Big Data

Impact of Climate Change Indicators on Global Trends

Table of Contents

1. Introduction

Project Overview:

- This project explores the impact of climate change indicators such as temperature trends and CO2 emissions on global patterns using the various skills learned in Big Data classes.

- By leveraging modern data architectures like Lakehouse and visualization tools like Power BI, we aim to provide actionable insights into climate trends.

Objective:

- To analyze climate change data using scalable data platforms and modern visualization frameworks to highlight actionable global trends and patterns.

2. Project Steps and Tools Used

Steps:

1. Data Ingestion: Import data from open sources such as Global Carbon Atlas, NOAA, and NASA.

2. Data Transformation: Use Databricks to clean, transform, and model data for analysis.

3. Data Visualization: Build insightful dashboards using Power BI.

Tools:

- Platforms: Databricks Community, Microsoft Fabric, and Power BI.

- Frameworks: PySpark for data processing and Azure for storage.

3. Sources and Data Collection

Data Sources:

1. NOAA: Weather and climate data.

2. NASA Earth Data: Atmospheric and temperature trends.

3. Global Carbon Atlas: CO2 emissions by sector and region.

Data Collection Process:

- Data was ingested into the Databricks platform using automated pipelines that connected directly to the sources.

1. Territorial Emissions in $MtCO_2$ per Year

- This dataset provides information on the annual territorial emissions of CO_2, measured in megatonnes ($MtCO_2$).
- It enables the analysis of greenhouse gas emissions by geographical regions or countries over time.
- These insights are crucial for understanding the contribution of different regions to global emissions, tracking compliance with international climate agreements, and assessing progress towards decarbonization goals.

2. Global Land Average Temperature Anomalies per Year

- This dataset contains data on the annual temperature anomalies recorded across global land areas, relative to a pre-industrial baseline.
- It highlights changes in average land surface temperatures, enabling the study of long-term global warming trends.
- By comparing these anomalies to CO_2 emissions and other climate indicators, researchers can evaluate the relationship between human activity and temperature variations, offering insights into the impacts of climate change on ecosystems and human societies.

4. Data Types and Formats

Formats:

- Datasets were provided in CSV format.

Key columns include:

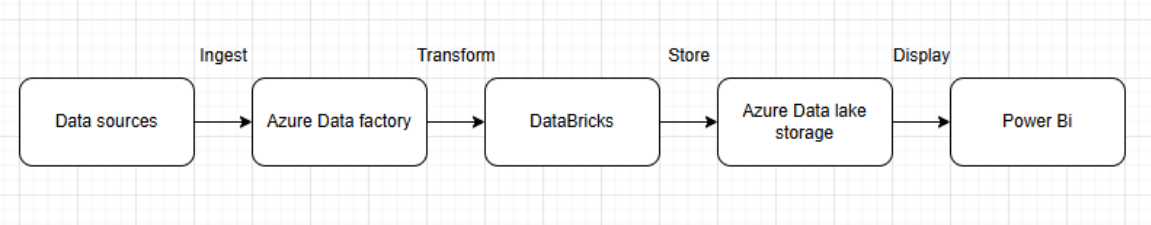
- Date

- Region

- Temperature Anomaly

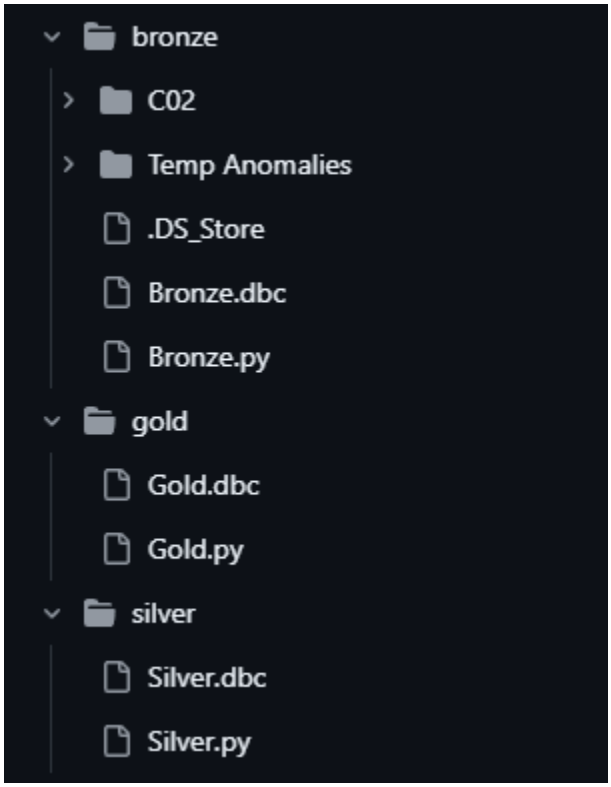
- CO2 Emissions

5. System Architecture Model



Lakehouse Architecture:

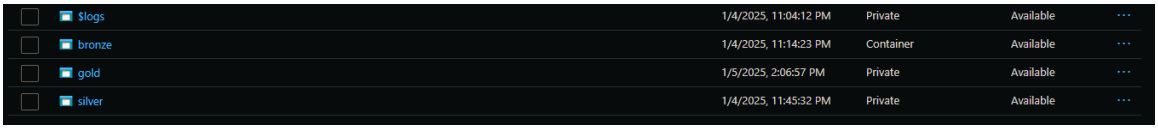
- Bronze Layer: Stores raw ingested datasets.
- Silver Layer: Contains cleaned and structured datasets.
- Gold Layer: Holds aggregated and analysis-ready data.



6. Azure and Databricks Implementation

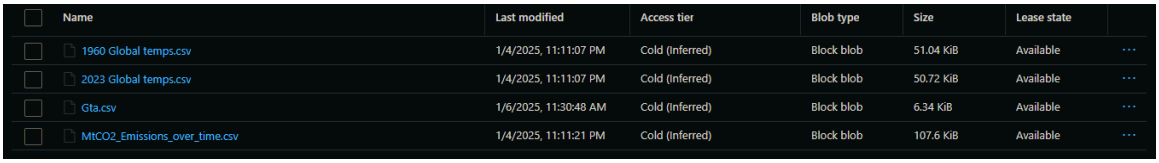
Azure Resources:

- Storage: Used Azure Data Lake for raw and processed data.



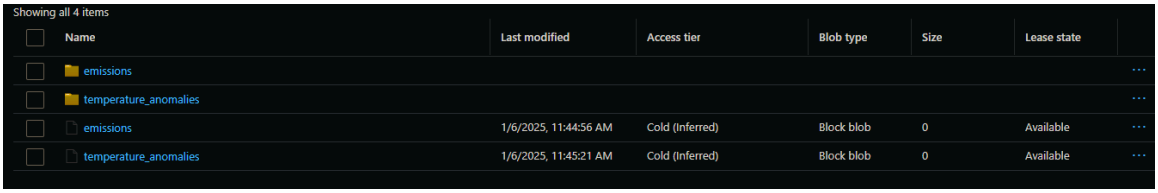
$logs	1/4/2025, 11:04:12 PM	Private	Available	···
bronze	1/4/2025, 11:14:23 PM	Container	Available	···
gold	1/5/2025, 2:06:57 PM	Private	Available	···
silver	1/4/2025, 11:45:32 PM	Private	Available	···

Bronze:



Name	Last modified	Access tier	Blob type	Size	Lease state	
1960 Global temps.csv	1/4/2025, 11:11:07 PM	Cold (Inferred)	Block blob	51.04 KiB	Available	···
2023 Global temps.csv	1/4/2025, 11:11:07 PM	Cold (Inferred)	Block blob	50.72 KiB	Available	···
Gta.csv	1/6/2025, 11:30:48 AM	Cold (Inferred)	Block blob	6.34 KiB	Available	···
MtCO2_Emissions_over_time.csv	1/4/2025, 11:11:21 PM	Cold (Inferred)	Block blob	107.6 KiB	Available	···

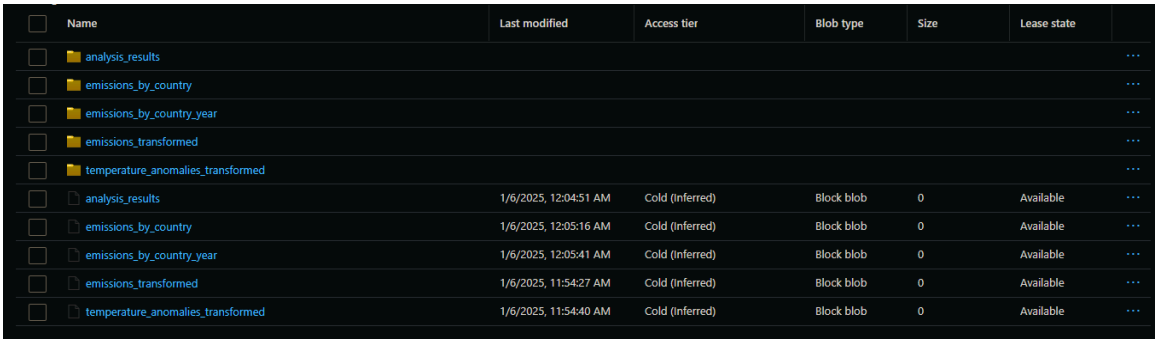
Silver:



Showing all 4 items

Name	Last modified	Access tier	Blob type	Size	Lease state	
emissions						···
temperature_anomalies						···
emissions	1/6/2025, 11:44:56 AM	Cold (Inferred)	Block blob	0	Available	···
temperature_anomalies	1/6/2025, 11:45:21 AM	Cold (Inferred)	Block blob	0	Available	···

Gold:



Name	Last modified	Access tier	Blob type	Size	Lease state	
analysis_results						···
emissions_by_country						···
emissions_by_country_year						···
emissions_transformed						···
temperature_anomalies_transformed						···
analysis_results	1/6/2025, 12:04:51 AM	Cold (Inferred)	Block blob	0	Available	···
emissions_by_country	1/6/2025, 12:05:16 AM	Cold (Inferred)	Block blob	0	Available	···
emissions_by_country_year	1/6/2025, 12:05:41 AM	Cold (Inferred)	Block blob	0	Available	···
emissions_transformed	1/6/2025, 11:54:27 AM	Cold (Inferred)	Block blob	0	Available	···
temperature_anomalies_transformed	1/6/2025, 11:54:40 AM	Cold (Inferred)	Block blob	0	Available	···

Databricks Implementation:

1. Set up pipelines for data ingestion and transformation.

2. Mounted Azure Data Lake containers to Databricks for direct access.

(Verifiable by looking at bronze silver or gold mounts in the python scripts)

7. Data Transformation and Modeling

Steps:

1. Cleaning Data: Removed duplicates and handled missing values.

2. Transformation:

- Aggregated metrics like temperature trends by decade.

- Standardized data formats for compatibility.

3. Data Modeling:

- Star schema design with:

- Fact table: Climate metrics (e.g., temperature, CO2 emissions).

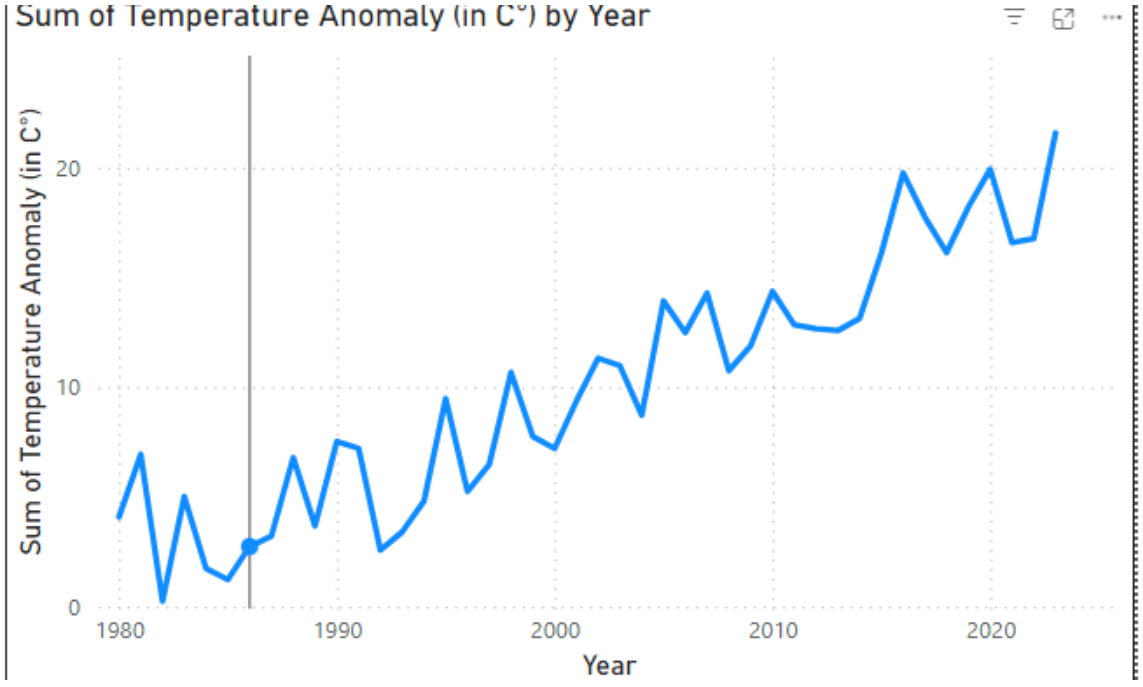
- Dimension tables: Time, region, and sector.

8. Data Visualization

Visualizations Built in Power BI:

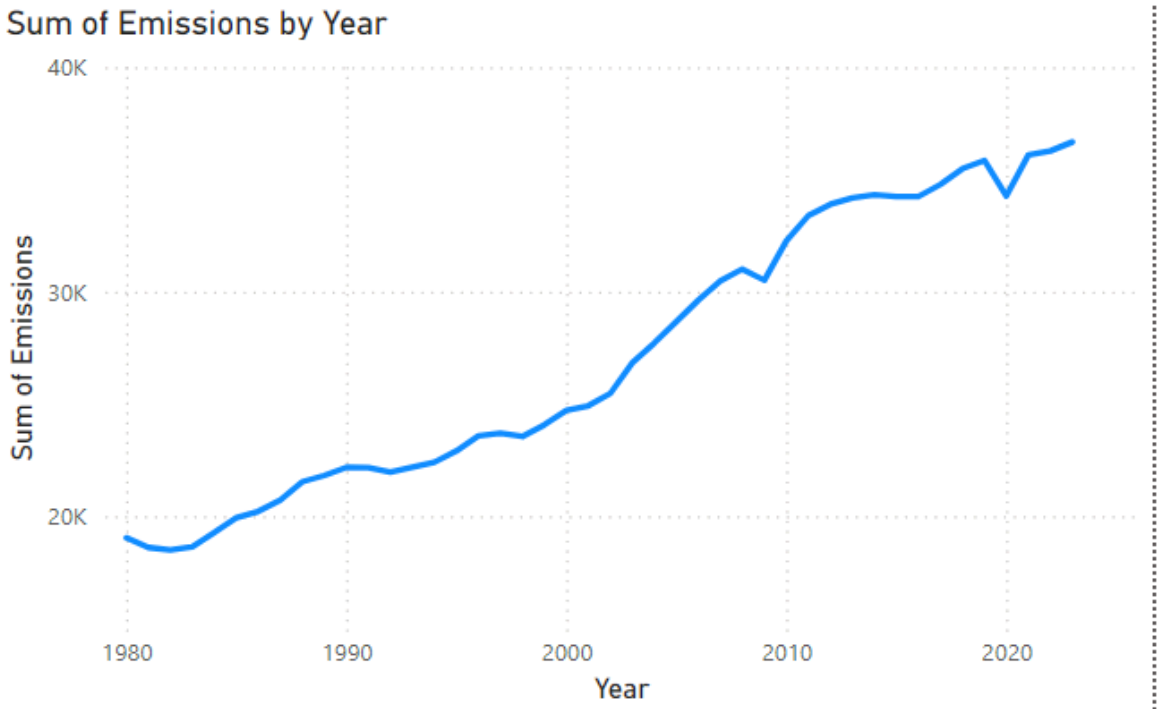
1. Global Temperature Trends:

- Line charts showing changes over decades.



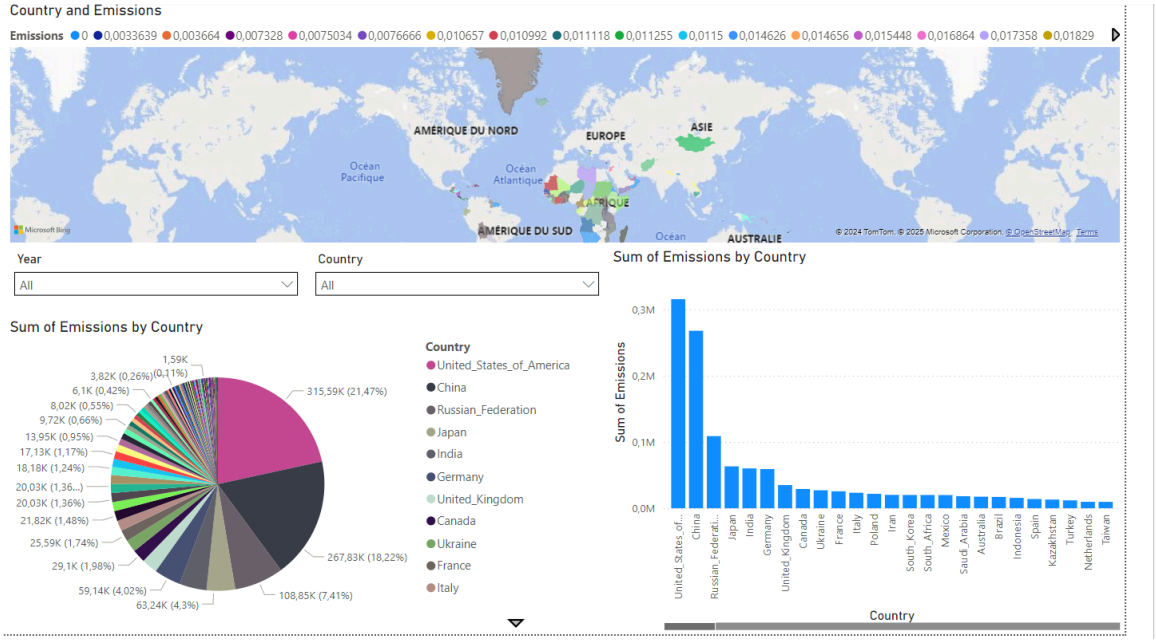
2. CO2 Emission:

- Bar charts for Global emission comparison.



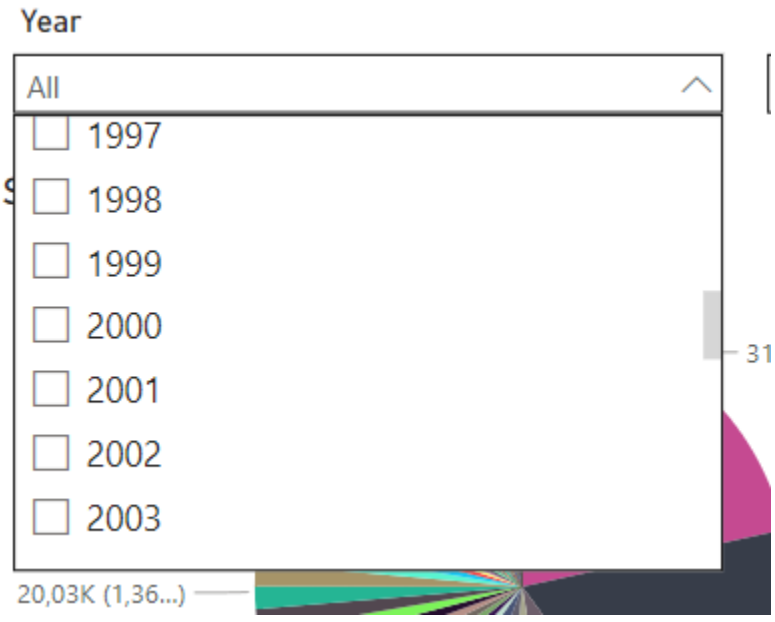
3. Geographic Distribution:

- Maps depicting the regional impact of climate changes.

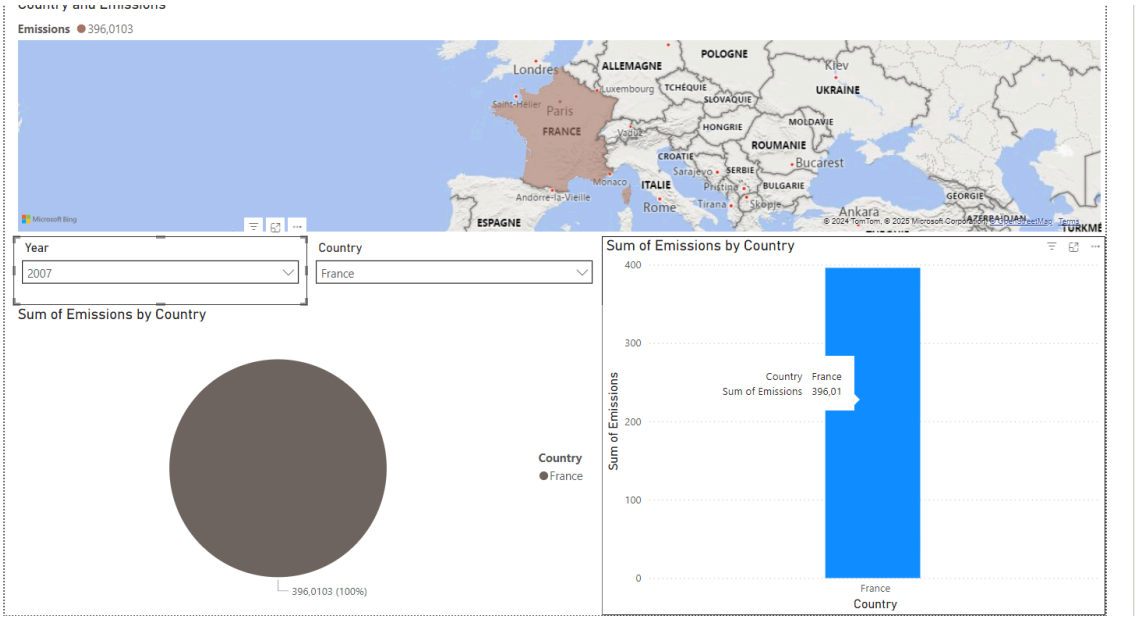


4.Use of filters and interactive visuals

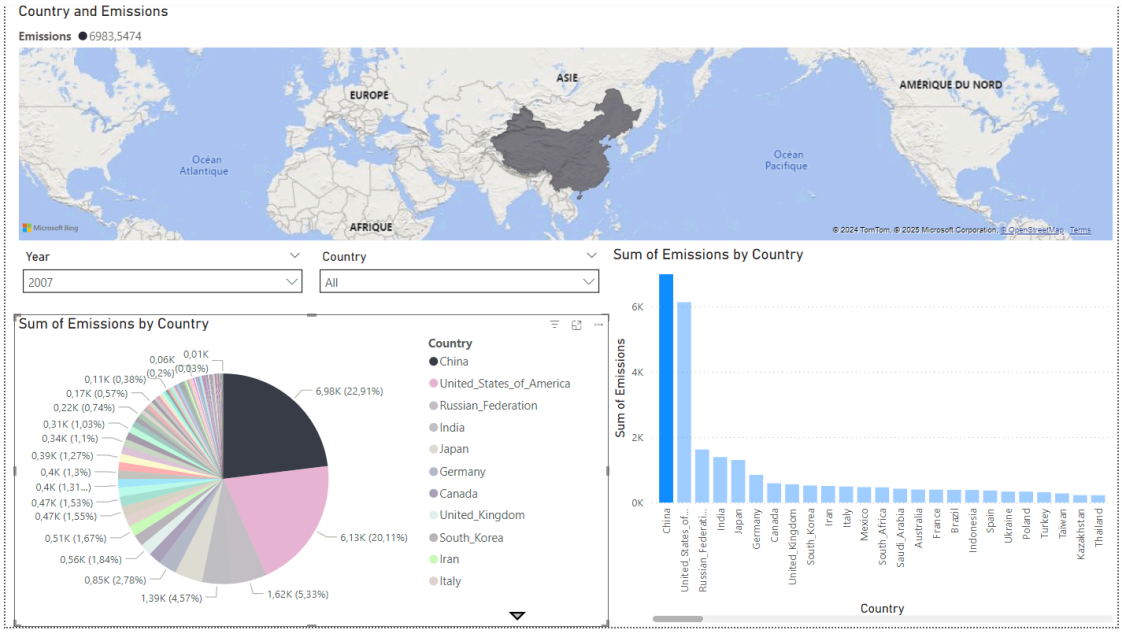
-Use of dropdowns for interactive visuals



France selected in the dropdown



China selected in the pie chart highlights automatically ll other aspects for easy data understanding



9. Conclusion

- This project highlights the increasing influence of climate indicators on global trends.

- Insights derived can aid policymakers and environmentalists in targeting areas for intervention.

- Future work could include:

- Real-time streaming for ongoing trends.

- Predictive analytics to forecast future scenarios.