).
  3. International Energy Agency, IEA World Energy Balances 2021.
  4. IPCC Guidelines for Greenhouse Gas Inventories 2006 (last accessed April 2023).
  5. WHO/UNICEF Joint Monitoring Program (last accessed July 2023).
  6. IUCN Red List of Threatened Species publication (last accessed May 2023).

- **Tertiary sources**
  7. Government(s), departments of statistics, and other agencies.

- **Quaternary sources**
  8. Data producers

4. Quality Assurance

Source Quality : Evaluation Criteria

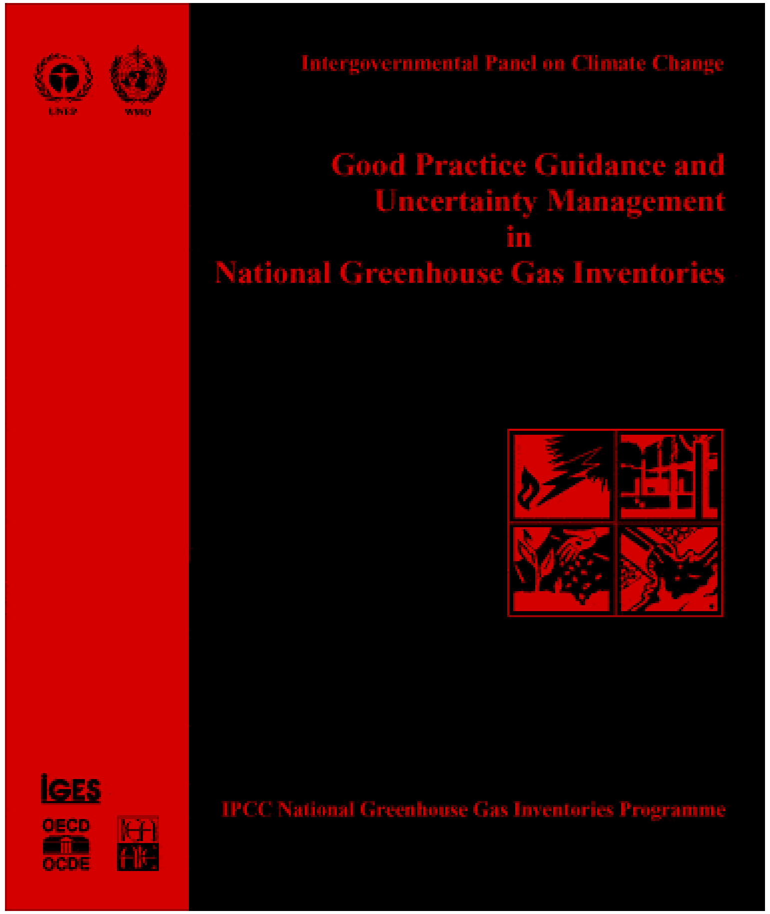


AIM	CATEGORIES	WHAT DO WE ACTUALLY NEED TO KNOW?	DIMENSIONS OF QUALITY
PROCESS QUALITY		Can I see how the data was obtained?	Appropriate sources Metadata and documentation
		Does the data handing process inimise/ check for data handling e arrors? Do the methods reflect best practice?	Sound methods and systems Ethical data practice Sustainability/cost-efficiency
		Are there processes to keep up to date with best practice?	Review procedures
INSTITUTIONAL QUALITY		Is the organization impartial and objective?	Impartiality and objectivity
		Is the organization transparent and credible?	Review procedures Data quality framework Statistical coordination

Source: https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/#UN-NQAF-Manual

4. Quality Assurance

Source Quality : Data Collection Methods



Contents

CO_2 EMISSIONS

The *IPCC Guidelines* provide two methods to estimate CO_2 emissions from Road Transport. The Tier 1, or 'top down' approach calculates CO_2 emissions by estimating fuel consumption in a common energy unit, multiplying by an emission factor to compute carbon content, computing the carbon stored, correcting for unoxidised carbon and finally converting oxidised carbon to CO_2 emissions. The approach is shown in Equation 2.4.

EQUATION 2.4

$$\text{Emissions} = \sum_j [(\text{Emission Factor}_j \bullet \text{Fuel consumed}_j) - \text{Carbon Stored}] \bullet \text{Fraction Oxidised}_j \bullet 44/12$$

Where j = fuel type

Alternatively, a Tier 2, or 'bottom-up' approach estimates emissions in two steps. The first step (Equation 2.5) is to estimate fuel consumed by vehicle type i and fuel type j.

EQUATION 2.5

$$\text{Fuel Consumption}_{ij} = n_{ij} \bullet k_{ij} \bullet e_{ij}$$

Where:

i = vehicle type

j = fuel type

n = number of vehicles

k = annual kilometres travelled per vehicle

e = average litres consumed per kilometre travelled

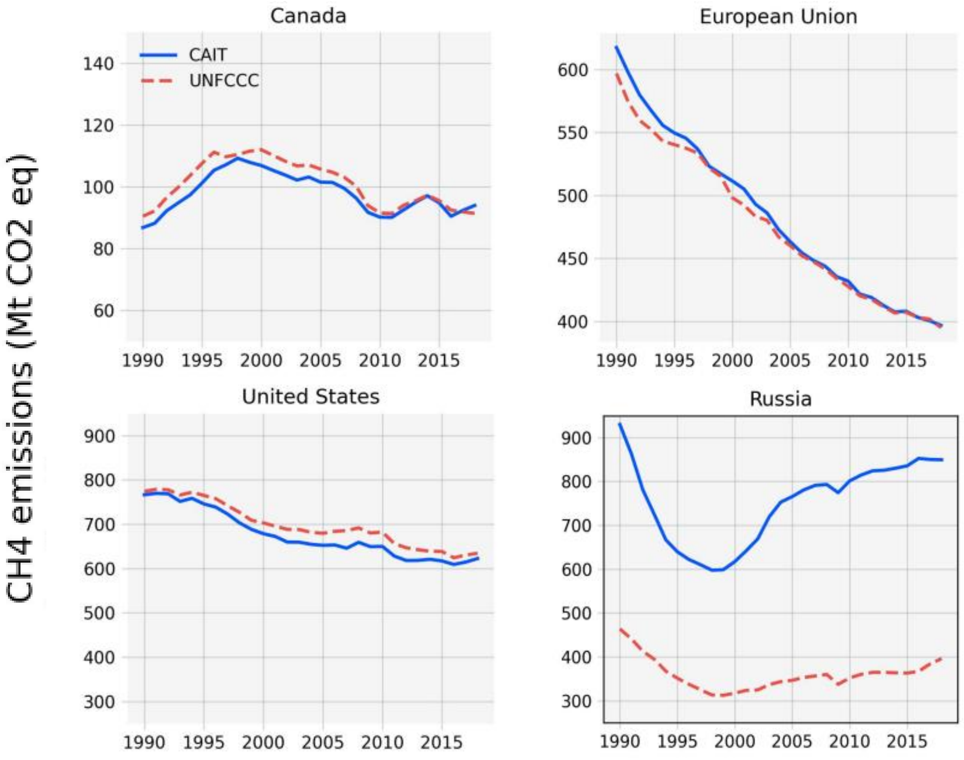
The second step is to estimate total CO_2 emissions by multiplying fuel consumption by an appropriate emission factor for the fuel type and vehicle type (Equation 2.6).

EQUATION 2.6

$$\text{Emissions} = \sum_i \sum_j (\text{Emission Factor}_{ij} \bullet \text{Fuel Consumption}_{ij})$$

Source: https://www.ipcc-nggip.iges.or.jp/public/gp/english/

Comparison of UNFCCC and CAIT Methane Emission Data



Data Source: Climate Watch Historical GHG Emissions
Access at https://bit.io/bitdotio/cop26_methane/

Can you spot something?

When "Reliable" Sources Publish Unreliable Data

Me:

Content Source: Deepnote.com/When Reliable Sources Publish Unreliable Data

4. Quality Assurance

Source Quality : Remediation

Re-mediation Strategies

- ✓ Compare different data sources.
- ✓ Search Google, relevant literature, and data source websites for any caveats or disclaimers about data use.
- ✓ Make sure we know exactly where the data came from: **the data producer may not be the same as the data publisher**.
- ✓ Clearly cite all data sources used.
- ✓ Archive a copy of any data used in case the original data source is modified or deleted.

Source: Deepnote.com/When Reliable Sources Publish Unreliable Data

4. Quality Assurance
Quality Assurance: Implementation

UNdata – A world of information

	A	B	C	D	E	F	G
1	Region/Country/Area	Land	Year	Series	Value	Footnotes	Source
2	8	Albania	1975	Emissions	4,524		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
3	8	Albania	1985	Emissions	7,145		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
4	8	Albania	2005	Emissions	3,980		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
5	8	Albania	2010	Emissions	4,074		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
6	8	Albania	2015	Emissions	3,975		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
7	8	Albania	2018	Emissions	4,525		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a

European Commission

(EDGAR)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	**Substance**	**EDGAR Country C**	**Country**	**1970**	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**1978**	**1979**	**1980**
2	CO2	ABW	Aruba	0.02521379	0.02882775	0.03947211	0.04428944	0.04346915	0.05739627	0.05642291	0.06709976	0.07193698	0.07591904	0.07977223
3	CO2	AFG	Afghanistan	1.73392027	1.73370991	1.69358448	1.73390479	2.19031807	2.02896685	1.89264158	2.28257446	1.93410676	2.05930222	2.00642298
4	CO2	AGO	Angola	8.93389904	8.5195129	10.3661043	11.3469957	11.8065608	10.904653	7.291981	12.0334732	14.2162696	14.1794043	14.3161297
5	CO2	AIA	Anguilla	0.00217759	0.00217769	0.00227319	0.00211848	0.00235984	0.00259365	0.00244415	0.00254677	0.00291135	0.00322276	0.00442236
6	CO2	AIR	International Aviation	169.900399	169.900399	179.759531	187.494406	180.478129	174.582471	174.907983	190.798505	197.866282	204.56754	203.458273
7	CO2	ALB	Albania	4.84498807	4.83812877	5.51829137	4.95194249	5.32961736	5.42411763	5.79657281	6.20093306	6.82099533	7.93670304	8.07349908
8	CO2	ANT	Curaçao	14.5157827	14.5157988	14.1139595	15.608674	14.124013	10.1967506	10.8491463	10.9831957	9.27181813	9.18243661	8.67775674
9	CO2	ARE	United Arab Emirates	19.4647816	25.4298124	27.7473592	35.8436021	36.0106836	36.6794437	45.580994	45.8764465	52.6113891	37.8795052	41.1384805

Our World in Data

(Oxford University)

Afghanista	2012	AFG	30560036	5.92E+10	0.029	0.001	10.035	-1.803	-15.23	10.085	-1.846	-15.472	0.33	0.17	0.25	0.328	0.17	0.249
Afghanista	2013	AFG	31622708	6.3E+10	0.036	0.001	9.229	-0.807	-8.04	9.21	-0.875	-8.679	0.291	0.146	0.275	0.292	0.146	0.276
Afghanista	2014	AFG	32792527	6.43E+10	0.029	0.001	9.086	-0.142	-1.544	9.041	-0.169	-1.837	0.276	0.141	0.318	0.277	0.141	0.32
Afghanista	2015	AFG	33831765	6.28E+10	0.041	0.001	9.67	0.584	6.429	9.803	0.763	8.437	0.29	0.156	0.299	0.286	0.154	0.295
Afghanista	2016	AFG	34700614	6.44E+10	0.076	0.002	8.906	-0.764	-7.899	9.108	-0.696	-7.095	0.262	0.141	0.305	0.257	0.138	0.298
Afghanista	2017	AFG	35688942	6.52E+10	0.045	0.001	9.677	0.771	8.657	14.573	5.465	60.005	0.408	0.224	0.493	0.271	0.148	0.327
Afghanista	2018	AFG	36743040	6.6E+10	0.057	0.002	10.602	0.924	9.553	13.308	-1.265	-8.682	0.362	0.202	0.392	0.289	0.161	0.313
Afghanista	2019	AFG	37856126	7.31E+10	0.038	0.001	10.825	0.223	2.106	19.743	6.436	48.361	0.522	0.27	0.651	0.286	0.148	0.357
Afghanista	2020	AFG	39068978	7.14E+10	0.061	0.002	11.606	0.781	7.213	16.833	-2.91	-14.741	0.431	0.236	0.614	0.297	0.163	0.424

SKiMS

4. Quality Assurance
Data Quality : Preparation for Evaluation

HOCHSCHULE FURTWANGEN UNIVERSITY HFU

Co2.csv

	A	B	C	D	E	F	G
1	Region/Country/Area	Land	Year	Series	Value	Footnotes	Source
2	8	Albania	1975	Emissions	4,524		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
3	8	Albania	1985	Emissions	7,145		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
4	8	Albania	2005	Emissions	3,980		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
5	8	Albania	2010	Emissions	4,074		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
6	8	Albania	2015	Emissions	3,975		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a
7	8	Albania	2018	Emissions	4,525		International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last a

```
Sub InsertRowsAndFillData()
    Dim ws As Worksheet
    Dim lastRow As Long
    Dim currentRow As Long
    Dim years As Variant
    Dim i As Integer

    ' Define the years to be inserted in column B
    years = Array(1975, 1985, 2005, 2010, 2015, 2018, 2019, 2020)

    ' Set the worksheet
    Set ws = ThisWorkbook.Sheets(4) ' Adjust if needed (e.g., Sheets("SheetName"))

    ' Find the last row with data in column A
    lastRow = ws.Cells(ws.Rows.Count, "A").End(xlUp).Row

    ' Loop through rows until the last row with text
    currentRow = 1
    Do While ws.Cells(currentRow, "A").Value <> ""
        ' Insert 7 rows below the current row
        ws.Rows(currentRow + 1 & ":" & currentRow + 6).Insert Shift:=xlDown

        ' Copy the value from column A of the current row to the 7 rows below
        ws.Range(ws.Cells(currentRow + 1, "A"), ws.Cells(currentRow + 7, "A")).Value = ws.Cells(currentRow, "A").Value

        ' Write the years into column B
        For i = 0 To UBound(years)
            ws.Cells(currentRow + i, "B").Value = years(i)
        Next i

        ' Move to the next set of rows
        currentRow = currentRow + 8
    Loop
End Sub
```

=IF(ISNA(

XLOOKUP(A3&B3,

Data!D4:D2259,

Data!F4:F2259)

),"Missing",

"Present")

	A	B
1	Country	Year
2	Albania	1975
3	Albania	1985
4	Albania	2005
5	Albania	2010
6	Albania	2015
7	Albania	2018
8	Albania	2019
9	Albania	2020
10	Angola	1975
11	Angola	1985
12	Angola	2005
13	Angola	2010
14	Angola	2015
15	Angola	2018
16	Angola	2019
17	Angola	2020
18	Armenia	1975
19	Armenia	1985
20	Armenia	2005
21	Armenia	2010
22	Armenia	2015
23	Armenia	2018
24	Armenia	2019
25	Armenia	2020
26	Austria	1975
27	Austria	1985
28	Austria	2005

Rules | Metrics | Expected | Data

4. Quality Assurance

Data Quality : Evaluation Results

Country	Year	Status in Dataset
Armenia	1975	Missing
Armenia	1985	Missing
Belarus	1975	Missing
Belarus	1985	Missing
Bosnia and Herzegovina	1975	Missing
Bosnia and Herzegovina	1985	Missing
Eritrea	1975	Missing
Eritrea	1985	Missing
Kazakhstan	1975	Missing
Kazakhstan	1985	Missing
Kosovo	1975	Missing
Kosovo	1985	Missing
Kyrgyzstan	1975	Missing
Kyrgyzstan	1985	Missing
Latvia	1975	Missing
Latvia	1985	Missing
Mongolia	1975	Missing
North Macedonia	1975	Missing
North Macedonia	1985	Missing
Republic of Moldova	1975	Missing
Republic of Moldova	1985	Missing
Russian Federation	1975	Missing
Russian Federation	1985	Missing
Serbia	1975	Missing
Serbia	1985	Missing

Example: Excel formula to compute distinct records under "COUNTRY" or "YEAR"

```
=SUM(IF(FREQUENCY(IF(LEN(Data!B3:B2258)>0,MATCH(Data!B3:B2258,Data!B3:B2258,0),""),IF(LEN(Data!B3:B2258)>0,MATCH(Data!B3:B2258,Data!B3:B2258,0),""))>0,1))
```

S. No.	Description	Value
1	Distinct Records under "COUNTRY"	149
2	Distinct Records under "YEAR"	8
3	Count of Expected Records under "COUNTRY-YEAR" Pair	1192
4	Count of Actual Records under "COUNTRY-YEAR" Pair	1132
5	Count of Missing Rows due to Unavailability of "COUNTRY" Data	0
6	Count of Missing Records due to Unavailability of "YEAR" Data	25

Consequence: Blank fields in dashboard if missing year is selected.

Potential Remediation Strategies:

1. Add row for missing year; replace missing Co2 value with '0'.
2. Add row for missing year; interpolate Co2 value with before/after.
3. Add row for missing year; fetch Co2 value from other source.
4. If not critical, leave the data set as it is; maintain status-quo. ✓

4. Quality Assurance
Source Quality : Download the Sources

- **Second Independent Reference Source**
  1. European Commission's Emission Database for Global Atmospheric Research (EDGAR)
     - **Link to Dataset**: EDGAR - The Emissions Database for Global Atmospheric Research
- **Third Independent Reference Source**
  2. Our World in Data (OWID) Group, Oxford University.
     - **Link to Dataset**: GitHub - owid/co2-data: Data on CO2 and greenhouse gas emissions by Our World in Data

Source Quality 2/8: Standardize Table Format

European Commission

Country	Year	SEARCH_STRING	Co2 (Megatons)
Aruba	1975	Aruba1975	0.057396273
Aruba	1985	Aruba1985	0.104138822
Aruba	2005	Aruba2005	0.384839427
Aruba	2010	Aruba2010	0.482206904
Aruba	2015	Aruba2015	0.462025593
Aruba	2018	Aruba2018	0.465881209
Aruba	2019	Aruba2019	0.557917266
Aruba	2020	Aruba2020	0.452553064

Our World in Data

Country	Year	SEARCH_STRING	Co2 (Megatons)
Afghanistan	1975	Afghanistan1975	2.121
Afghanistan	1985	Afghanistan1985	3.501
Afghanistan	2005	Afghanistan2005	1.89
Afghanistan	2010	Afghanistan2010	8.365
Afghanistan	2015	Afghanistan2015	9.67
Afghanistan	2018	Afghanistan2018	10.602
Afghanistan	2019	Afghanistan2019	10.825
Afghanistan	2020	Afghanistan2020	11.606

4. Quality Assurance
Source Quality : Standardization Procedure

Country	Year	SEARCH_STRING	Co2 (Kilotons)
Albania	1975	Albania1975	4,524
Albania	1985	Albania1985	7,145
Albania	2005	Albania2005	3,980
Albania	2010	Albania2010	4,074
Albania	2015	Albania2015	3,975
Albania	2018	Albania2018	4,525
Albania	2019	Albania2019	4,200
Albania	2020	Albania2020	3,512
Algeria	1975	Algeria1975	13,691
Algeria	1985	Algeria1985	42,446
Algeria	2005	Algeria2005	78,045
Algeria	2010	Algeria2010	96,452
Algeria	2015	Algeria2015	131,690
Algeria	2018	Algeria2018	138,496
Algeria	2019	Algeria2019	143,569
Algeria	2020	Algeria2020	135,599

UNdata
A world of information

(Kiloton of Co2 (10^3) / 1000) => Megaton of Co2 (10^6)

Country	Year	SEARCH_STRING	Co2 (Kilotons)	Co2 (Megatons)
Albania	1975	Albania1975	4,524	4.524
Albania	1985	Albania1985	7,145	7.145
Albania	2005	Albania2005	3,980	3.98
Albania	2010	Albania2010	4,074	4.074
Albania	2015	Albania2015	3,975	3.975
Albania	2018	Albania2018	4,525	4.525
Albania	2019	Albania2019	4,200	4.2
Albania	2020	Albania2020	3,512	3.512
Algeria	1975	Algeria1975	13,691	13.691
Algeria	1985	Algeria1985	42,446	42.446
Algeria	2005	Algeria2005	78,045	78.045
Algeria	2010	Algeria2010	96,452	96.452
Algeria	2015	Algeria2015	131,690	131.69
Algeria	2018	Algeria2018	138,496	138.496
Algeria	2019	Algeria2019	143,569	143.569
Algeria	2020	Algeria2020	135,599	135.599

Uniform Country Name across Datasets (examples):

- People’s Democratic Republic of Korea => North Korea
- United States of America => United States
- Democratic Republic of Congo => DR Congo
- Russian Federation => Russia
- Brunei Darussalam => Brunei
- Côte d’Ivoire => Ivory Coast
- Türkiye => Turkey

4. Quality Assurance
Source Quality : Standardization Procedure

	A	B	C	D	E	F	G	H	I	J
1	Country	Year	1975	1985	2005	2010	2015	2018	2019	2020
2	Aruba		0.057396273	0.104138822	0.384839427	0.482206904	0.462025593	0.465881209	0.557917266	0.452553064
3	Afghanistan		2.02896685	4.123374917	1.266340324	7.788893115	8.346520776	7.932004694	7.249068567	7.054132973
4	Angola		10.90465296	14.1037418	15.63885114	22.73989087	33.09749892	26.25888652	27.57321601	20.71091767
5	Anguilla		0.002593654	0.003680496	0.013679755	0.024024484	0.028026636	0.028247295	0.027604422	0.022803655
6	International Aviation		174.5824705	225.9081363	429.0246419	466.6570758	536.2136805	615.9375423	625.1414353	298.6556776
7	Albania		5.424117634	8.073940436	4.148476443	4.556850488	4.863525143	5.305582509	4.953544001	4.560776421
8	Curaçao		10.19675057	4.499205783	5.91286189	4.288051276	5.741944054	3.236726024	2.511433241	2.05481112
9	United Arab Emirates		36.67944373	51.81151718	122.4618737	171.9537309	206.6195196	187.0554933	197.6422573	195.49085
10	Argentina		90.91669608	94.21036792	153.6797539	177.404229	196.2521887	189.8389393	183.2869034	170.1202783
11	Armenia		13.84218973	18.25174887	4.462479612	4.298827714	5.61664906	5.950878354	6.424075918	6.913823302
12	Antigua and Barbuda		0.135544059	0.104658108	0.172036425	0.278491334	0.290809023	0.294943085	0.340322647	0.27590049
13	Australia		193.5950414	237.9354424	384.0856836	415.2731165	400.5424269	409.2066671	407.8603828	392.5190523

European Commission

```vba
Sub InsertAndTranspose()
    Dim ws As Worksheet
    Dim currentRow As Long, newRow As Long
    Dim lastRow As Long
    Dim dataRow As Long

    ' Set the worksheet
    Set ws = ThisWorkbook.Sheets(5) ' Adjust the sheet index or name as needed

    ' Start from row 2
    currentRow = 2

    Do While ws.Cells(currentRow, 1).Value <> "" ' Continue until column A is empty
        ' Insert 7 rows below currentRow
        ws.Rows(currentRow + 1 & ":" & currentRow + 7).Insert Shift:=xlDown

        ' Copy A2 into A3:A9
        ws.Range(ws.Cells(currentRow, 1), ws.Cells(currentRow, 1)).Copy _
            Destination:=ws.Range(ws.Cells(currentRow + 1, 1), ws.Cells(currentRow + 7, 1))

        ' Transpose range C1:J1 into range B2:B9
        ws.Range("C1:J1").Copy
        ws.Cells(currentRow, 2).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks:=False, Transpose:=Tru

        ' Transpose range D2:J2 into range C3:C8
        ws.Range(ws.Cells(currentRow, 4), ws.Cells(currentRow, 10)).Copy
        ws.Cells(currentRow + 1, 3).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks:=False, Transpose:

        ' Calculate the next starting row
        currentRow = currentRow + 8

        ' Check if A cell of the next row has data
        If ws.Cells(currentRow, 1).Value = "" Then Exit Do
    Loop

    ' Clear the clipboard
    Application.CutCopyMode = False
```

	A	B	C	D
1	Country	Year	SEARCH_STRING	Co2 (Megatons)
2	Aruba	1975	Aruba1975	0.057396273
3	Aruba	1985	Aruba1985	0.104138822
4	Aruba	2005	Aruba2005	0.384839427
5	Aruba	2010	Aruba2010	0.482206904
6	Aruba	2015	Aruba2015	0.462025593
7	Aruba	2018	Aruba2018	0.465881209
8	Aruba	2019	Aruba2019	0.557917266
9	Aruba	2020	Aruba2020	0.452553064
10	Afghanistan	1975	Afghanistan1975	2.02896685
11	Afghanistan	1985	Afghanistan1985	4.123374917
12	Afghanistan	2005	Afghanistan2005	1.266340324

4. Quality Assurance

Source Quality Merge Tables into Single Sheet

	A	B	D	E	F
1	Country	Year	Co2_S1_UN	Co2_S2_EU	Co2_S3_OX
2	Albania	1975	4.524	5.424	4.591
3	Albania	1985	7.145	8.074	7.882
4	Albania	2005	3.98	4.148	4.261
5	Albania	2010	4.074	4.557	4.784
6	Albania	2015	3.975	4.864	4.712
7	Albania	2018	4.525	5.306	4.895
8	Albania	2019	4.2	4.954	4.827
9	Albania	2020	3.512	4.561	4.711
10	Algeria	1975	13.691	36.407	32.001
11	Algeria	1985	42.446	71.684	72.609
12	Algeria	2005	78.045	101.397	110.993
13	Algeria	2010	96.452	118.615	118.353
14	Algeria	2015	131.69	164.862	160.087
15	Algeria	2018	138.496	174.836	172.705
16	Algeria	2019	143.569	181.649	179.907
17	Algeria	2020	135.599	171.893	168.178
18	Angola	1975	2.799	10.905	4.409
19	Angola	1985	3.819	14.104	4.676
20	Angola	2005	7.51	15.639	15.387
21	Angola	2010	15.863	22.740	22.983
22	Angola	2015	23.293	33.097	27.507
23	Angola	2018	18.989	26.259	23.305
24	Angola	2019	20.748	27.573	22.047
25	Angola	2020	16.939	20.711	16.533
26	Argentina	1975	85.498	90.917	94.845
27	Argentina	1985	88.181	94.210	100.531
28	Argentina	2005	143.25	153.680	161.589
29	Argentina	2010	164.171	177.404	185.946
30	Argentina	2015	181.809	196.252	191.744
31	Argentina	2018	173.309	189.839	180.599
32	Argentina	2019	164.341	183.287	178.508
33	Argentina	2020	150.666	170.120	165.121
34	Armenia	2005	4.163	4.462	4.376
35	Armenia	2010	4.118	4.299	4.253
36	Armenia	2015	5.222	5.617	5.469
37	Armenia	2018	5.522	5.951	5.822
38	Armenia	2019	5.986	6.424	6.284
39	Armenia	2020	6.464	6.914	6.795
40	Australia	1975	182.195	193.595	173.102
41	Australia	1985	223.305	237.935	224.556
42	Australia	2005	368.484	384.086	386.176

Comparision | UN | EU

4. Quality Assurance

Source Quality : Add Measures to Compute Variation

	A	B	D	E	F	G	H	I	J	K
1	Country	Year	Co2_S1_UN	Co2_S2_EU	Co2_S3_OX	Abs(S1 - S2)	%(S1 - S2)	Abs(S1 - S3)	%(S1 - S3)	Anomaly Flag
2	Albania	1975	4.524	5.424	4.591	0.900	20%	0.067	1%	-
3	Albania	1985	7.145	8.074	7.882	0.929	13%	0.737	10%	-
4	Albania	2005	3.98	4.148	4.261	0.168	4%	0.281	7%	-
5	Albania	2010	4.074	4.557	4.784	0.483	12%	0.71	17%	-
6	Albania	2015	3.975	4.864	4.712	0.889	22%	0.737	19%	-
7	Albania	2018	4.525	5.306	4.895	0.781	17%	0.37	8%	-
8	Albania	2019	4.2	4.954	4.827	0.754	18%	0.627	15%	-
9	Albania	2020	3.512	4.561	4.711	1.049	30%	1.199	34%	-
10	Algeria	1975	13.691	36.407	32.001	22.716	166%	18.31	134%	Check!
11	Algeria	1985	42.446	71.684	72.609	29.238	69%	30.163	71%	-
12	Algeria	2005	78.045	101.397	110.993	23.352	30%	32.948	42%	-
13	Algeria	2010	96.452	118.615	118.353	22.163	23%	21.901	23%	-
14	Algeria	2015	131.69	164.862	160.087	33.172	25%	28.397	22%	-
15	Algeria	2018	138.496	174.836	172.705	36.340	26%	34.209	25%	-
16	Algeria	2019	143.569	181.649	179.907	38.080	27%	36.338	25%	-
17	Algeria	2020	135.599	171.893	168.178	36.294	27%	32.579	24%	-
18	Angola	1975	2.799	10.905	4.409	8.106	290%	1.61	58%	Check!
19	Angola	1985	3.819	14.104	4.676	10.285	269%	0.857	22%	Check!
20	Angola	2005	7.51	15.639	15.387	8.129	108%	7.877	105%	Check!
21	Angola	2010	15.863	22.740	22.983	6.877	43%	7.12	45%	-
22	Angola	2015	23.293	33.097	27.507	9.804	42%	4.214	18%	-
23	Angola	2018	18.989	26.259	23.305	7.270	38%	4.316	23%	-
24	Angola	2019	20.748	27.573	22.047	6.825	33%	1.299	6%	-
25	Angola	2020	16.939	20.711	16.533	3.772	22%	0.406	2%	-
26	Argentina	1975	85.498	90.917	94.845	5.419	6%	9.347	11%	-
27	Argentina	1985	88.181	94.210	100.531	6.029	7%	12.35	14%	-
28	Argentina	2005	143.25	153.680	161.589	10.430	7%	18.339	13%	-
29	Argentina	2010	164.171	177.404	185.946	13.233	8%	21.775	13%	-
30	Argentina	2015	181.809	196.252	191.744	14.443	8%	9.935	5%	-
31	Argentina	2018	173.309	189.839	180.599	16.530	10%	7.29	4%	-
32	Argentina	2019	164.341	183.287	178.508	18.946	12%	14.167	9%	-
33	Argentina	2020	150.666	170.120	165.121	19.454	13%	14.455	10%	-
34	Armenia	2005	4.163	4.462	4.376	0.299	7%	0.213	5%	-
35	Armenia	2010	4.118	4.299	4.253	0.181	4%	0.135	3%	-
36	Armenia	2015	5.222	5.617	5.469	0.395	8%	0.247	5%	-
37	Armenia	2018	5.522	5.951	5.822	0.429	8%	0.3	5%	-
38	Armenia	2019	5.986	6.424	6.284	0.438	7%	0.298	5%	-
39	Armenia	2020	6.464	6.914	6.795	0.450	7%	0.331	5%	-
40	Australia	1975	182.195	193.595	173.102	11.400	6%	9.093	5%	-
41	Australia	1985	223.305	237.935	224.556	14.630	7%	1.251	1%	-
42	Australia	2005	368.484	384.086	386.176	15.602	4%	17.692	5%	-

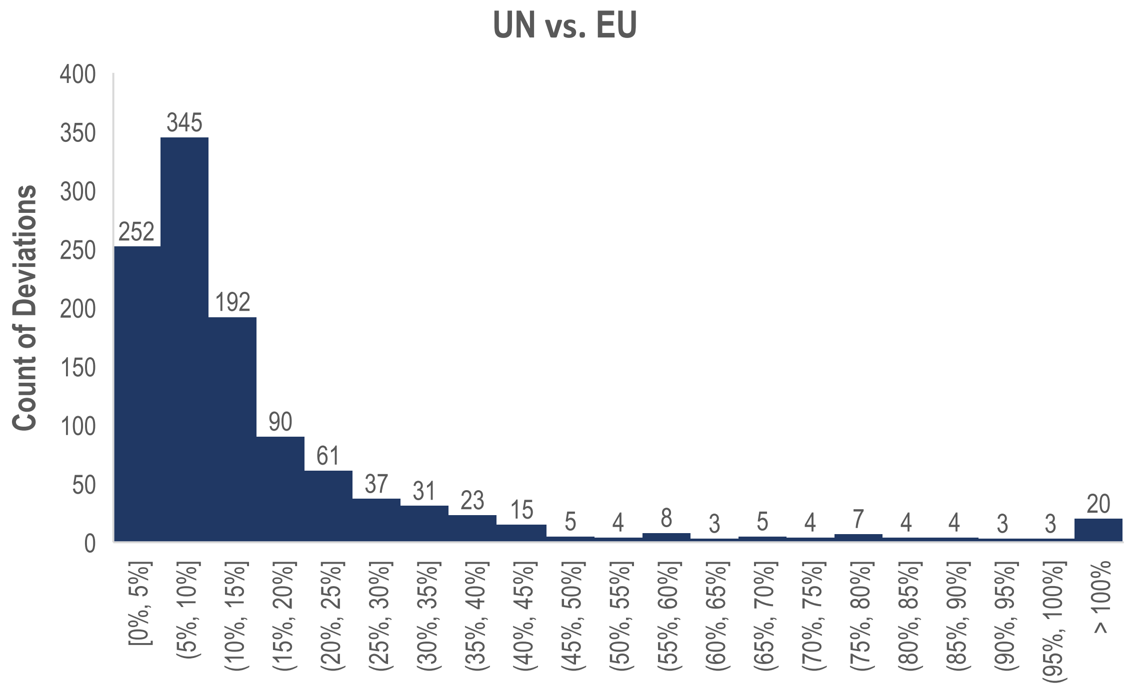
Comparision | UN | EU | OX | Example | +

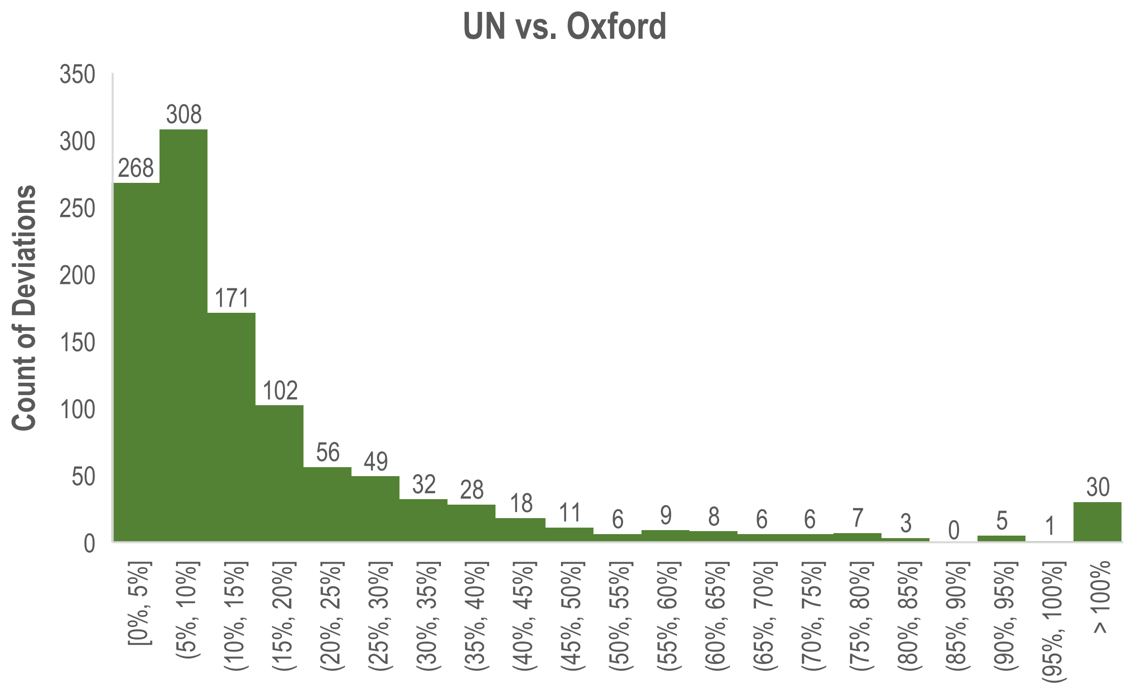
Column	Variation Measure
H	Absolute value of (Co2 value in source 1 – Co2 value in Source 2)
G	% Difference between Co2 value in source 1 and Co2 value in Source 2
I	Absolute value of (Co2 value in source 1 – Co2 value in Source 3)
J	% Difference between Co2 value in source 1 and Co2 value in Source 2

Anomaly Flag

=IF(OR(H10>100%,J10>100%),"Check!","-")

4. Quality Assurance
Source Quality : Results & Potential Reasons

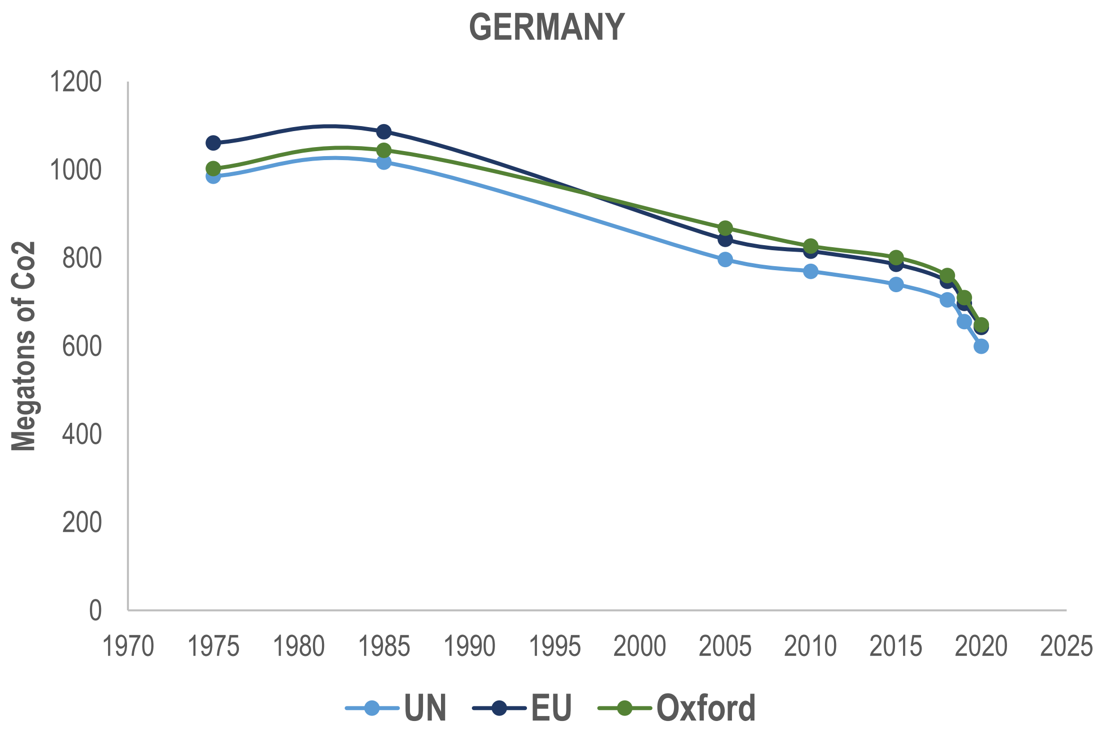


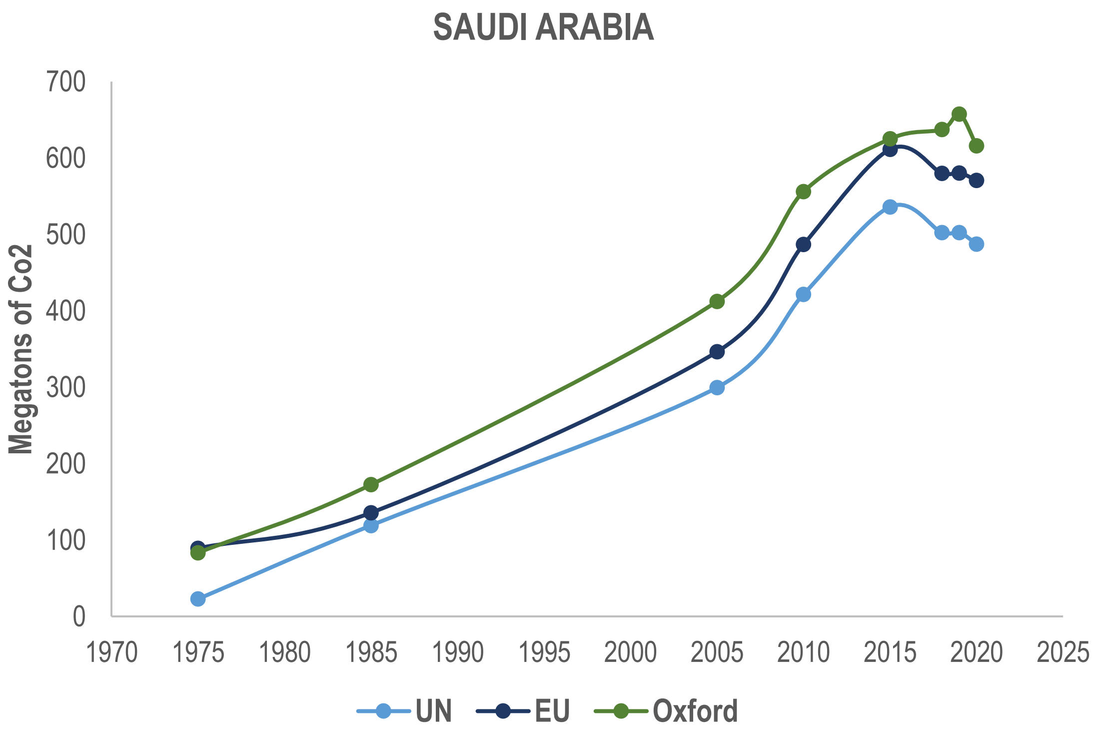


Potential reasons for variation:

1. Clubbed reporting: for example, Serbia AND Montenegro.
2. Variation in estimates: Not all figures are actuals; rather expected values extrapolated from samples onto the population.

4. Quality Assurance
Source Quality : Comparison





- Highly transparent with strict reporting requirements and practices.

- Opaque system with restricted access to sensitive data such as Co2 emissions.

4. Quality Assurance
Source Quality : Deviation Example

Country	Year	UN (S1)	EU (S2)	Oxford (S3)	Abs(S1 - S2)	%(S1 - S2)	Abs(S1 - S3)	%(S1 - S3)	Anomaly Flag
Saudi Arabia	1975	22.7	88.9	83.3	66.1	291%	61	266%	Check!
Saudi Arabia	1985	119.1	135.5	172.4	16.4	14%	53	45%	-
Saudi Arabia	2005	299.6	345.9	412.4	46.4	15%	113	38%	-
Saudi Arabia	2010	421.7	486.5	556.0	64.8	15%	134	32%	-
Saudi Arabia	2015	536.1	611.0	625.0	74.9	14%	89	17%	-
Saudi Arabia	2018	502.4	580.0	637.5	77.7	15%	135	27%	-
Saudi Arabia	2019	502.5	579.9	657.6	77.4	15%	155	31%	-
Saudi Arabia	2020	487.5	571.3	616.1	83.8	17%	129	26%	-

5. Design & Implementation
KPI Definition

Emissions

- **CO_2 Emission Growth Rate**: Annual % change.
- **Per Capita CO_2 Emissions**: Emissions by population.
- **Top Emitting Countries**: Identify leaders by emissions.

Water & Sanitation

- **Clean Water Access:** % population with clean water.
- **Sanitation Coverage:** % with improved facilities.
- **Water Stress Levels**: Regions with higher water demand.

Biodiversity

- **Threatened Species Count** Species at risk by category.
- **Change in Threatened Species:** Annual increase/decrease.
- **Biodiversity Hotspots**: Regions with high threat levels.

Land use

- **Land Distribution:** % for agriculture, urban, conversation.
- **Deforestation Rate:** Annual forest area change.
- **Urbanization Growth**: Urban land % increase.

5. Design & Implementation

KPI Definition

NAME OF KEY FIGURE	CHANGE IN CO_2 EMISSIONS PER COUNTRY PER YEAR
DESCRIPTION	**DETAILS**
1. Identification	CO_2 emissions rate.
2. Description	This KPI measures the annual percentage change in CO_2 emissions between two years.
3. Addressee	Government, NGOs, Researchers & Environmental Strategists.
4. Formula	[(CO_2emissionsinyear N – CO_2emissionsinyear N – 1) / CO_2emissionsinyear N-1] x 100
5. Target value	Reduction target: ≥ -2% annually per country.
6. Unit	Percentage (%)
7. Periodicity	Year
8. Dimensions	Year, Series, Land
9. Aggregations behavior	Average, min, max, trend
10. Planned values	Predetermined yearly targets set by international agreements (e.g., Paris Agreement goals).
11. Tolerance values	±0.5% from the target.
12. Escalation norms	Trigger intervention if annual change deviates by more than ±0.5% for consecutive years.
13. Validity/periodicity	01.01.1975 – 31.12.9999
14. Dependency	International Energy Agency, IEA World Energy Balances 2021 and 2006 IPCC Guidelines for Greenhouse Gas Inventories, last accessed April 2023.
15. Person in charge	Government and people.
16. Example	If the CO_2 emissions per year in 2020 were 599,289 metric tons and 2019 were 655,181 metric tons, the change would be: (599,289-655,181)/599,289 x 100 = **− 9.32 %**

5. Design & Implementation
KPI Definition

NAME OF KEY FIGURE	CHANGE OF THREATENED SPECIES BETWEEN TWO YEARS IN %
DESCRIPTION	**DETAILS**
1. Identification	Threatened species rate.
2. Description	This KPI tracks the increase or decrease in the number of threatened species in a given region between two years.
3. Addressee	Government, NGOs, Researchers & Environmental Strategists.
4. Formula	[(#threatenedspeciesinyear N - #threatenedspeciesinyear N-1)/ #threatenedspeciesinyear N-1] x 100
5. Target Value	Reduction target: ≥ 10% fewer threatened species annually per region/country.
6. Unit	Percentage (%)
7. Periodicity	Year
8. Dimensions	Year, Series, Land
9. Aggregations Behavior	Average, min, max, trend
10. Planned Values	Planned reduction targets aligned with biodiversity conservation goals (e.g., Convention on Biological Diversity).
11. Tolerance Values	±0.5% from target values
12. Escalation Norms	Trigger intervention if no significant improvement is observed over consecutive years.
13. Validity/Periodicity	01.01.2004 – 31.12.9999
14. Dependency	World Conservation Union (IUCN), Gland and Cambridge, IUCN Red List of Threatened Species publication, last accessed May 2023.
15. Person in Charge	Government
16. Example	If a region had 98 threatened species in 2022 and 61 in 2021, the change in number will be [(98-61)/61]x100= **60.6%**

SKiMS

5. Design & Implementation

KPI Definition

NAME OF KEY FIGURE	REGIONAL DISPARITIES IN WATER & SANITATION ACCESS (%)
DESCRIPTION	**DETAILS**
1. Identification	Disparities % in water and sanitation in different regions.
2. Description	This KPI measures the proportion of rural population, having access to water and sanitation facilities compare to urban population.
3. Addressee	Government, NGOs, Researchers & Environmental Strategists.
4. Formula	(proportionofruralpopulationwithaccesstofacilities / proportionofurbanpopulationwithaccesstofacilities) x 100
5. Target Value	Universal access (100%) in all regions by 2030, aligned with SDG 6.
6. Unit	Percentage (%).
7. Periodicity	Annual
8. Dimensions	Year, Series, Land
9. Aggregations Behavior	Average, minimum, maximum, trend analysis.
10. Planned Values	Progressive targets (e.g., 90% by 2025, 100% by 2030).
11. Tolerance Values	± 2% from annual improvement targets
12. Escalation Norms	Immediate action if access decreases or stagnates for two consecutive years.
13. Validity/Periodicity	01.01.2006 – 31.12.9999
14. Dependency	World Health Organization (WHO) and United Nations Children's Fund (UNICEF), Geneva and New York, the WHO/UNICEF Joint Monitoring Programme for the Water and Sanitation database, last accessed July 2023.
15. Person in Charge	Government
16. Example	If 60% of the rural population has access to facilities and 90% of the urban population has access, the formula calculates the disparity between urban and rural access facilities, which is 90/60 x 100= 66.6%

5. Design & Implementation

KPI Definition

NAME OF KEY FIGURE	CHANGE IN FOREST COVER AS A PERCENTAGE OF TOTAL LAND AREA BETWEEN TWO YEARS
DESCRIPTION	**DETAILS**
1. Identification	Rate of increase/decrease of forest cover.
2. Description	This KPI tracks the change in forest cover between two years, expressed as a percentage of the total land area, to assess deforestation and reforestation trends.
3. Addressee	Government, NGOs, Researchers & Environmental Strategists.
4. Formula	[(Forestcoveripercentageinyear N – Forestcoverpercentageinyear N – 1) / Forestcoverpercentageinyear N-1] x 100.
5. Target Value	Net positive change (≥ 0%) annually.
6. Unit	Percentage (%).
7. Periodicity	Year
8. Dimensions	Year, Series, Land
9. Aggregations Behavior	Average, minimum, maximum, trend analysis.
10. Planned Values	Planned reforestation or afforestation targets per region/country.
11. Tolerance Values	±0.5% deviation from target values
12. Escalation Norms	Escalation triggered if forest loss exceeds 1% for two consecutive years.
13. Validity/Periodicity	01.01.2005 – 31.12.9999
14. Dependency	Food and Agriculture Organization of the United Nations (FAO), Rome, FAOSTAT data last accessed December 2022.
15. Person in Charge	Government
16. Example	If forest cover 61.2% in 2010 and 59.4% in 2020, the change would be [(59.4 – 61.2)/61.2] x 100 = **- 2.94 %**

LAND DATA TYPE

ATTRIBUTE NAME	ATTRIBUTE CATEGORY	DATA TYPE	EXAMPLE
Country_ID	Nominal	Int	356
Land_Name	Nominal	Varchar	India
Year	Ordinal	Int	2020
Series	Nominal	Varchar	1. Land (1000 Hectares) 2. Forest Cover (1000 Hectares) 3. Permanent crops (1000 Hectares) 4. Total land cover (1000 Hectares)
Value	Continuous	Decimal	155.37

5. Design & Implementation

Data Category & Data Types

CO_2 DATA TYPE

ATTRIBUTE NAME	ATTRIBUTE CATEGORY	DATA TYPE	EXAMPLE
Country_ID	Nominal	Int	826
Land_Name	Nominal	Varchar	United Kingdom
Year	Ordinal	Int	2020
Series	Nominal	Varchar	1. Emissions (1000 metrics tons of CO_2) 2. Emissions per capita (metric tons of CO_2)
Value	Continuous	Decimal	306.974

SPECIES DATA TYPE

ATTRIBUTE NAME	ATTRIBUTE CATEGORY	DATA TYPE	EXAMPLE
Country_ID	Nominal	Int	76
Land_Name	Nominal	Varchar	Brazil
Year	Ordinal	Int	2022
Series	Nominal	Varchar	1. Threatened species: Plants (number) 2. Threatened species: Vertebrates (number) 3. Threatened species: Invertebrates (number) 4. Threatened species: Total (number)
Value	Discrete	Int	1373

5. Design & Implementation

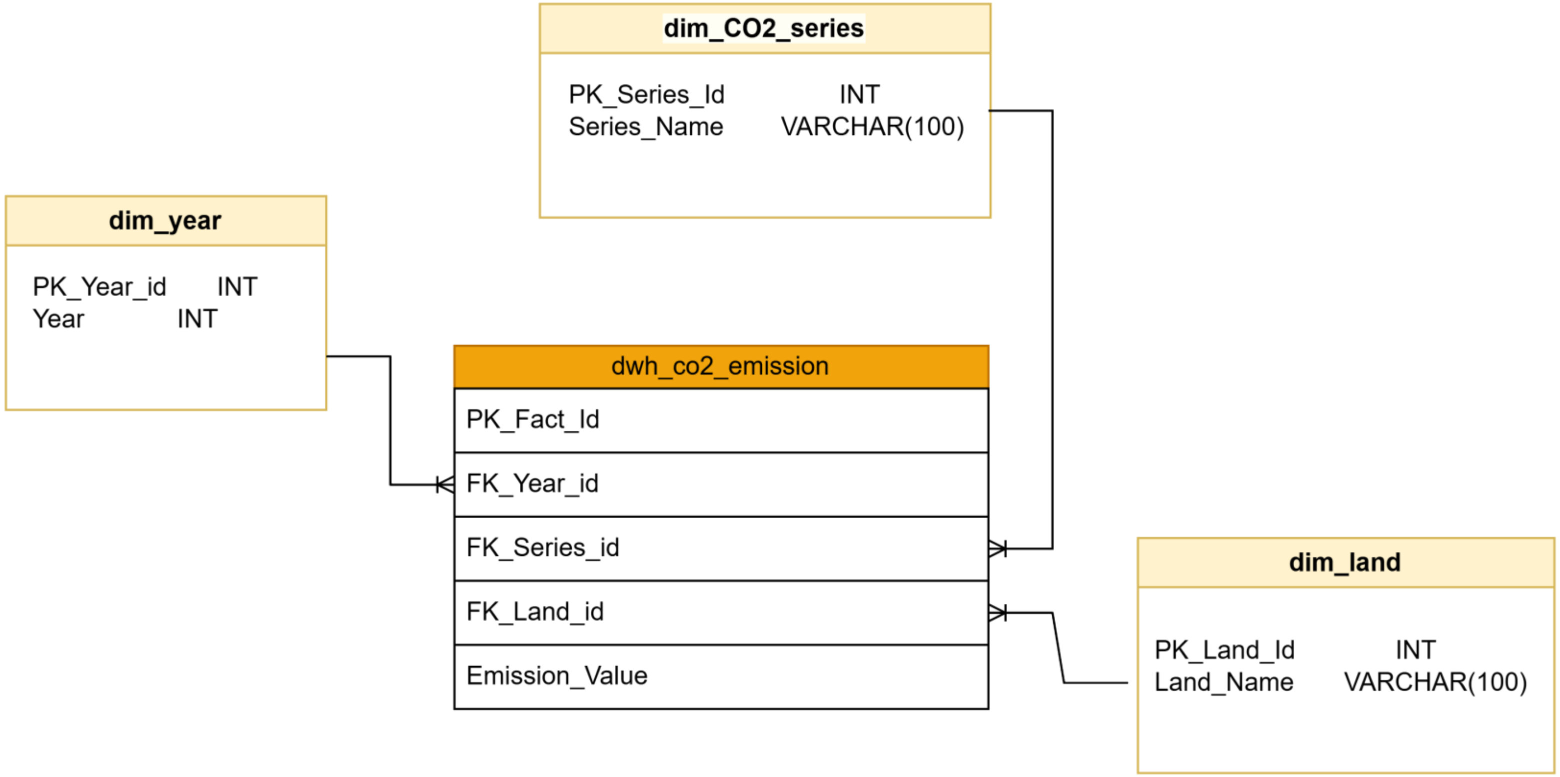
Data Category & Data Types

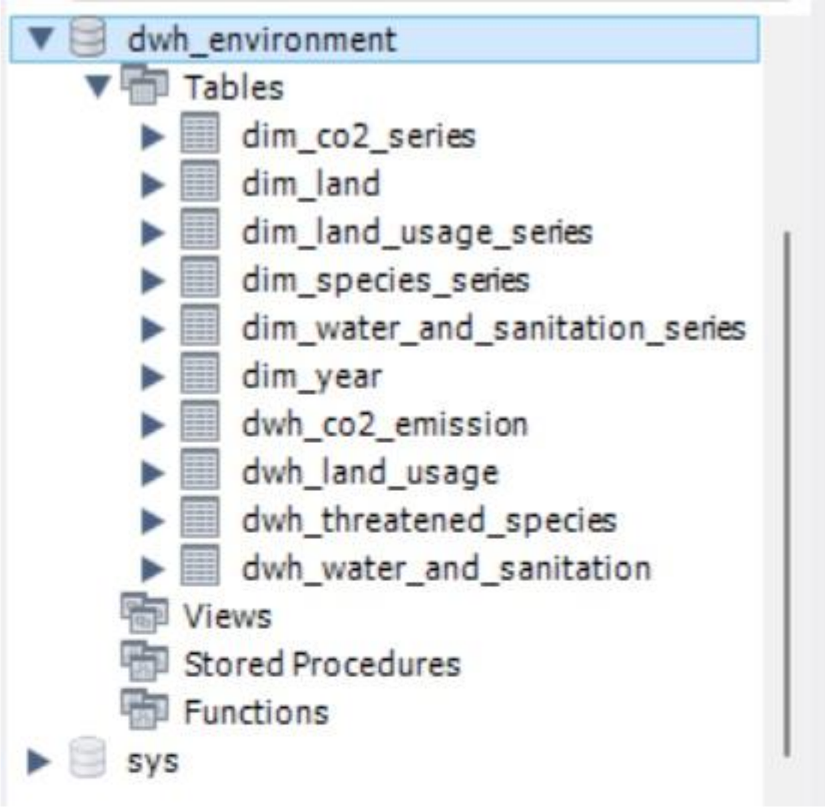
DRINKING WATER AND SANITATION FACILITIES DATA TYPE

ATTRIBUTE NAME	ATTRIBUTE CATEGORY	DATA TYPE	EXAMPLE
Country_ID	Nominal	Int	276
Land_Name	Nominal	Varchar	Germany
Year	Ordinal	Int	2020
Series	Nominal	Varchar	1. Drinking water access: Rural (% of population) 2. Drinking water access: Urban (% of population) 3. Drinking water access: Total (% of population) 4. Sanitation facilities: Rural (% of population) 5. Sanitation facilities: Urban (% of population) 6. Sanitation facilities: Total (% of population)
Value	Continuous	Decimal	99.9

5. Design & Implementation

Entity Relationship Diagram





5. Design & Implementation

ETL Process

Data Cleaning

```python
# -*- coding: utf-8 -*-
"""
Created on Tue Nov 12 10:01:40 2024

@author: subhashri ravichandran
"""

import pandas as pd

df = pd.read_csv("C:\\Users\\subha\\OneDrive\\Dokumente\\BCM\\WinterSemester\\Business Intelligence\\Project\\Final_Files\\fact tables\\psa\\Co2.csv",encoding='ISO-8859-1')
df['Value'] = df['Value'].str.replace(',','')
df.to_csv("C:\\Users\\subha\\OneDrive\\Dokumente\\BCM\\WinterSemester\\Business Intelligence\\Project\\Final_Files\\fact tables\\psa\\New folder\\Co2Updated.csv", index = False)
print("process complete")
df2 = pd.read_csv("C:\\Users\\subha\\OneDrive\\Dokumente\\BCM\\WinterSemester\\Business Intelligence\\Project\\Final_Files\\fact tables\\psa\\Land.csv",encoding='ISO-8859-1')
df2['Value'] = df2['Value'].str.replace(',','')
df2.to_csv("C:\\Users\\subha\\OneDrive\\Dokumente\\BCM\\WinterSemester\\Business Intelligence\\Project\\Final_Files\\fact tables\\psa\\New folder\\LandUpdated.csv", index = False)
print("process complete")
df3 = pd.read_csv("C:\\Users\\subha\\OneDrive\\Dokumente\\BCM\\WinterSemester\\Business Intelligence\\Project\\Final_Files\\fact tables\\psa\\Species.csv",encoding='ISO-8859-1')
df3['Value'] = df3['Value'].str.replace(',','')
df3.to_csv("C:\\Users\\subha\\OneDrive\\Dokumente\\BCM\\WinterSemester\\Business Intelligence\\Project\\Final_Files\\fact tables\\psa\\New folder\\SpeciesUpdated.csv", index = False)
print("process complete")
```

Before – co2.csv

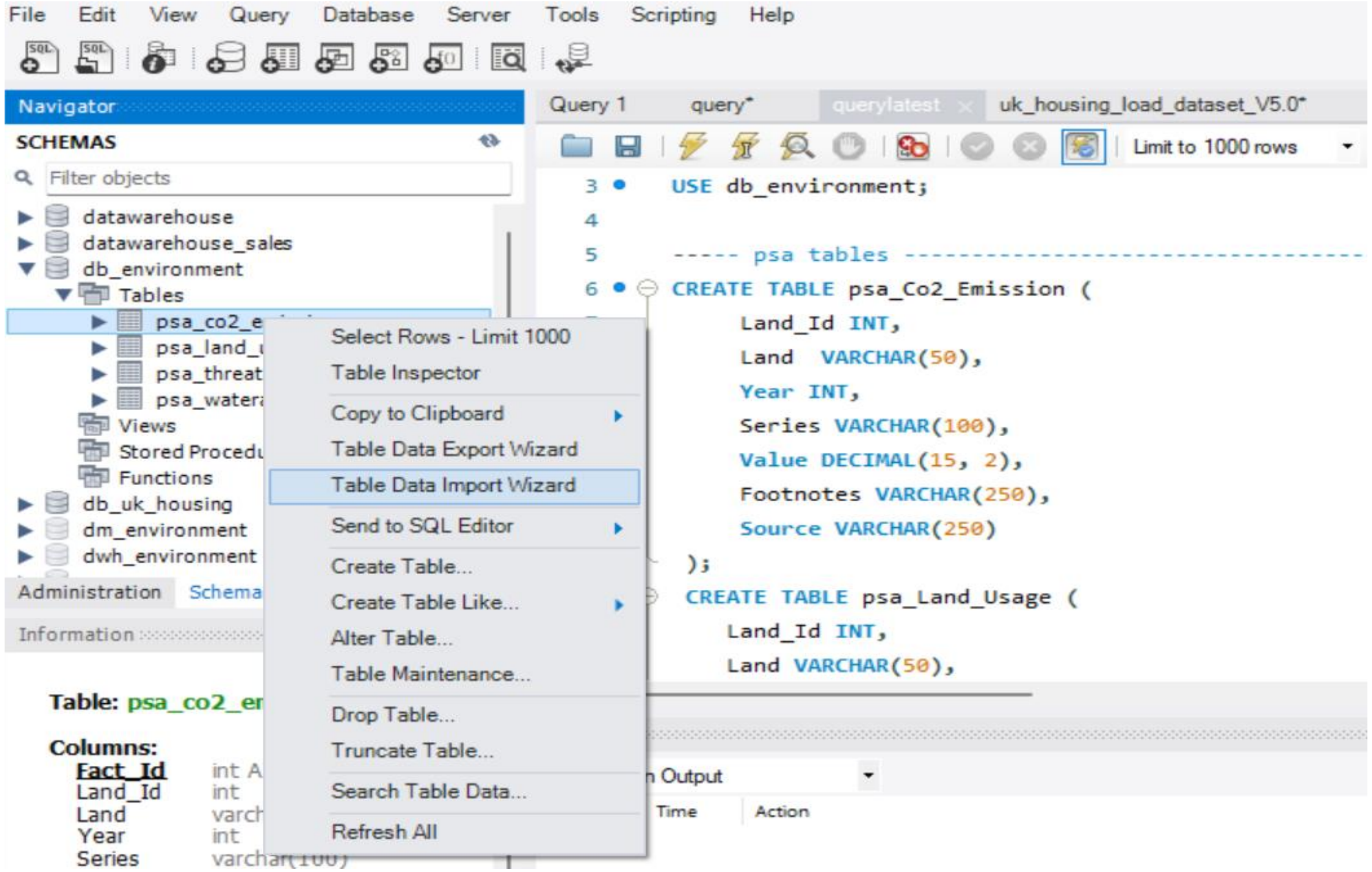
	A	B	C	D	E	F	G
1	Region/Country/Area	Land	Year	Series	Value	Footnotes	Source
2	8	Albania	1975	Emissions (thou	4,524		Internatio
3	8	Albania	1985	Emissions (thou	7,145		Internatio
4	8	Albania	2005	Emissions (thou	3,980		Internatio
5	8	Albania	2010	Emissions (thou	4,074		Internatio
6	8	Albania	2015	Emissions (thou	3,975		Internatio
7	8	Albania	2018	Emissions (thou	4,525		Internatio
8	8	Albania	2019	Emissions (thou	4,200		Internatio
9	8	Albania	2020	Emissions (thou	3,512		Internatio

After – co2.csv

	A	B	C	D	E	F	G
1	Land_Id	Land	Year	Series	Value	Footnotes	Source
2	8	Albania	1975	Emissions	4524		International
3	8	Albania	1985	Emissions	7145		International
4	8	Albania	2005	Emissions	3980		International
5	8	Albania	2010	Emissions	4074		International
6	8	Albania	2015	Emissions	3975		International
7	8	Albania	2018	Emissions	4525		International
8	8	Albania	2019	Emissions	4200		International
9	8	Albania	2020	Emissions	3512		International

ETL Process

PSA-Level



Loading data into PSA tables

5. Design & Implementation

ETL Process

PSA-Level

Limit to 1000 rows

```sql
----- adding primary keys to psa tables ----------------------------------------------------------
ALTER TABLE psa_Co2_Emission ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;
ALTER TABLE psa_Land_Usage ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;
ALTER TABLE psa_ThreatenedSpecies ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;
ALTER TABLE psa_WaterandSanitation ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;

----- verifying if the data is inserted to psa tables ------------------------------------------------
SELECT * FROM psa_Co2_Emission;
SELECT count(*) FROM psa_Co2_Emission;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch r

Fact_Id	Land_Id	Land	Year	Series	Value	Footnotes	Source
1	72	Botswana	2010	Emissions per capita (metric tons of carbon dioxi...	1.70		International Er
2	72	Botswana	2015	Emissions per capita (metric tons of carbon dioxi...	3.30		International Er
3	72	Botswana	2018	Emissions per capita (metric tons of carbon dioxi...	3.60		International Er
4	72	Botswana	2019	Emissions per capita (metric tons of carbon dioxi...	3.10		International Er
5	72	Botswana	2020	Emissions per capita (metric tons of carbon dioxi...	2.50		International Er
6	76	Brazil	1975	Emissions (thousand metric tons of carbon dioxi...	141202.00		International Er
7	76	Brazil	1985	Emissions (thousand metric tons of carbon dioxi...	166948.00		International Er

Data in raw format –psa_co2_emission

Limit to 1000 rows

```sql
ALTER TABLE psa_Co2_Emission ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;
ALTER TABLE psa_Land_Usage ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;
ALTER TABLE psa_ThreatenedSpecies ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;
ALTER TABLE psa_WaterandSanitation ADD COLUMN Fact_Id INT  AUTO_INCREMENT PRIMARY KEY FIRST;

----- verifying if the data is inserted to psa tables ----------------------------------------
SELECT * FROM psa_Co2_Emission;
SELECT count(*) FROM psa_Co2_Emission;
SELECT * FROM psa_Land_Usage;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(*)
2264

Count of records in psa_co2_emission

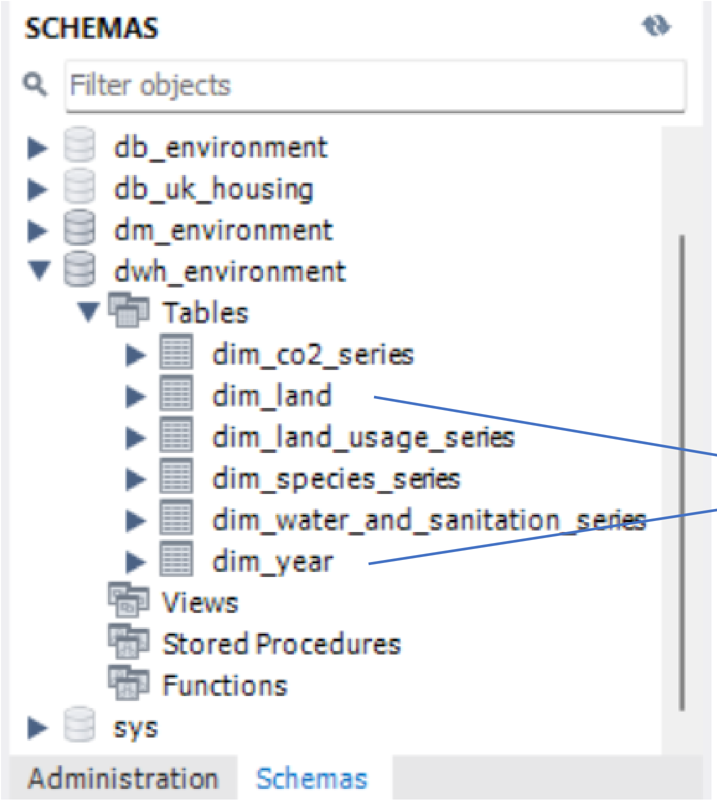
5. Design & Implementation

ETL Process

DWH-Level

```sql
CREATE SCHEMA dwh_Environment;
USE dwh_Environment;

----- dimension tables ------
CREATE TABLE Dim_Land (
    Land_Id INT PRIMARY KEY,
    Land_Name VARCHAR(100) NOT NULL
);
CREATE TABLE dim_Year (
    Year_Id INT PRIMARY KEY AUTO_INCREMENT,
    Year INT NOT NULL
);
CREATE TABLE Dim_CO2_Series (
    Series_Id INT PRIMARY KEY AUTO_INCREMENT,
    Series_Name VARCHAR(100) NOT NULL
);
CREATE TABLE Dim_Land_Usage_Series (
    Series_Id INT PRIMARY KEY AUTO_INCREMENT,
    Series_Name VARCHAR(100) NOT NULL
);
CREATE TABLE Dim_Species_Series (
    Series_Id INT PRIMARY KEY AUTO_INCREMENT,
    Series_Name VARCHAR(100) NOT NULL
);
CREATE TABLE Dim_Water_and_Sanitation_Series (
    Series_Id INT PRIMARY KEY AUTO_INCREMENT,
    Series_Name VARCHAR(100) NOT NULL
);
```



Common for all fact tables

Creating 6 dim tables in datawarehouse

5. Design & Implementation

ETL Process

DWH-Level

Common dimension table 1 - dim_land

```sql
----- insert values from psa tables to dimension tables ------
INSERT INTO dim_Land (Land_Id, Land_Name)
SELECT Land_Id, Land
FROM (
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_Co2_Emission
    UNION
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_Land_Usage
    UNION
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_ThreatenedSpecies
    UNION
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_WaterandSanitation
) AS unique_Lands;
SELECT * FROM dim_Land;
SELECT count(*) FROM dim_Land;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Co

Land_Id	Land_Name
1	Total, all countries or areas
2	Africa
4	Afghanistan
5	South America
8	Albania

Primary key – Land_Id

Inserting values into dim_land

```sql
INSERT INTO dim_Land (Land_Id, Land_Name)
SELECT Land_Id, Land
FROM (
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_Co2_Emission
    UNION
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_Land_Usage
    UNION
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_ThreatenedSpecies
    UNION
    SELECT DISTINCT Land_Id, Land FROM db_Environment.psa_WaterandSanitation
) AS unique_Lands;
SELECT * FROM dim_Land;
SELECT count(*) FROM dim_Land;
INSERT INTO dim_Year (Year)
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(*)
290

Count of records in dim_land

SKiMS

5. Design & Implementation
ETL Process
DWH-Level

HOCHSCHULE FURTWANGEN UNIVERSITY HFU

Common dimension Table 2 - dim_year

```sql
SELECT count(*) FROM dim_Land;
INSERT INTO dim_Year (Year)
SELECT DISTINCT Year
FROM (
    SELECT Year FROM db_Environment.psa_Co2_Emission
    UNION
    SELECT Year FROM db_Environment.psa_Land_Usage
    UNION
    SELECT Year FROM db_Environment.psa_ThreatenedSpecies
    UNION
    SELECT Year FROM db_Environment.psa_WaterandSanitation
) AS unique_Years;
SELECT * FROM dim_Year;
SELECT count(*) FROM dim_Year;
```

Year_Id	Year
1	2010
2	2015
3	2018
4	2019
5	2020

Primary key – Year_Id

Inserting values into dim_year

```sql
INSERT INTO dim_Year (Year)
SELECT DISTINCT Year
FROM (
    SELECT Year FROM db_Environment.psa_Co2_Emission
    UNION
    SELECT Year FROM db_Environment.psa_Land_Usage
    UNION
    SELECT Year FROM db_Environment.psa_ThreatenedSpecies
    UNION
    SELECT Year FROM db_Environment.psa_WaterandSanitation
) AS unique_Years;
SELECT * FROM dim_Year;
SELECT count(*) FROM dim_Year;
```

count(*)
16

Count of records in dim_year

5. Design & Implementation

ETL Process

DWH-Level

Fact table specific dimension tables - dim_co2_series, dim_land_usage_series, dim_species series, dim_water_and_sanitation_series

```sql
INSERT INTO dim_CO2_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_Co2_Emission;

SELECT * FROM Dim_CO2_Series;
SELECT count(*) FROM Dim_CO2_Series;

INSERT INTO Dim_Land_Usage_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_Land_Usage;

INSERT INTO Dim_Species_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_ThreatenedSpecies;

INSERT INTO Dim_Water_and_Sanitation_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_WaterandSanitation;
```

Inserting values into dimension tables

```sql
INSERT INTO dim_CO2_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_Co2_Emission;

SELECT * FROM Dim_CO2_Series;
SELECT count(*) FROM Dim_CO2_Series;

INSERT INTO Dim_Land_Usage_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_Land_Usage;

INSERT INTO Dim_Species_Series (Series_Name)
SELECT DISTINCT Series
FROM db_Environment.psa_ThreatenedSpecies;
```

Series_Id	Series_Name
1	Emissions per capita (metric tons of carbon dioxide)
2	Emissions (thousand metric tons of carbon dioxide)
NULL	NULL

Primary key – Series_Id

Records in dim_co2_series

5. Design & Implementation

ETL Process

DWH-Level

Fact tables - dwh_co2_series, dwh_land_usage_series, dwh_species series, dwh_water_and_sanitation_series

```sql
FROM db_Environment.psa_WaterandSanitation;

----- dwh tables ------------------
CREATE TABLE dwh_Co2_Emission (
    Fact_Id INT PRIMARY KEY,
    Land_Id INT,
    Year_Id INT,
    Series_Id INT,
    Emission_Value DECIMAL(15, 2),
    FOREIGN KEY (Land_Id) REFERENCES dwh_Environment.Dim_Land(Land_Id),
    FOREIGN KEY (Year_Id) REFERENCES dwh_Environment.dim_Year(Year_Id),
    FOREIGN KEY (Series_Id) REFERENCES dwh_Environment.Dim_CO2_Series(Series_Id)
);
CREATE TABLE dwh_Land_Usage (
    Fact_Id INT PRIMARY KEY,
    Land_Id INT,
    Year_Id INT,
    Series_Id INT,
    Land_Metric_Value DECIMAL(15, 2),
    FOREIGN KEY (Land_Id) REFERENCES dwh_Environment.Dim_Land(Land_Id),
    FOREIGN KEY (Year_Id) REFERENCES dwh_Environment.dim_Year(Year_Id),
    FOREIGN KEY (Series_Id) REFERENCES dwh_Environment.Dim_Land_Usage_Series(Series_Id)
);
```

Foreign keys to dim tables

```sql
    FOREIGN KEY (Series_Id) REFERENCES dwh_Environment.Dim_Land_Usage_Series(Series_Id)
);
CREATE TABLE dwh_Threatened_Species (
    Fact_Id INT PRIMARY KEY,
    Land_Id INT,
    Year_Id INT,
    Series_Id INT,
    Threatened_Species_Count INT,
    FOREIGN KEY (Land_Id) REFERENCES dwh_Environment.Dim_Land(Land_Id),
    FOREIGN KEY (Year_Id) REFERENCES dwh_Environment.dim_Year(Year_Id),
    FOREIGN KEY (Series_Id) REFERENCES dwh_Environment.Dim_Species_Series(Series_Id)
);
CREATE TABLE dwh_Water_and_Sanitation (
    Fact_Id INT PRIMARY KEY,
    Land_Id INT,
    Year_Id INT,
    Series_Id INT,
    Service_Coverage_Percentage DECIMAL(15, 2),
    FOREIGN KEY (Land_Id) REFERENCES dwh_Environment.Dim_Land(Land_Id),
    FOREIGN KEY (Year_Id) REFERENCES dwh_Environment.dim_Year(Year_Id),
    FOREIGN KEY (Series_Id) REFERENCES dwh_Environment.Dim_Water_and_Sanitation_Series(Series_Id)
);
```

Creating fact tables

5. Design & Implementation

ETL Process

Limit to 1000 rows

```sql
----- insert values from psa tables to datawarehouse tables ------
----- Insert data into dwh_Co2_Emission ------
INSERT INTO dwh_Co2_Emission (Fact_Id, Land_Id, Year_Id, Series_Id, Emission_Value)
SELECT
    psa.Fact_Id,
    dim_L.Land_Id,
    dim_Y.Year_Id,
    dim_S.Series_Id,
    psa.Value
FROM
    db_Environment.psa_Co2_Emission AS psa
JOIN
    dwh_Environment.Dim_Land AS dim_L ON psa.Land_Id = dim_L.Land_Id
JOIN
    dwh_Environment.dim_Year AS dim_Y ON psa.Year = dim_Y.Year
JOIN
    dwh_Environment.Dim_CO2_Series AS dim_S ON psa.Series = dim_S.Series_Name;
```

Inserting values from psa to fact table

Limit to 1000 rows

```sql
SELECT
    psa.Fact_Id,
    dim_L.Land_Id,
    dim_Y.Year_Id,
    dim_S.Series_Id,
    psa.Value
FROM
    db_Environment.psa_Co2_Emission AS psa
JOIN
    dwh_Environment.Dim_Land AS dim_L ON psa.Land_Id = dim_L.Land_Id
JOIN
    dwh_Environment.dim_Year AS dim_Y ON psa.Year = dim_Y.Year
JOIN
    dwh_Environment.Dim_CO2_Series AS dim_S ON psa.Series = dim_S.Series_Name;

SELECT COUNT(*) AS TotalRecords FROM dwh_Co2_Emission;
```

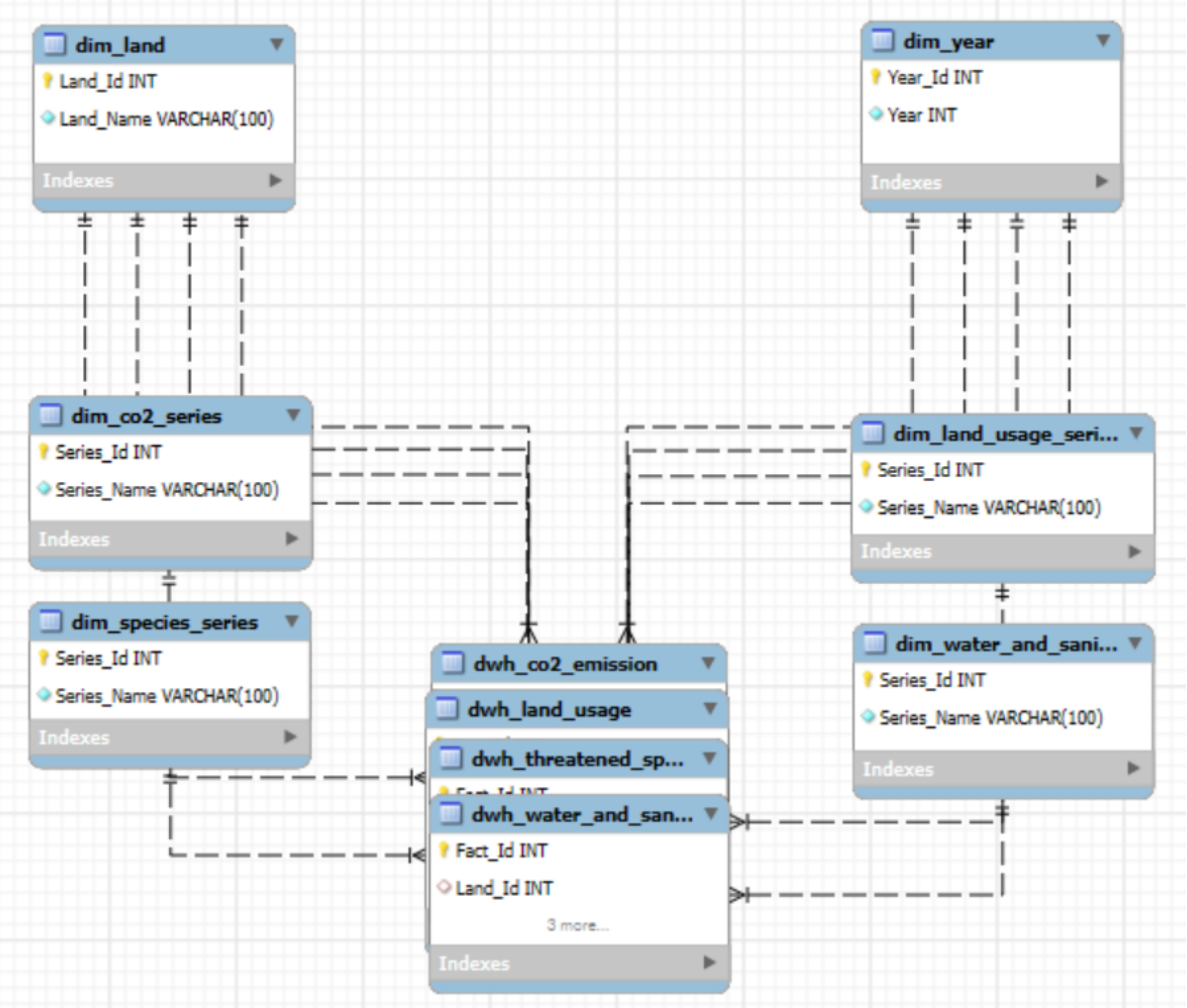
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

TotalRecords
2264

Count of records in dwh_co2_emission

5. Design & Implementation

EER Diagram



Snowflake schema

Fact Tables - 4

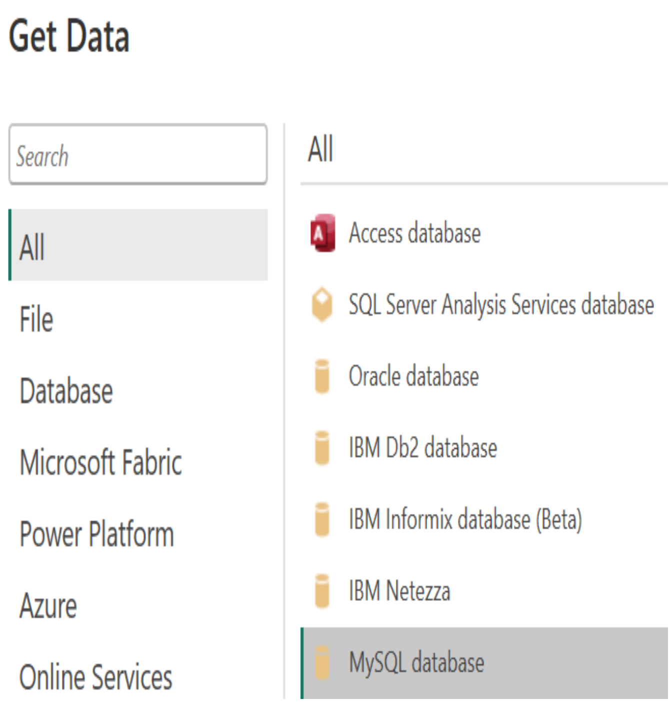
dwh_co2_emission
dwh_land_usage
dwh_threatened_species
dwh_water_and_sanitation

Dimension Tables - 6

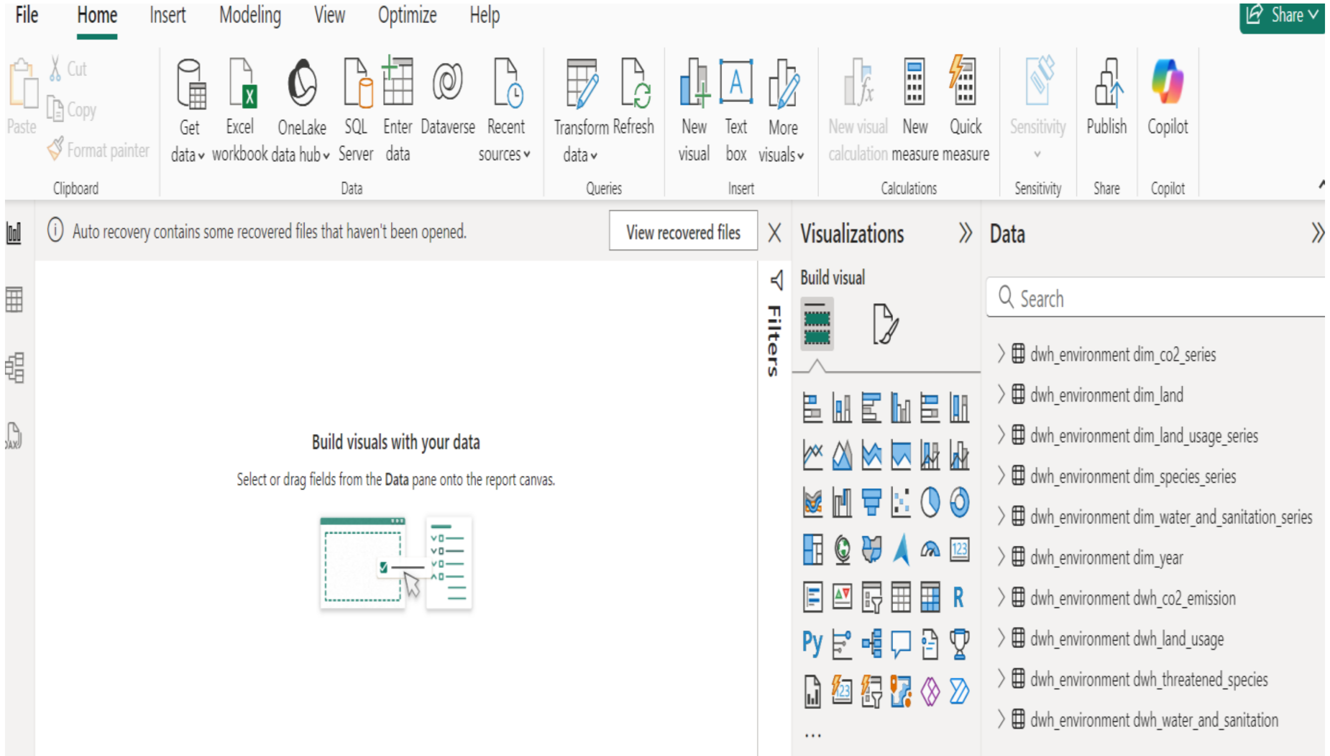
dim_co2_series
dim_land_usage_series
dim_species_series
dim_water_and_sanitation_series
dim_land
dim_year

5. Design & Implementation

Load Data into Power BI

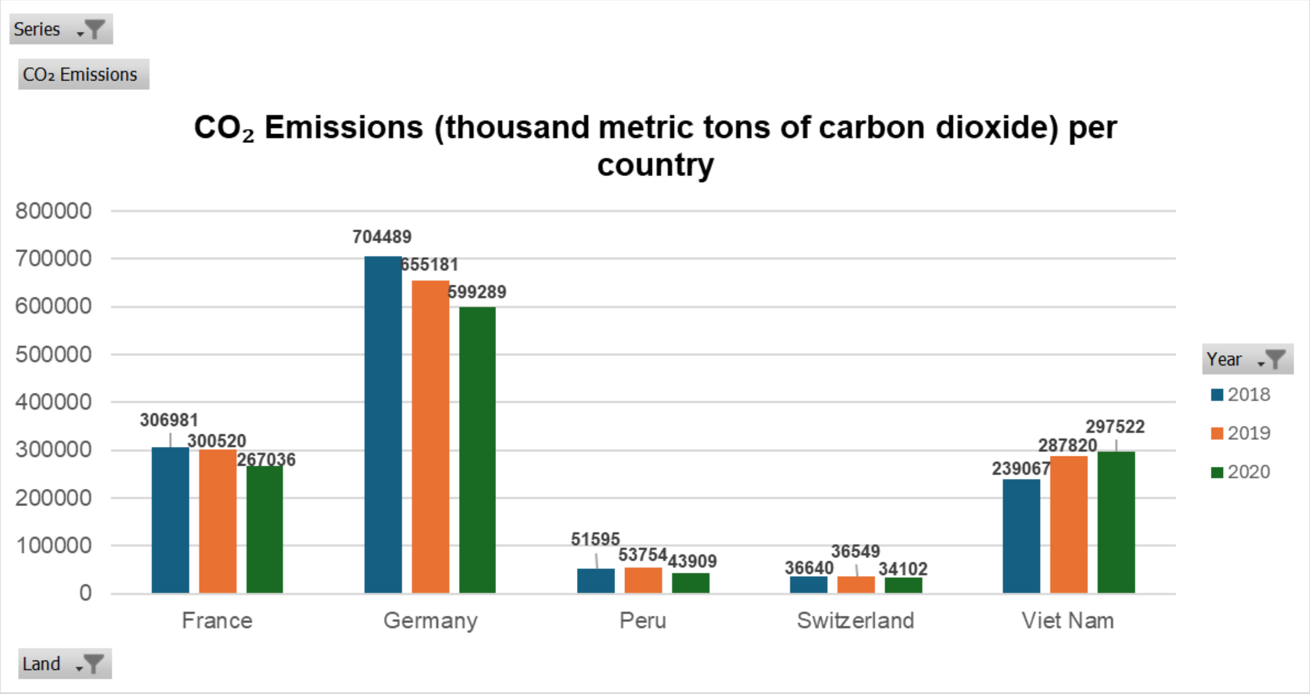


DB Selection



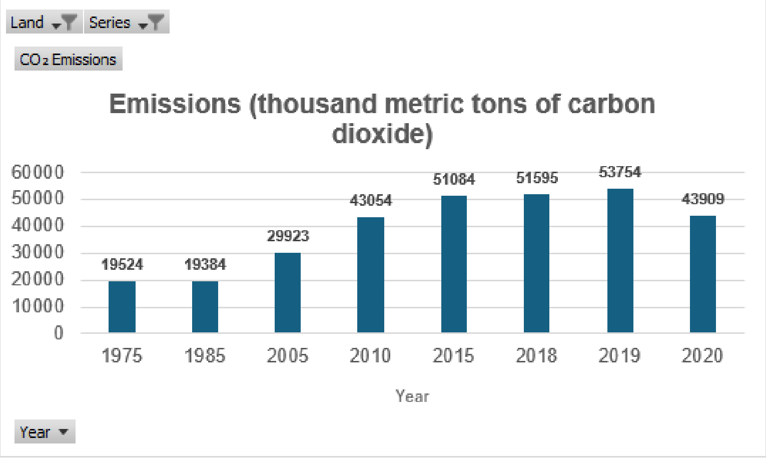
Data successfully loaded without errors.

KPI 01: CHANGE IN CO2 EMISSIONS BETWEEN TWO YEARS PER COUNTRY (%)



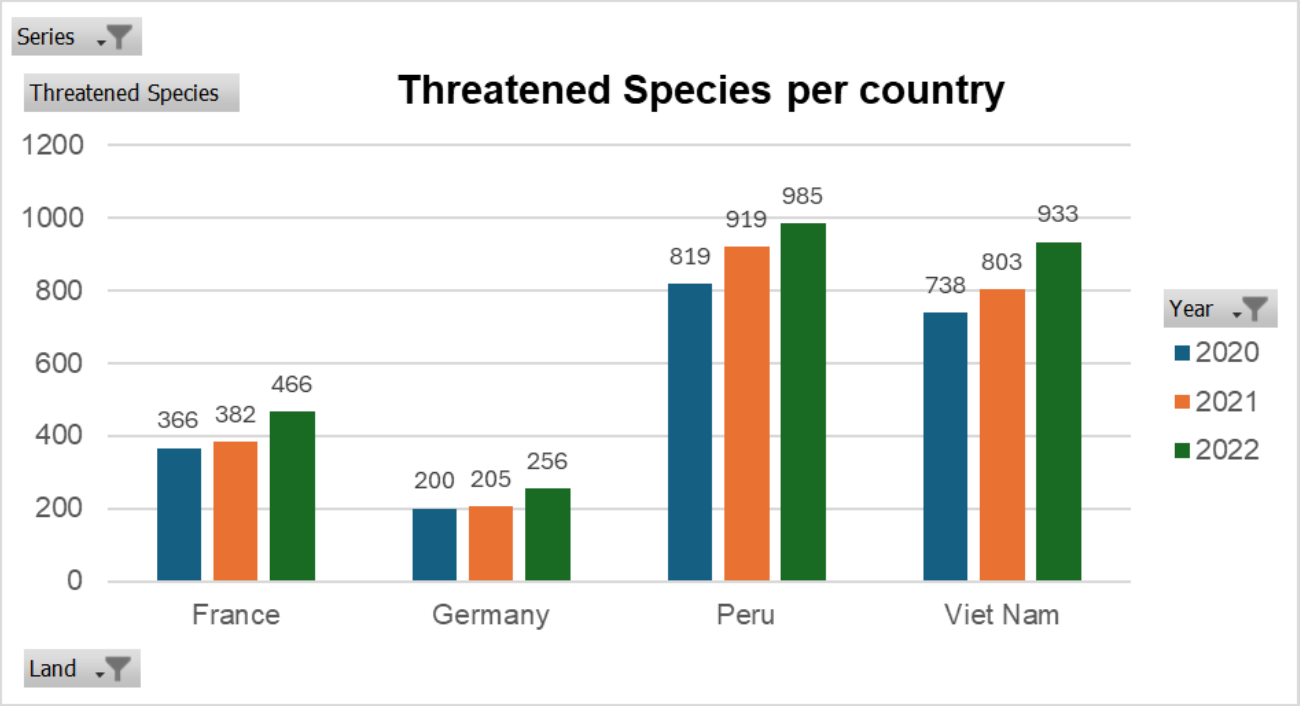
CO_2 Emissions (thousand metric tons of carbon dioxide) in Peru

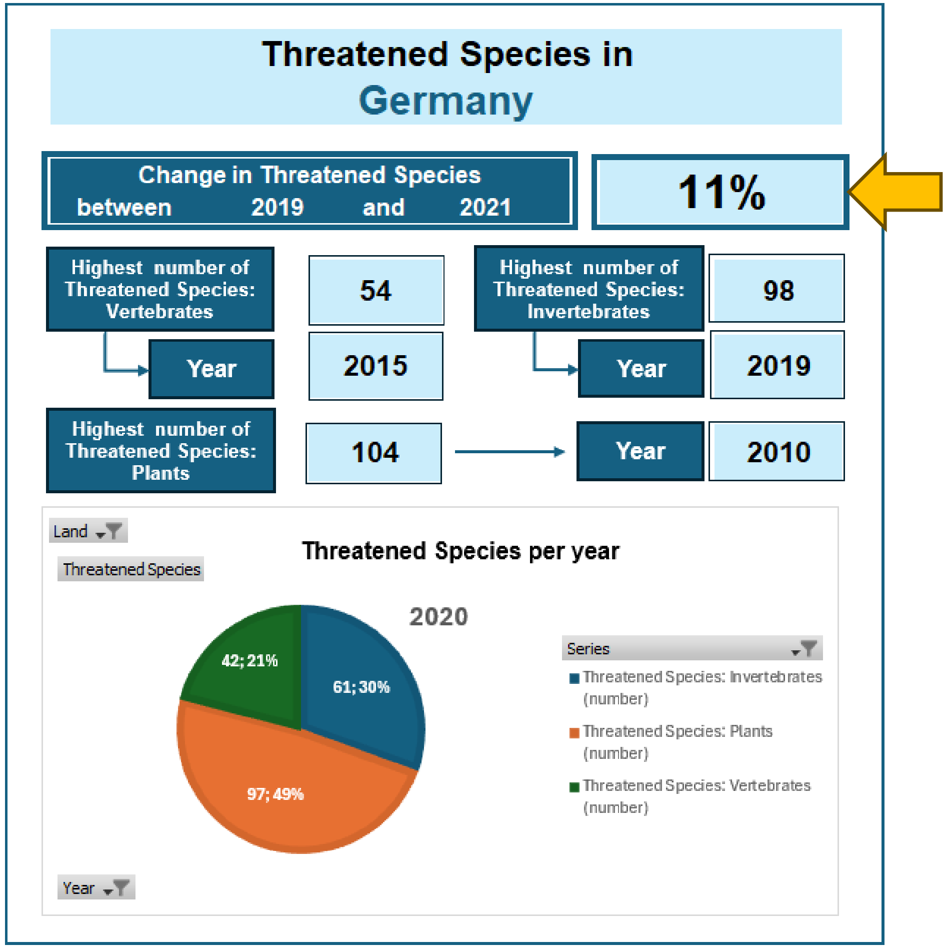
Change in CO_2 emissions between 2019 and 2020	-18%
Average CO_2 Emissions per capita	1.33
Highest CO_2 Emission Amount	53754
Year	2019



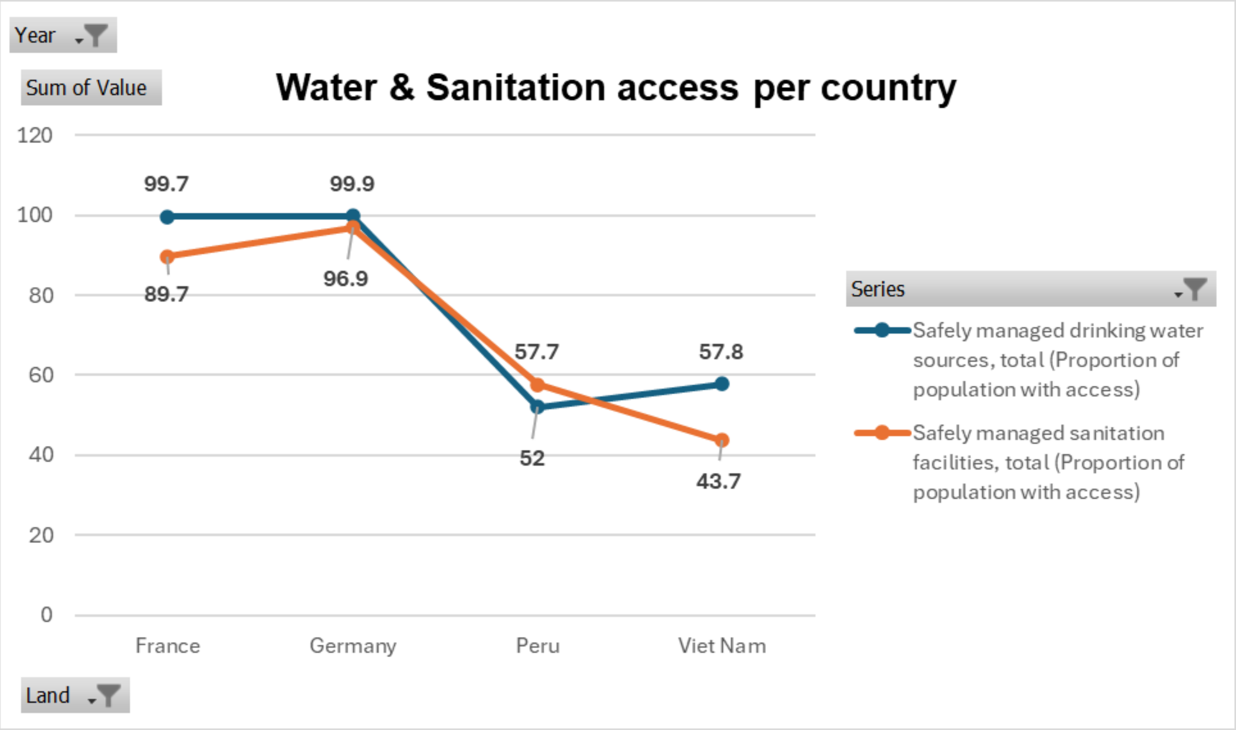
5. Design & Implementation
Excel Prototype

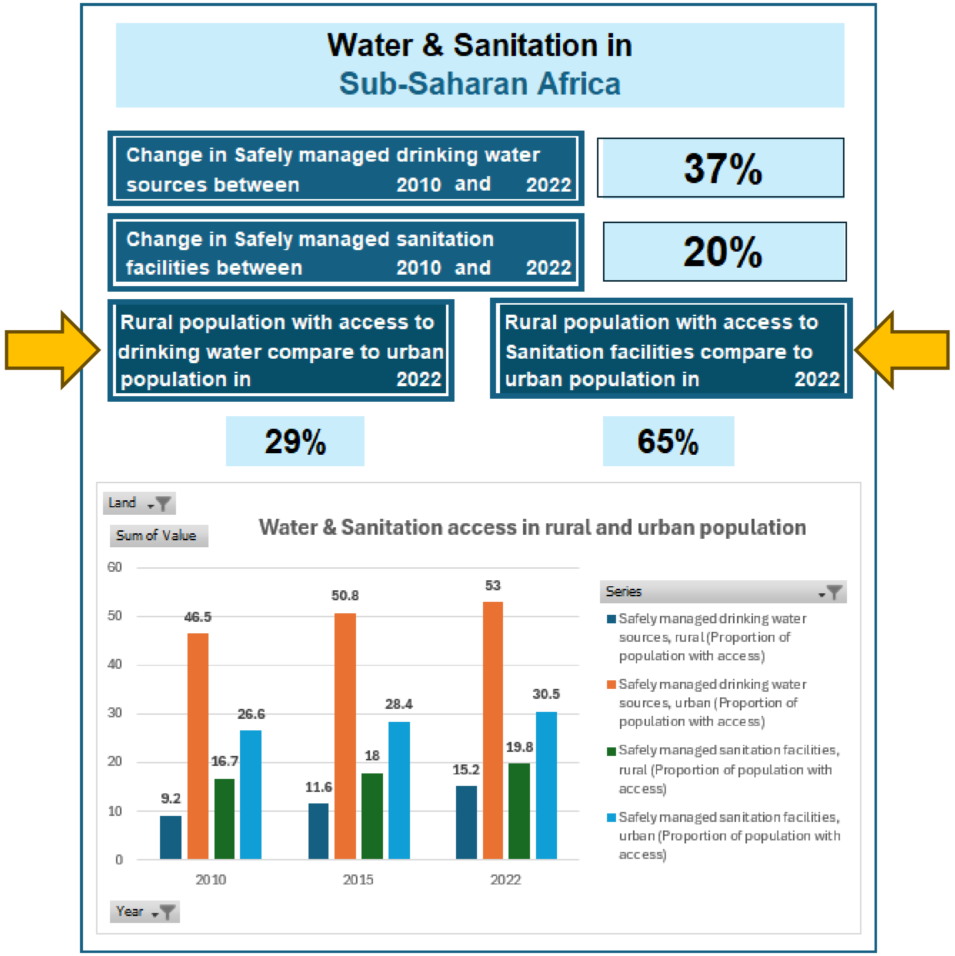
KPI 02: CHANGE IN THREATENED SPECIES BETWEEN TWO YEARS PER REGION (%)



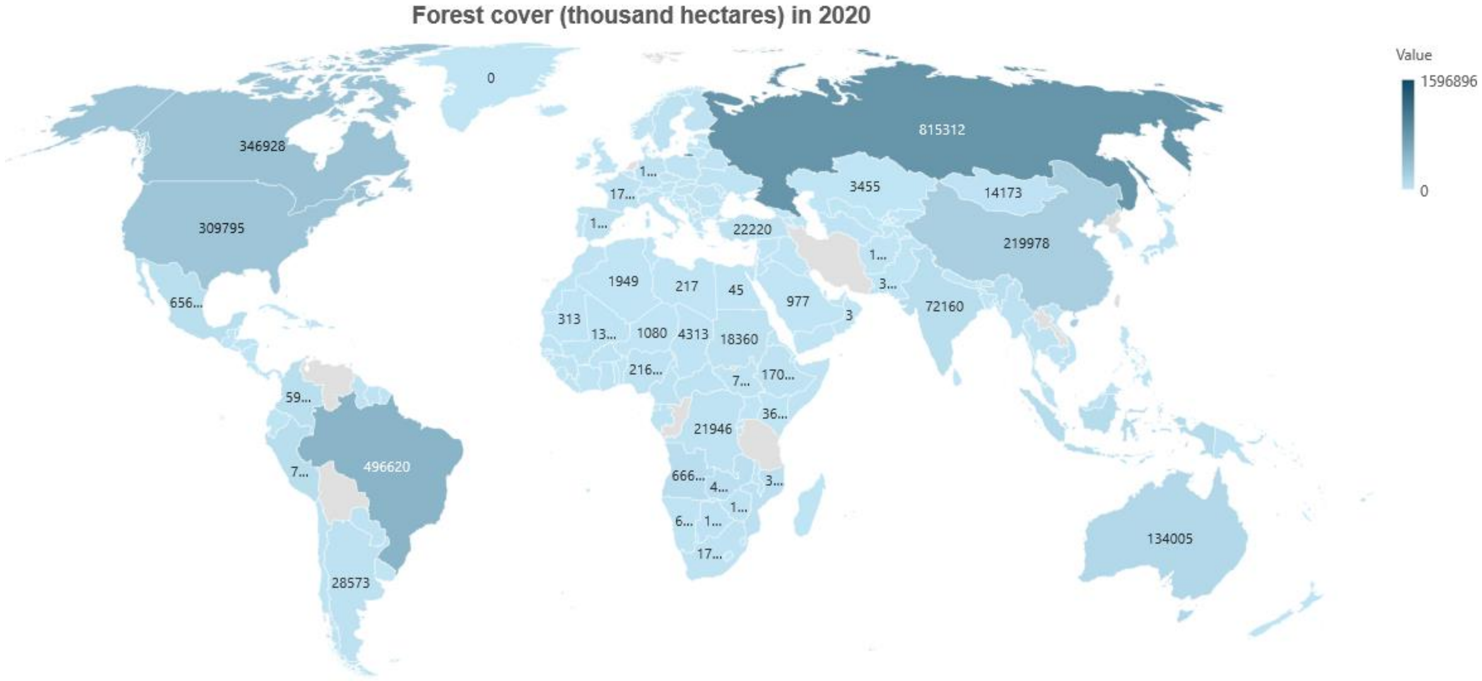


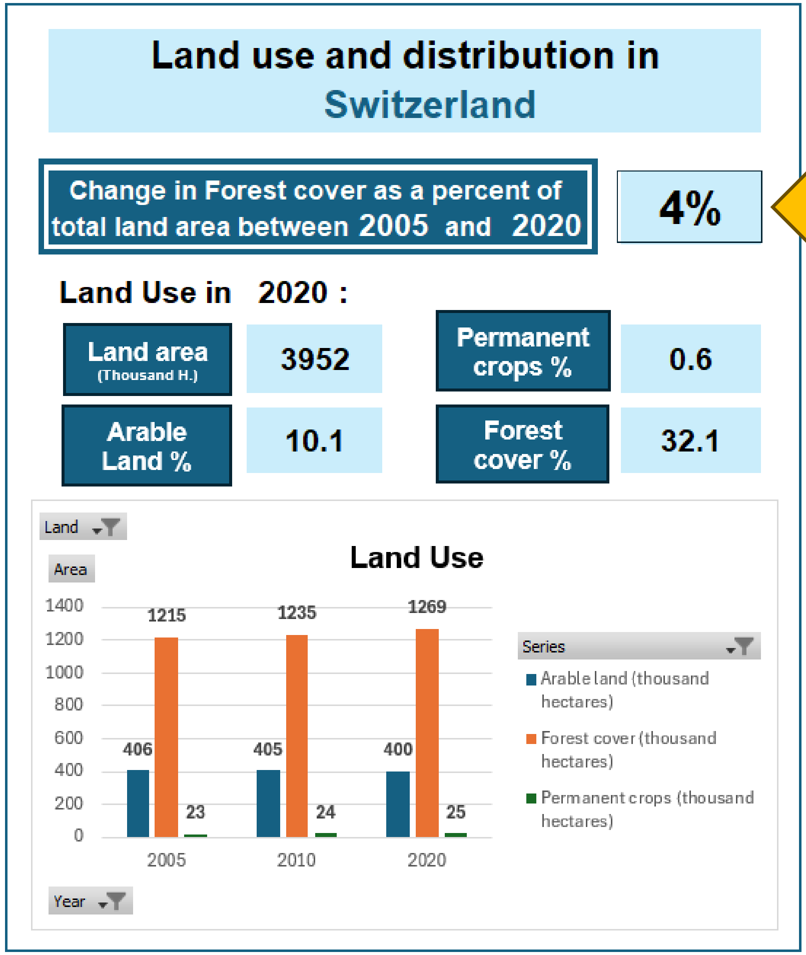
KPI 03: REGIONAL DISPARITIES IN WATER & SANITATION ACCESS (%)





KPI 04: CHANGE IN FOREST COVER AS A PERCENTAGE OF TOTAL LAND AREA BETWEEN TWO YEARS (%)





Roadmap to next steps

Initial Phase

- Project Definition
- Project Planning
- Business Case
- Dataset Selection
- Tools

Progress So Far

- KPI definitions
- Dataset analysis
- Tools setup and integration
- Quality Assurance
- ETL process completed & verified
- Data warehouse design
- Star Schema
- Excel Prototype

Next Steps

- BI Dashboard Development
- Data Insights Analysis
- Testing
- Surprise
- Final Report

5. Design & Implementation

Recap Business Case

What is the Business Case About?

- Development of a Data Warehouse (DW) and Business Intelligence (BI) system to centralize, analyze, and visualize critical global sustainability data.
- Focuses on emissions, water resources, land use, and biodiversity to uncover patterns and support informed decision-making on environmental challenges.

BIA Tool Used: Power BI

Importance of the project:

- To address urgent issues like emissions, water & sanitation, threatened species and land use.
- Provide a centralized platform for data-driven decisions
- Drives actionable insights to address global environmental issues.

Who Benefits?

- **Policymakers:**
  Data-backed decisions for climate policies.
- **Businesses:**
  Sustainable operational strategies.
- **NGOs:**
  Advocacy and monitoring tools for environmental efforts.

How It Serves:

- Identifies global trends and regional disparities.
- Enables targeted interventions for sustainability goals.
- Facilitates collaboration among stakeholders for impactful action.

5. Design & Implementation

Recap KPIs

CHANGE IN CO_2 EMISSIONS PER COUNTRY PER YEAR

This KPI measures the annual percentage change in CO_2 emissions between two years.

Formula

$$\frac{(CO_2 \text{ } emissions \text{ } in \text{ } year \text{ } N - CO_2 \text{ } emissions \text{ } in \text{ } year \text{ } N - 1)}{CO_2 \text{ } emissions \text{ } in \text{ } year \text{ } N - 1} \, x \, 100$$

CHANGE OF THREATENED SPECIES BETWEEN TWO YEARS IN %

This KPI tracks the increase or decrease in the number of threatened species in a given region between two years.

Formula

$$\frac{(threatened \text{ } species \text{ } in \text{ } year \text{ } N - threatened \text{ } species \text{ } in \text{ } year \text{ } N - 1)}{threatened \text{ } species \text{ } in \text{ } year \text{ } N - 1} \, x \, 100$$

REGIONAL DISPARITIES IN WATER & SANITATION ACCESS (%)

This KPI measures the proportion of rural population, having access to water and sanitation facilities compare to urban population.

Formula

$$\frac{proportion \text{ } of \text{ } rural \text{ } population \text{ } with \text{ } access \text{ } to \text{ } facilities}{proportion \text{ } of \text{ } urban \text{ } population \text{ } with \text{ } access \text{ } to \text{ } facilities} \, x \, 100$$

CHANGE IN FOREST COVER AS A PERCENTAGE OF TOTAL LAND AREA BETWEEN TWO YEARS

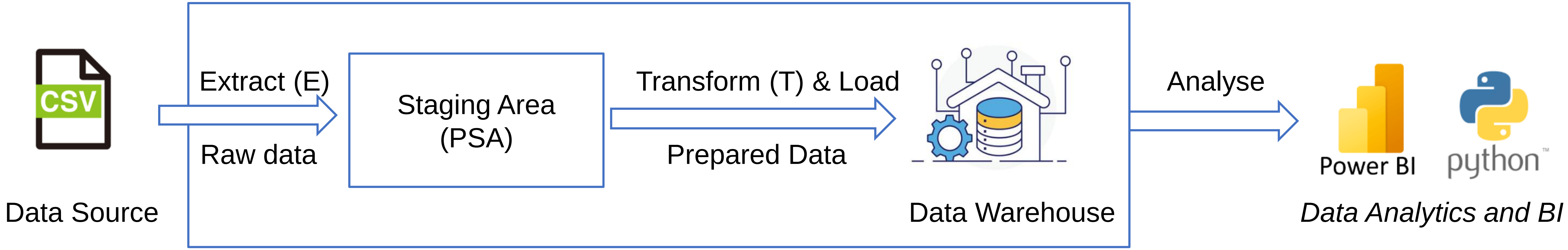
This KPI tracks the change in forest cover between two years, expressed as a percentage of the total land area, to assess deforestation and reforestation trends.

Formula

$$\frac{(forest \text{ } cover \text{ } \% \text{ } in \text{ } year \text{ } N - forest \text{ } cover \text{ } \% \text{ } in \text{ } year \text{ } N - 1)}{forest \text{ } cover \text{ } \% \text{ } in \text{ } year \text{ } N - 1} \, x \, 100$$

5. Design & Implementation

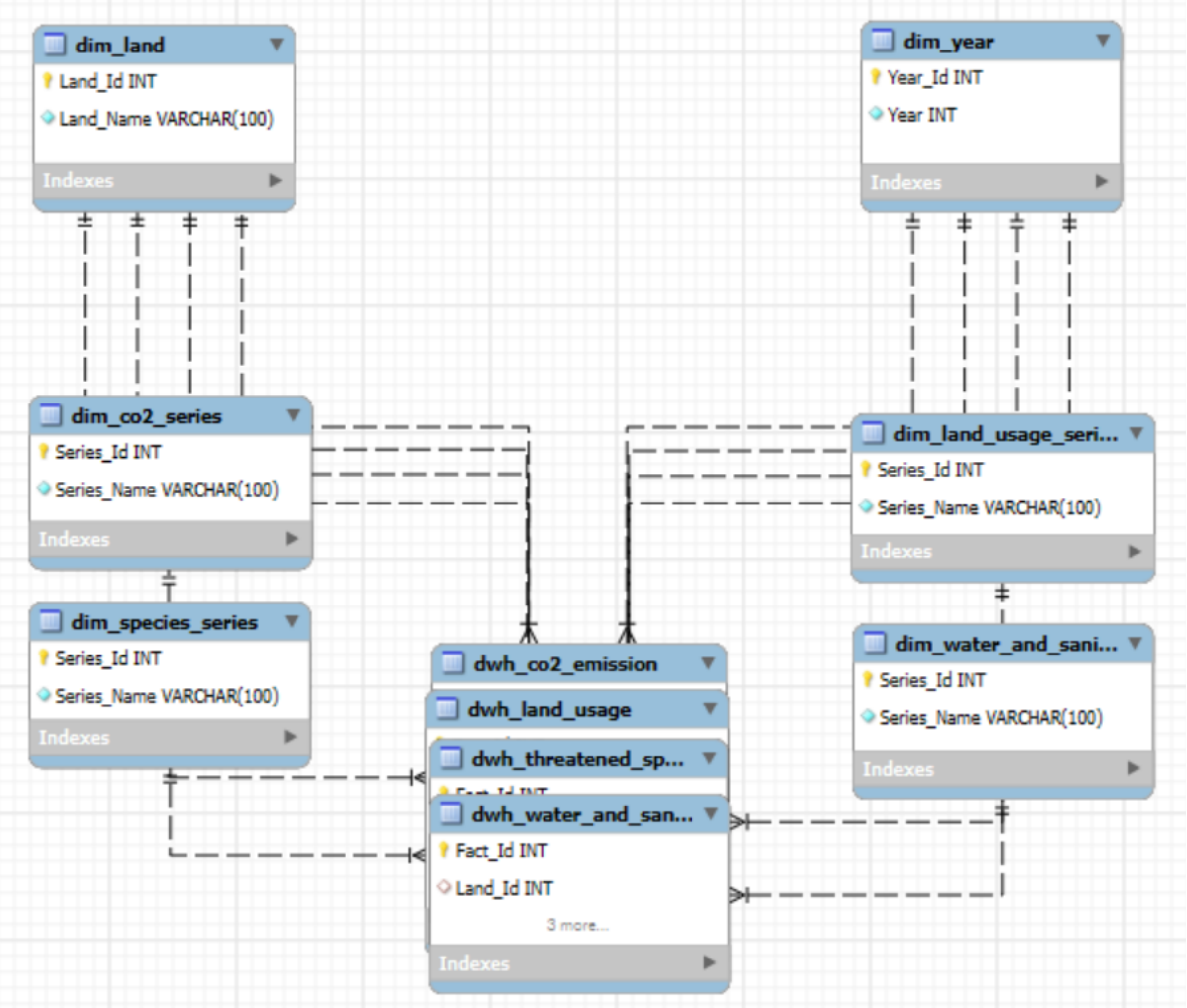
Recap Datawarehouse Design & ETL



ETL Pipeline

5. Design & Implementation

Recap Datawarehouse Design & ETL



Snowflake schema

Fact Tables - 4

dwh_co2_emission
dwh_land_usage
dwh_threatened_species
dwh_water_and_sanitation

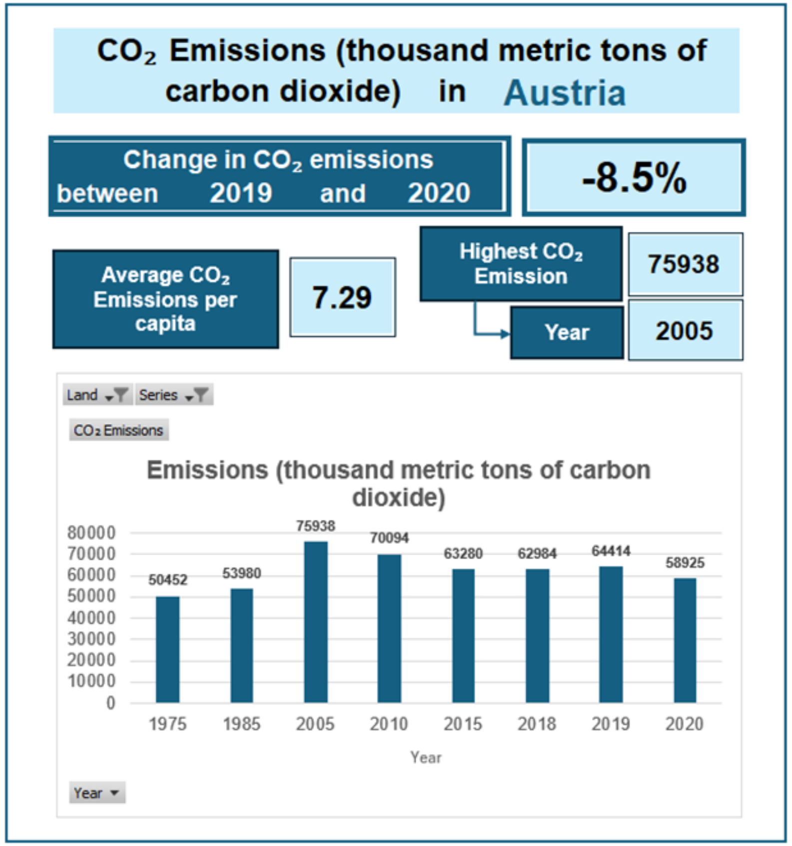
Dimension Tables - 6

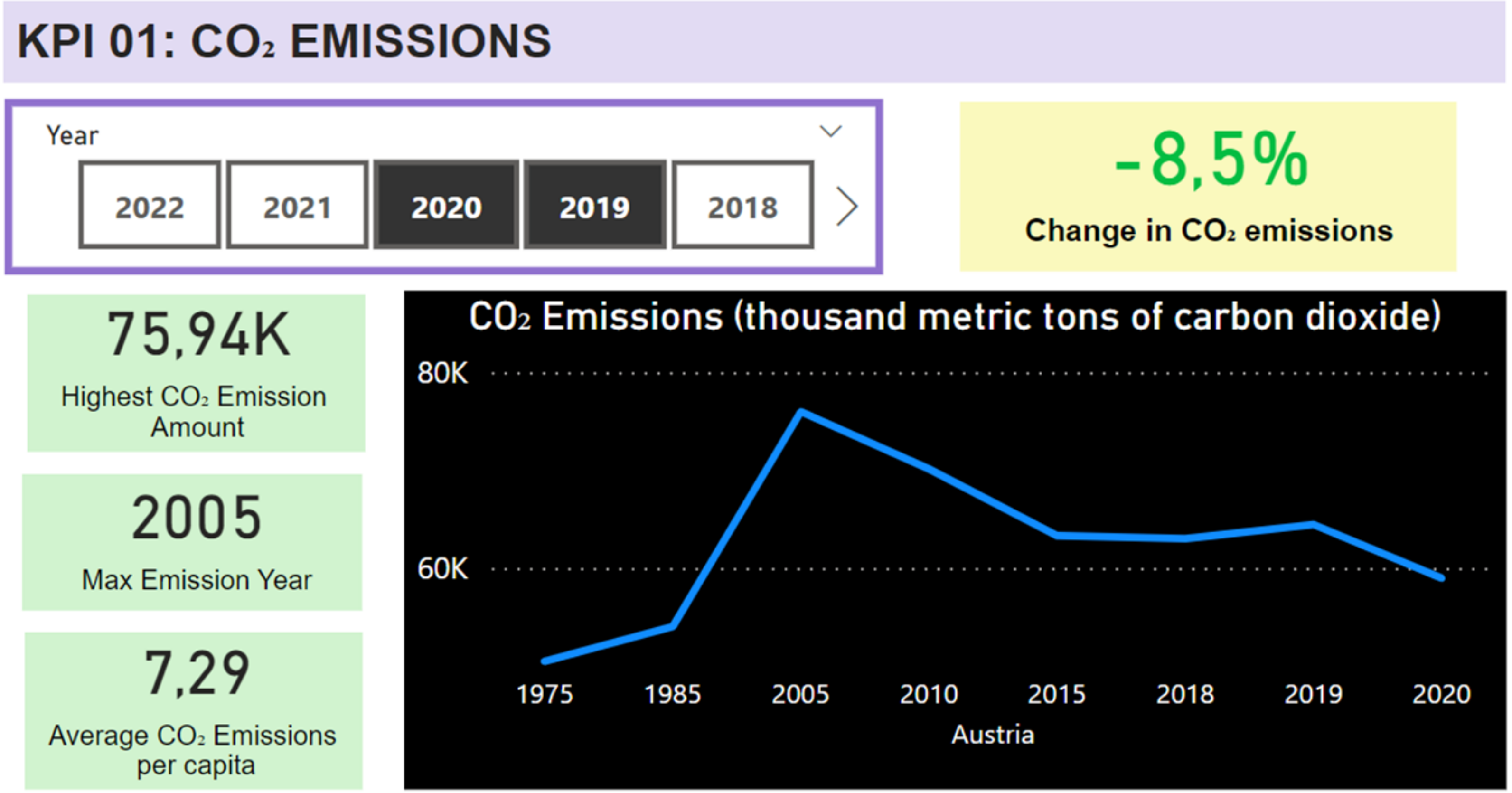
dim_co2_series
dim_land_usage_series
dim_species_series
dim_water_and_sanitation_series
dim_land
dim_year

5. Design & Implementation

Dashboard

5. Design & Implementation
Dashboard





KPI 01: CO_2 EMISSIONS

Change in CO2 emissions

```
Change in CO₂ emissions =
VAR SelectedMinYear = MIN('dwh_environment dim_year'[Year])
VAR SelectedMaxYear = MAX('dwh_environment dim_year'[Year])
VAR EmissionForMaxYear =
    CALCULATE(
        SUM('dwh_environment dwh_co2_emission'[Emission_Value]),
        'dwh_environment dim_year'[Year] = SelectedMaxYear
    )
VAR EmissionForMinYear =
    CALCULATE(
        SUM('dwh_environment dwh_co2_emission'[Emission_Value]),
        'dwh_environment dim_year'[Year] = SelectedMinYear
    )
RETURN
IF(
    EmissionForMinYear > 0,
    (EmissionForMaxYear - EmissionForMinYear) / EmissionForMinYear,
    BLANK()
)
```

Average CO2 emissions per capita

```
AverageEmissionsPerCapitaSeries =
CALCULATE(
    AVERAGE('dwh_environment dwh_co2_emission'[Emission_Value]), -- Calculate the average of Emission_Value
    REMOVEFILTERS('dwh_environment dim_year'), -- Ignore Year slicer or filters
    KEEPFILTERS('dwh_environment dim_land'[Land_Name]), -- Keep Country filter active
    FILTER(
        ALL('dwh_environment dim_co2_series'), -- Ensure no external filters affect the Series table
        'dwh_environment dim_co2_series'[Series_Name] = "Emissions per capita (metric tons of carbon dioxide)" -- Filter for the specific Series_Name
    )
)
```

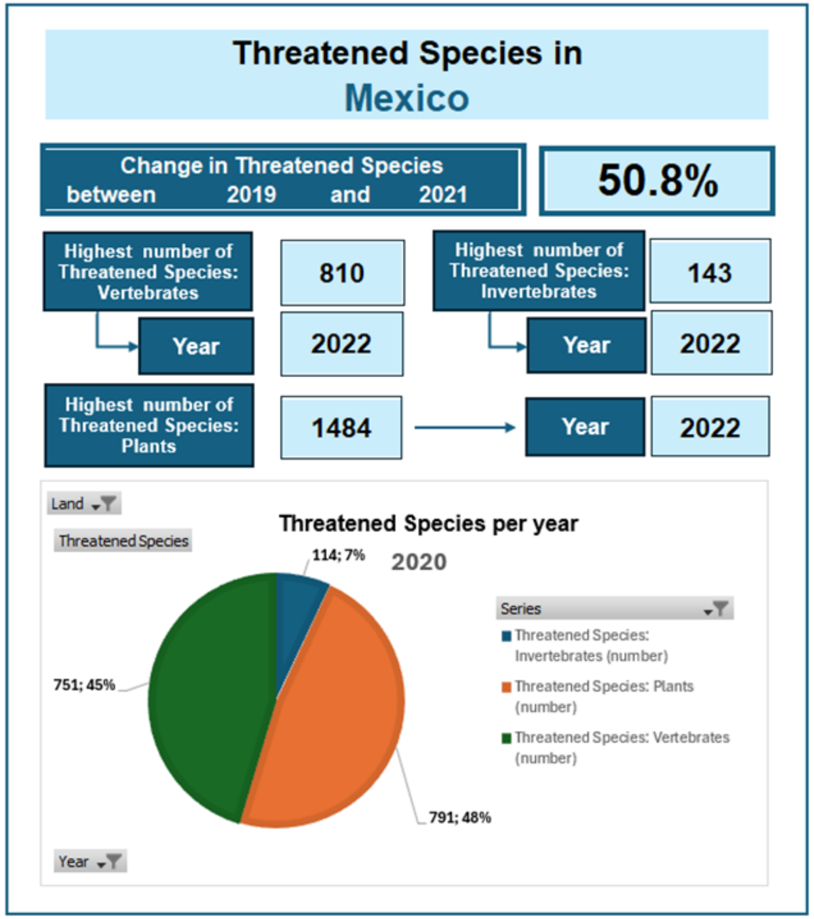
Highest CO2 Emission Amount

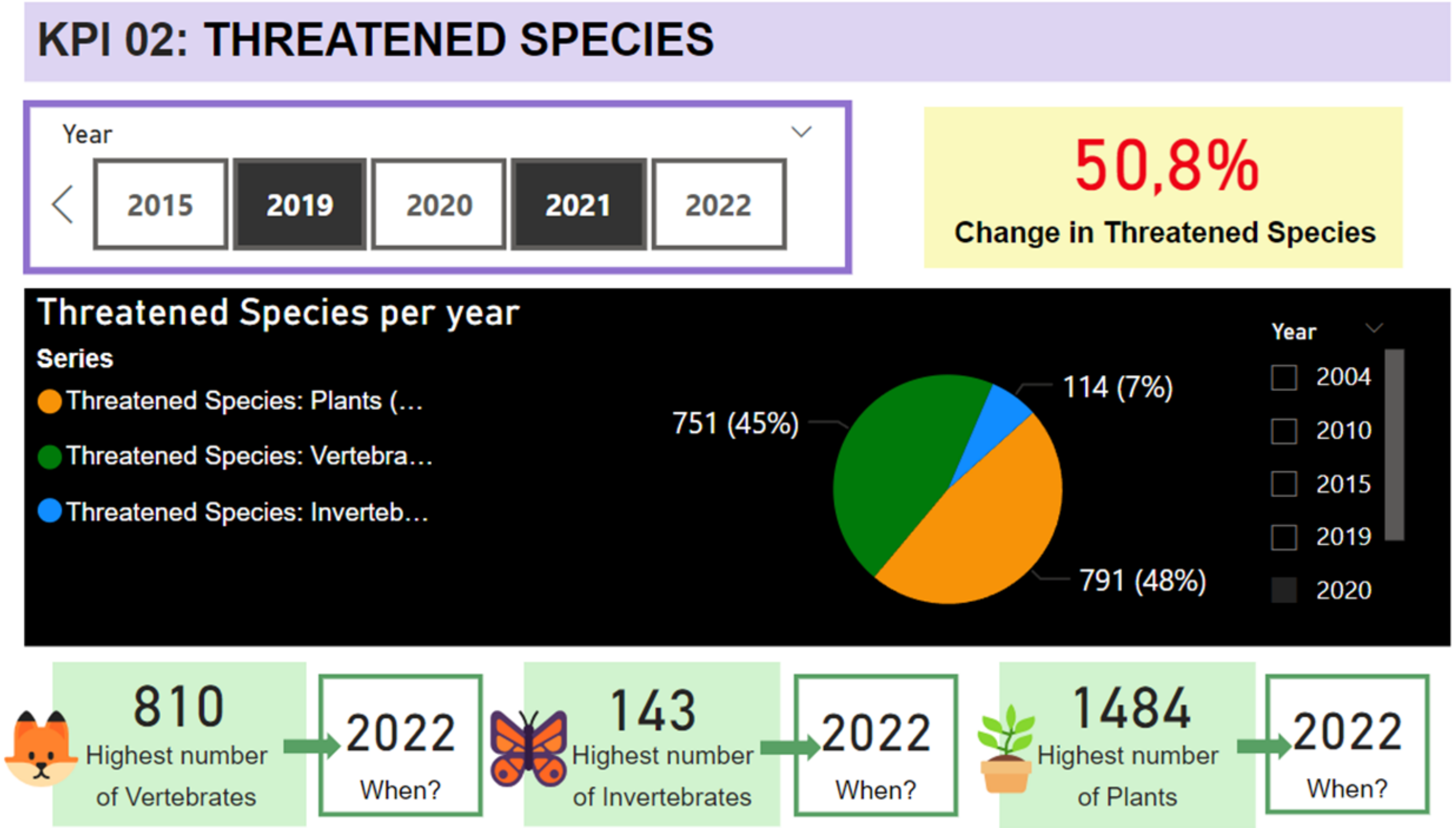
```
Highest CO₂ Emission Amount =
CALCULATE(
    MAX('dwh_environment dwh_co2_emission'[Emission_Value]),
    ALL('dwh_environment dim_year'[Year])
)
```

Year of the highest CO2 Emission

```
Max Emission Year =
VAR MaxEmissionValue =
    CALCULATE(
        MAX('dwh_environment dwh_co2_emission'[Emission_Value]), -- Find maximum emission value
        REMOVEFILTERS('dwh_environment dim_year') -- Ignore Year slicer
    )
RETURN
CALCULATE(
    MAX('dwh_environment dim_year'[Year]), -- Retrieve the year corresponding to the maximum emission value
    FILTER(
        ALL('dwh_environment dwh_co2_emission'), -- Consider all rows in the table
        'dwh_environment dwh_co2_emission'[Emission_Value] = MaxEmissionValue -- Match rows with maximum emission value
    )
)
```

5. Design & Implementation
Dashboard





5. Design & Implementation

Dashboard

Data Analysis Expressions (DAX) in Power BI

KPI 02: THREATENED SPECIES

Change in Threatened Species

```
Change in Threatened Species =
VAR SelectedMinYear = MIN('dwh_environment dim_year'[Year])
VAR SelectedMaxYear = MAX('dwh_environment dim_year'[Year])
VAR EmissionForMaxYear =
    CALCULATE(
        SUM('dwh_environment dwh_threatened_species'[Threatened_Species_Count]),
        'dwh_environment dim_year'[Year] = SelectedMaxYear
    )
VAR EmissionForMinYear =
    CALCULATE(
        SUM('dwh_environment dwh_threatened_species'[Threatened_Species_Count]),
        'dwh_environment dim_year'[Year] = SelectedMinYear
    )
RETURN
IF(
    EmissionForMinYear > 0,
    (EmissionForMaxYear - EmissionForMinYear) / EmissionForMinYear,
    BLANK()
)
```

Highest number of Threatened Species

```
Highest TS =
CALCULATE(
    MAX('dwh_environment dwh_threatened_species'[Threatened_Species_Count]),
    ALL('dwh_environment dim_year'[Year])
)
```

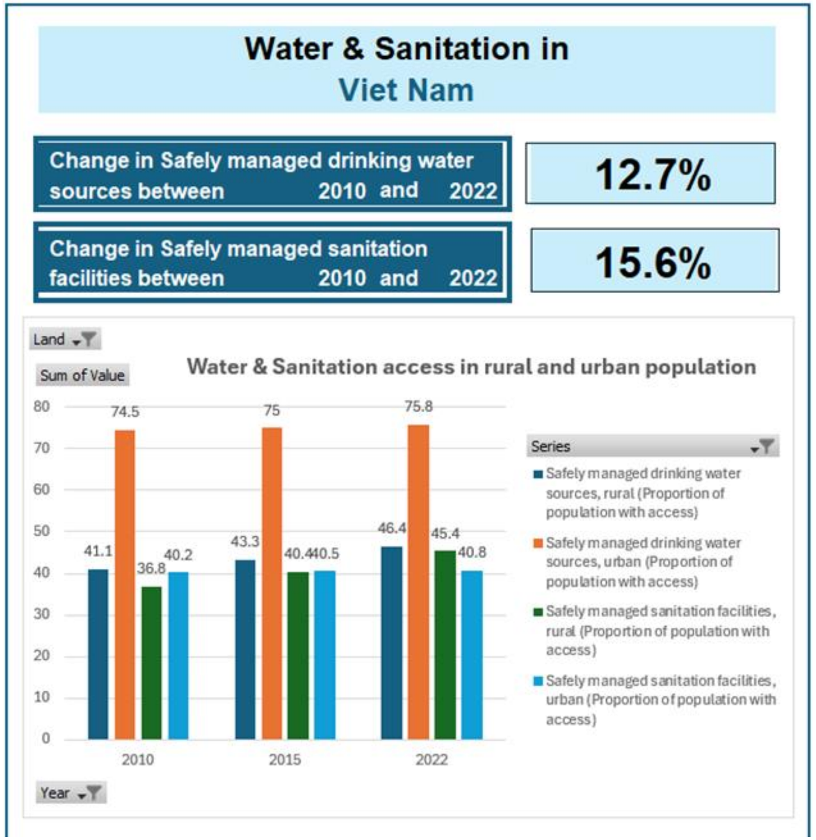
Using this formula and filtering you can obtain highest number of Vertebrates, Invertebrates and Plants

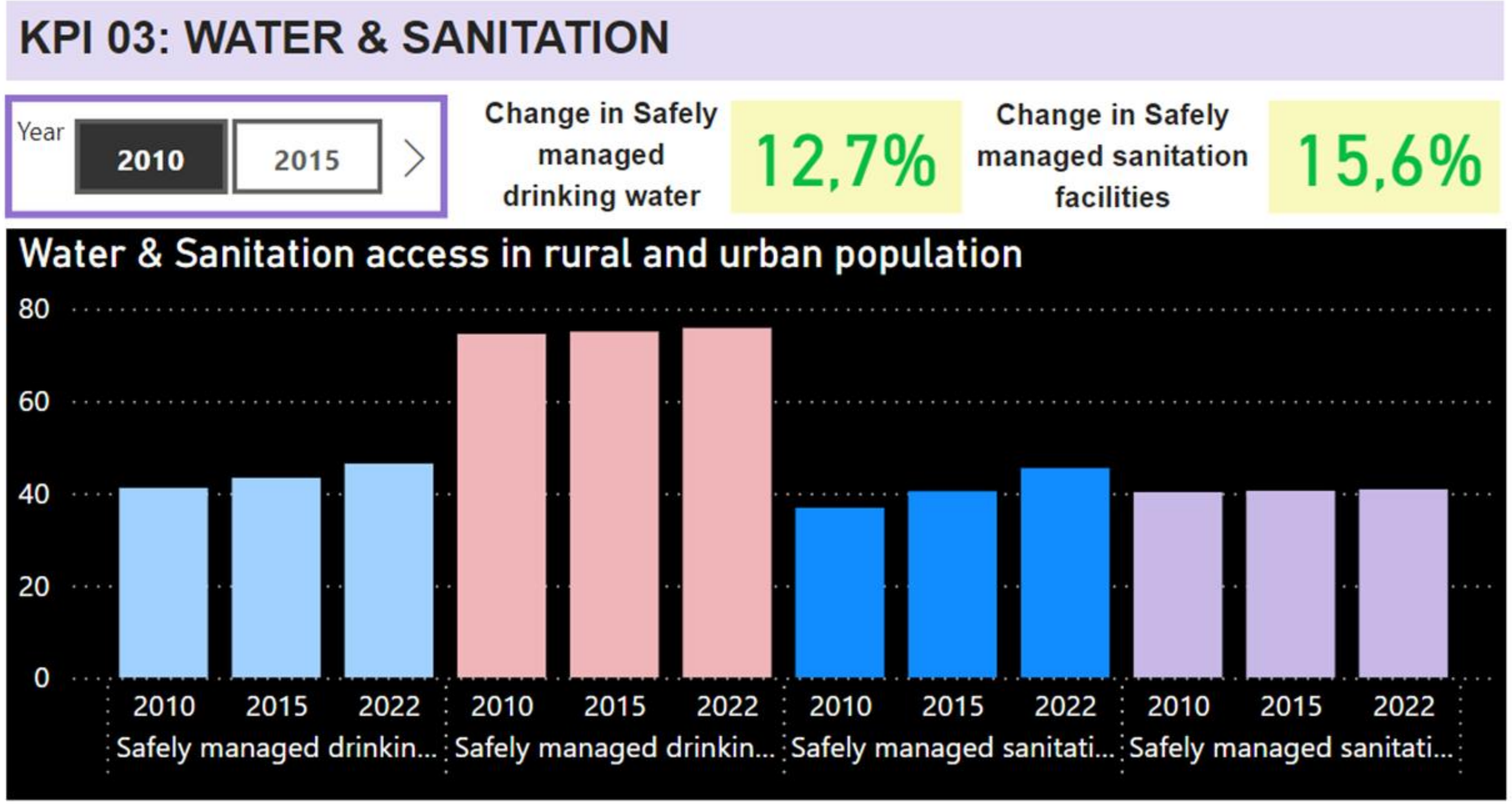
Year of the highest number of Threatened Species

```
Year of Max Threatened Species Count =
VAR MaxTSCount =
    CALCULATE(
        MAX('dwh_environment dwh_threatened_species'[Threatened_Species_Count]), -- Find maximum count
        REMOVEFILTERS('dwh_environment dim_year') -- Ignore Year slicer
    )
RETURN
CALCULATE(
    MAX('dwh_environment dim_year'[Year]), -- Retrieve the year corresponding to the maximum count
    FILTER(
        ALL('dwh_environment dwh_threatened_species'), -- Consider all rows in the table
        'dwh_environment dwh_threatened_species'[Threatened_Species_Count] = MaxTSCount -- Match rows with maximum count
    )
)
```

Using this formula and filtering you can obtain the year of the highest number of Vertebrates, Invertebrates and Plants

5. Design & Implementation
Dashboard



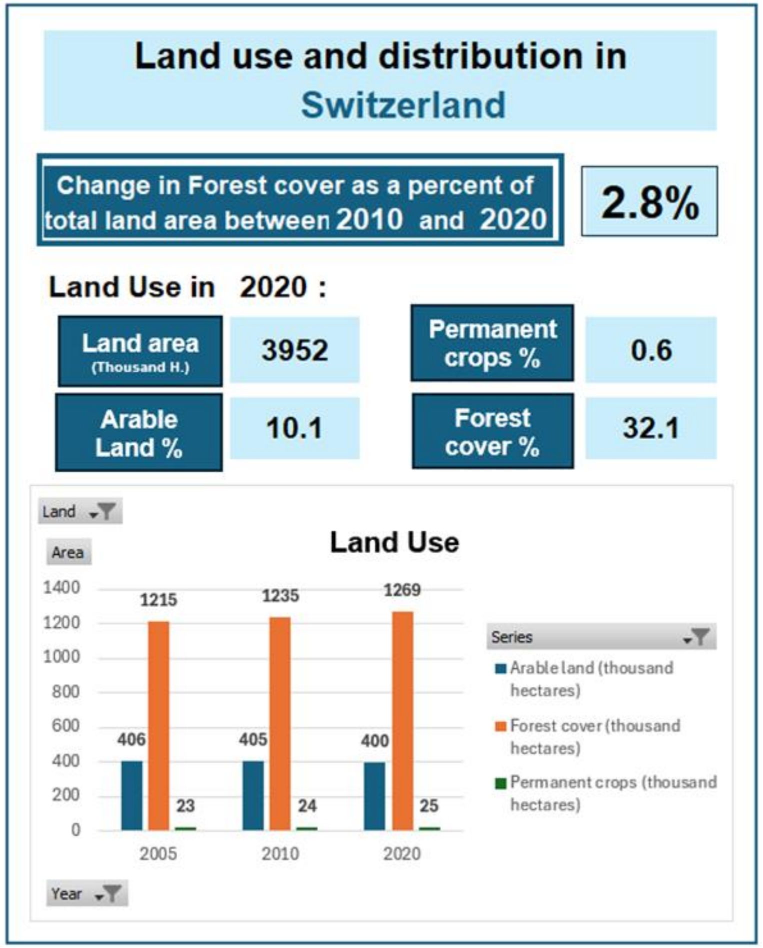


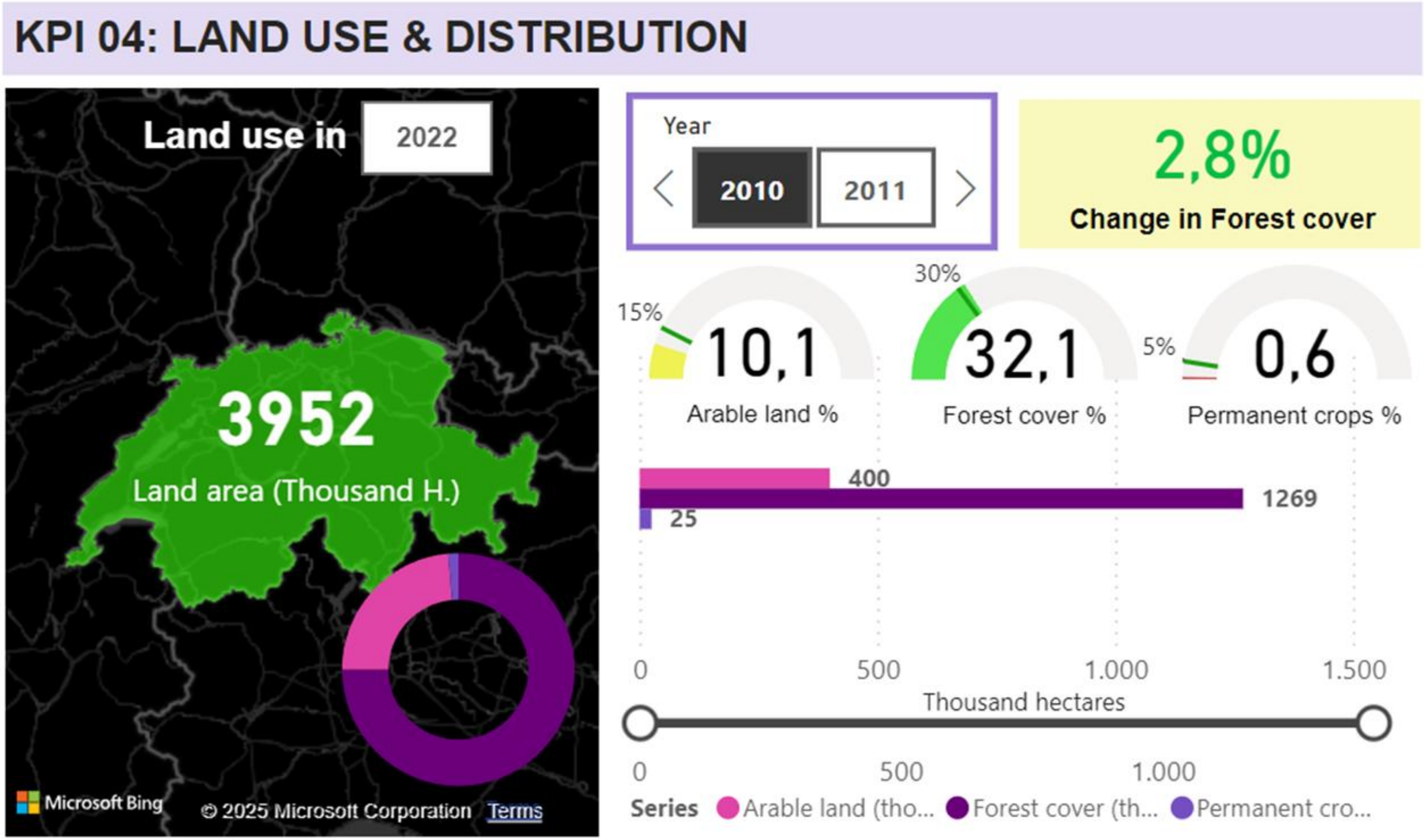
KPI 03: WATER & SANITATION

Change in Safely managed drinking water or sanitation facilities

```
Change in Water and sanitation =
VAR SelectedMinYear = MIN('dwh_environment dim_year'[Year])
VAR SelectedMaxYear = MAX('dwh_environment dim_year'[Year])
VAR WaterSanitationMaxYear =
    CALCULATE(
        SUM('dwh_environment dwh_water_and_sanitation'[Service_Coverage_Percentage]),
        'dwh_environment dim_year'[Year] = SelectedMaxYear
    )
VAR WaterSanitationForMinYear =
    CALCULATE(
        SUM('dwh_environment dwh_water_and_sanitation'[Service_Coverage_Percentage]),
        'dwh_environment dim_year'[Year] = SelectedMinYear
    )
RETURN
IF(
    WaterSanitationForMinYear > 0,
    (WaterSanitationMaxYear - WaterSanitationForMinYear) / WaterSanitationForMinYear,
    BLANK()
)
```

5. Design & Implementation
Dashboard





KPI 04: LAND USE & DISTRIBUTION

Change in Forest cover as e percent of total land

```
Change in Forest cover =
VAR SelectedMinYear = MIN('dwh_environment dim_year'[Year])
VAR SelectedMaxYear = MAX('dwh_environment dim_year'[Year])
VAR ForestCoverForMaxYear =
    CALCULATE(
        SUM('dwh_environment dwh_land_usage'[Land_Metric_Value]),
        'dwh_environment dim_year'[Year] = SelectedMaxYear
    )
VAR ForestCoverForMinYear =
    CALCULATE(
        SUM('dwh_environment dwh_land_usage'[Land_Metric_Value]),
        'dwh_environment dim_year'[Year] = SelectedMinYear
    )
RETURN
IF(
    ForestCoverForMinYear > 0,
    (ForestCoverForMaxYear - ForestCoverForMinYear) / ForestCoverForMinYear,
    BLANK()
)
```

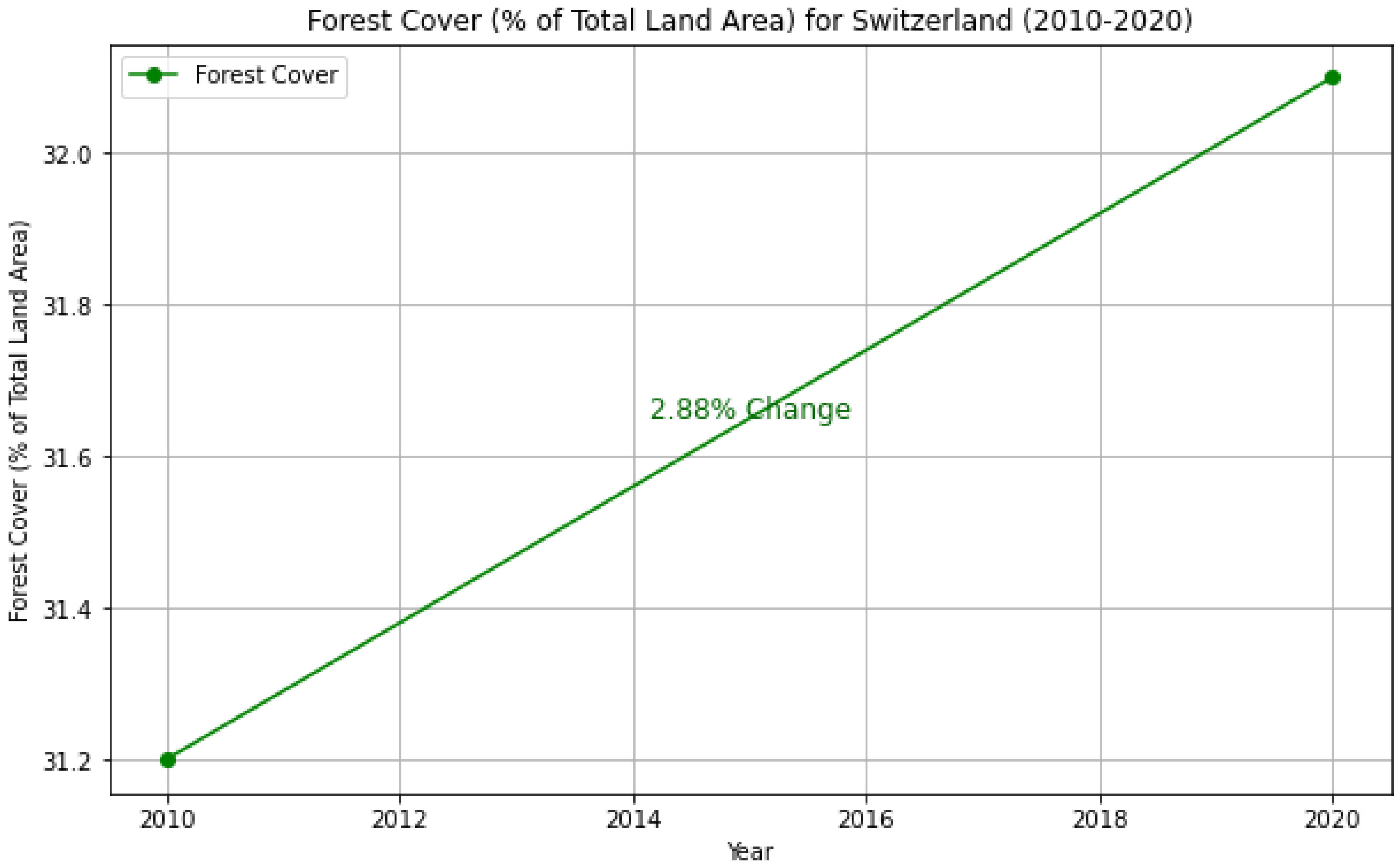
Land Use and Distribution

```
Land_Use_and_Distribution =
CALCULATE(
    SUM('dwh_environment dwh_land_usage'[Land_Metric_Value]),
    FILTER(
        'dwh_environment dim_land',
        'dwh_environment dim_land'[Land_Name] IN VALUES('dwh_environment dim_land'[Land_Name])
    ),
    FILTER(
        'dwh_environment dim_year',
        'dwh_environment dim_year'[Year] IN VALUES('dwh_environment dim_year'[Year])
    )
)
```

Using this formula and filtering you can obtain Arable Land%, Permanent crops% and Forest cover%

5. Design & Implementation
Testing data accuracy with Python

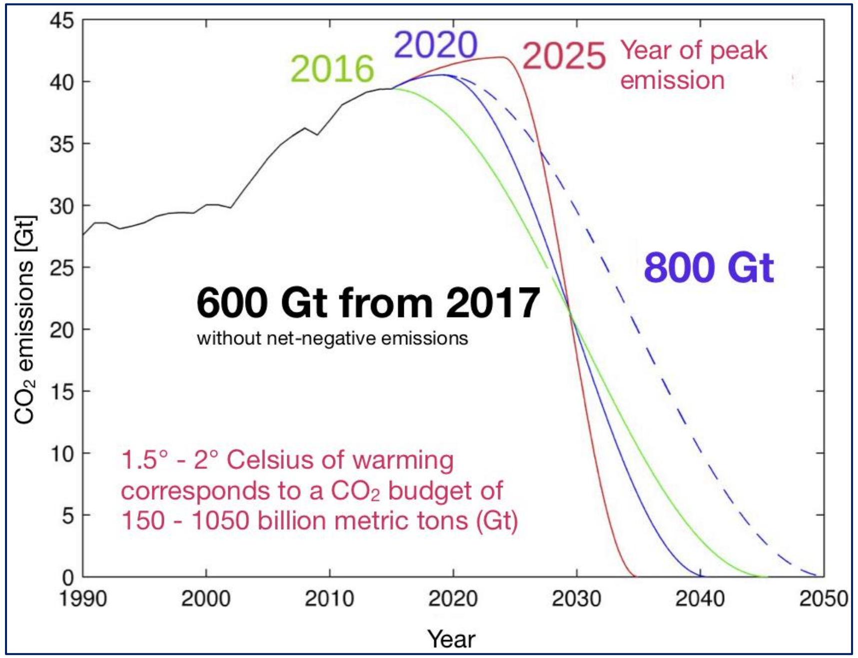
- Use of Python for data accuracy testing and visualization to showcase the robustness of the analysis.
- Forest cover increased by **2.8%** during this period as shown on Power BI dashboard and also on Excel Prototype.
- Result can be trusted.

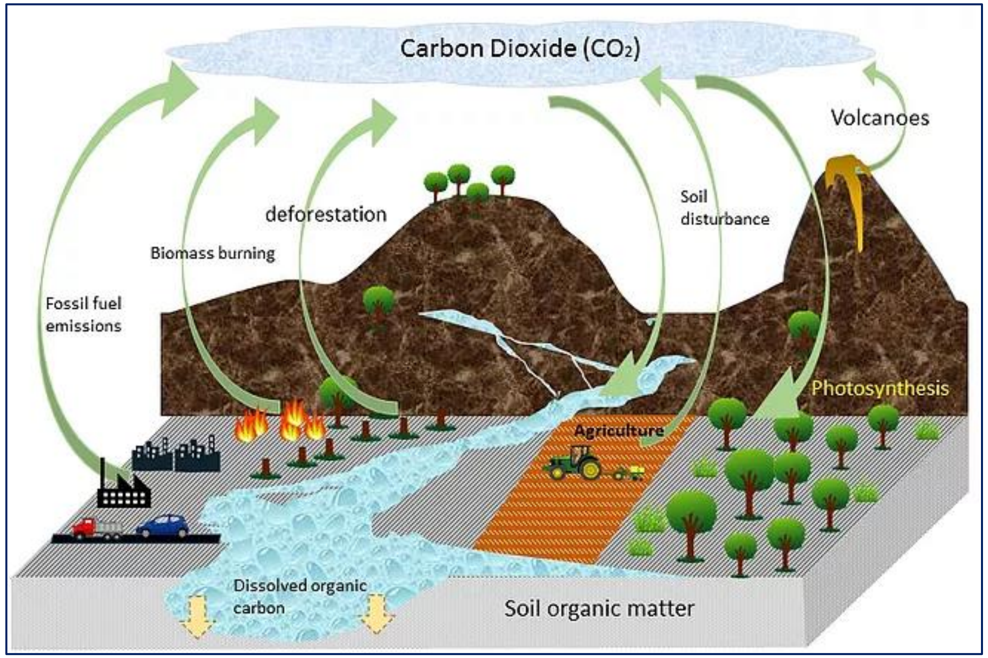


Key Findings

CO_2 Emissions: Context

The Paris Agreement | UNFCCC





Carbon Budget | Net Zero | Co2 vs. GDP | Co2 by Sector

Image Source: https://winrock.org/greenhouse-gas-emissions-from-land-use-and-land-use-change/

Key Findings

CO_2 Emissions: Trends (2020)

Top 5 countries with highest levels of annual Co2 emissions in 2020

COUNTRY	KILOTONS OF CO2
China	10,190,639
United States of America	4,324,698
India	2,135,190
Russian Federation	1,562,930
Japan	998,658

Top 5 countries with highest levels of annual Co2 emissions per Capita in 2020

COUNTRY	KILOTONS OF CO2
Qatar	29.2
Brunei Darussalam	21.1
Kuwait	21
Gibraltar	19
Bahrain	18.9

5. Design & Implementation

Key Findings

CO_2 Emissions: Trends (2019-2020)

Top 5 countries with highest growth in annual CO2 emissions between 2019 - 2020

COUNTRY	KILOTONS OF CO2
China	148722
Viet Nam	9702
Kazakhstan	6948
United Arab Emirates	2779
Brunei Darussalam	2547

Top 5 countries with highest growth in annual CO2 emissions per Capita between 2019 - 2020

COUNTRY	KILOTONS OF CO2
Brunei Darussalam	5.7
Kazakhstan	0.2
United Arab Emirates	0.1
China	0.1
Kosovo	0.1

Top 5 countries with highest % growth in annual CO2 emissions between 2019 - 2020

COUNTRY	% GROWTH
Brunei Darussalam	37.97%
Nepal	15.12%
Armenia	7.99%
Ghana	7.35%
Benin	6.94%

Top 5 countries with highest % growth in annual CO2 emissions per Capita between 2019 - 2020

COUNTRY	% GROWTH
Brunei Darussalam	37.01%
Tajikistan	14.29%
Armenia	5.00%
Viet Nam	3.45%
Kosovo	2.08%

5. Design & Implementation

Key Findings

CO_2 Emissions: Trends (1975-2020)

Top 5 countries with highest increase in annual CO2 emissions between 1975 - 2020

COUNTRY	KILOTONS OF CO2
China	9,076,155
India	1,891,457
Iran (Islamic Republic of)	504,264
Indonesia	496,516
Republic of Korea	471,306

Top 5 countries with highest increase in annual CO2 emissions per Capita between 2019 - 2020

COUNTRY	KILOTONS OF CO2
Gibraltar	16.9
Brunei Darussalam	12.4
Oman	11.8
Saudi Arabia	10.9
United Arab Emirates	9.3

Top 5 countries with highest % increase in annual CO2 emissions between 2019 - 2020

COUNTRY	% INCREASE
Oman	8702.05%
United Arab Emirates	3566.06%
Saudi Arabia	2045.64%
Qatar	1608.92%
Viet Nam	1405.98%

Top 5 countries with highest % increase in annual CO2 emissions per Capita between 2019 - 2020

COUNTRY	% INCREASE
Ukraine	3600.00%
Estonia	2550.00%
Kosovo	1533.33%
Oman	1475.00%
Serbia	1000.00%

5. Design & Implementation

Key Findings

CO_2 Emissions: Trends (2015-2020)

Top 5 countries with highest reduction in annual CO2 emissions between 2015 - 2020

COUNTRY	KILOTONS OF CO2
United States of America	- 679815
Japan	- 163773
Germany	- 140601
Mexico	- 93259
United Kingdom	- 91815

Top 5 countries with highest reduction in annual CO2 emissions per Capita between 2015 - 2020

COUNTRY	KILOTONS OF CO2
Curaçao	- 23
Estonia	- 5.5
Trinidad and Tobago	- 4.4
Luxembourg	- 3.7
Bahrain	- 3

Top 5 countries with highest % decrease in annual CO2 emissions between 2015 - 2020

COUNTRY	% REDUCTION
Curaçao	- 64.42%
Venezuela	- 61.28%
Estonia	- 49.00%
Iceland	- 29.47%
Trinidad and Tobago	- 27.75%

Top 5 countries with highest % decrease in CO2 emissions per Capita between 2015 - 2020

COUNTRY	% REDUCTION
Curaçao	- 63.89%
Venezuela	- 59.57%
Estonia	- 50.93%
Angola	- 37.50%
Zimbabwe	- 37.50%

6. Conclusion and Lessons Learned

Conclusion

What went well?

- Strong teamwork and mutual support.
- Subhashri's expertise in MySQL configuration & dataset loading into Power BI.

What went wrong?

- UN dataset challenges, limited column count and missing data.
- Spent time on forecasting CO_2 emissions till 2030, then realized the dataset is not suitable. Limited points with 8 years/country.
- Challenges with MySQL configuration and loading in BIA tool.
- We used 4 star schema, one per each of datasets.

How we solved these issues?

- Filled missing data with 0 to maintain data consistency.
- Excluded datasets not fit for forecasting.
- Reconfigured MySQL settings & optimized queries.
- Iteratively tested and resolved tool integration issues.
- Modified the schema to a snowflake schema, halfway to the project.

6. Conclusion and Lessons Learned

Lessons Learned

Lessons learned from challenges:

- Always check data quality to avoid errors.
- Evaluate datasets thoroughly before use.
- Early testing of tools and configurations prevents delays.

Benefits of these lessons…

- Able to conduct data quality assurance for the datasets.
- Build dashboards and warehouses efficiently.
- Ensure datasets are suitable for forecasting to avoid inaccuracies.

VIELEN DANK!