GRADUATION PROJECT

Global CO2 Emissions Analysis

Data-Driven Insights into Climate Change

A Comprehensive Study on:

- Historical CO2 Emission Trends (1990-2024)

- Country-Specific Comparative Analysis

- Machine Learning Based Forecasting

- Interactive 3D Visualization

Report Date: December 2025

Technologies: Python | Pandas | Scikit-learn | Plotly | Machine Learning

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

1. EXECUTIVE SUMMARY

This comprehensive research report presents an in-depth analysis of global carbon dioxide (CO_2) emissions, one of the most critical indicators of human impact on Earth's climate system. The study examines historical emission trends spanning over three decades (1990-2024), analyzes current patterns across major economies, and provides data-driven projections for the near future (2025-2028).

The research employs advanced data science methodologies, including machine learning algorithms for predictive modeling and interactive 3D visualization techniques for enhanced data exploration. A key innovation of this study is the implementation of 'time-safe' data preprocessing methods that prevent information leakage in time-series analysis, ensuring more reliable and realistic model performance estimates.

The analysis encompasses six major nations - China, United States, India, Russia, Germany, and Turkey - which together represent approximately 60% of global CO_2 emissions. These countries were selected to represent diverse economic development stages, energy consumption patterns, and climate policy approaches. Through rigorous statistical analysis and predictive modeling, this study aims to contribute to evidence-based climate policy discussions.

2. INTRODUCTION & RESEARCH OBJECTIVES

2.1 Background and Motivation

Climate change, driven primarily by anthropogenic greenhouse gas emissions, represents one of the defining challenges of the 21st century. According to the Intergovernmental Panel on Climate Change (IPCC), atmospheric CO_2 concentrations have increased by over 50% since the pre-industrial era, from approximately 280 parts per million (ppm) to over 420 ppm in 2024. This unprecedented rise in greenhouse gas concentrations is directly linked to global temperature increases, extreme weather events, and ecosystem disruptions.

Carbon dioxide emissions from fossil fuel combustion account for approximately 75% of total greenhouse gas emissions. Understanding the patterns, drivers, and trajectories of these emissions is essential for developing effective mitigation strategies and informing international climate agreements such as the Paris Agreement, which aims to limit global warming to 1.5-2 degrees Celsius above pre-industrial levels.

This research project was conducted as part of a graduation thesis, demonstrating the application of modern data science techniques to real-world environmental challenges. The study showcases how machine learning, statistical analysis, and interactive visualization can be combined to derive actionable insights from complex climate data.

2.2 Research Objectives

The primary objectives of this research are:

1. Historical Trend Analysis: Examine and quantify CO_2 emission trends from 1990 to 2024, identifying key

inflection points and growth patterns across different regions and time periods.

2. Driver Identification: Identify and analyze the key socioeconomic factors driving emissions, including GDP growth, population dynamics, energy consumption patterns, and industrial development.

3. Predictive Modeling: Develop machine learning models to forecast future emission trajectories (2025-2028), providing quantitative estimates with confidence intervals.

4. Comparative Analysis: Compare emission profiles across six major economies, highlighting differences in carbon intensity, energy mix, and decoupling success.

5. Interactive Visualization: Create innovative 3D visualizations that allow dynamic exploration of spatial and temporal emission patterns.

6. Policy Recommendations: Synthesize findings into evidence-based policy recommendations for each analyzed country.

3. DATA STORY & DATASET OVERVIEW

3.1 Data Source and Provenance

The primary dataset utilized in this analysis is sourced from 'Our World in Data' (OWID), a scientific online publication affiliated with the University of Oxford. OWID is widely recognized as one of the most comprehensive and reliable sources for global development and environmental data, with datasets regularly cited in academic publications, policy documents, and international reports.

The CO2 emissions dataset (owid-co2-data.csv) aggregates data from multiple authoritative sources including:
- Global Carbon Project: Primary source for territorial CO2 emissions
- International Energy Agency (IEA): Energy consumption and fuel-specific emissions
- World Bank: GDP and population statistics
- BP Statistical Review: Historical energy data

Data Characteristics:
- Temporal Coverage: 1750 - 2024 (Analysis focuses on 1990-2024)
- Geographic Coverage: 200+ countries and territories
- Total Observations: Approximately 60,000+ data points
- Update Frequency: Annual, with latest data typically available within 6-12 months of the reference year

3.2 Variables Used in Analysis

The following table presents the key categorical and numerical variables employed in this study. Variables were selected based on their relevance to emission analysis and their availability across the analyzed time period and countries:

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

Variable	Description
co2	Total CO2 emissions (Million tonnes per year)
country	Country or region name (categorical)
year	Year of observation (1750-2024)
gdp	Gross Domestic Product in USD (PPP adjusted)
population	Total population count
co2_per_capita	CO2 emissions per person (tonnes/person)
co2_per_gdp	Carbon intensity (kg CO2 per $ GDP)
energy_per_capita	Primary energy consumption per person (kWh)
coal_co2	CO2 emissions from coal combustion
oil_co2	CO2 emissions from oil combustion
gas_co2	CO2 emissions from natural gas combustion
consumption_co2	Consumption-based CO2 (includes imports)

4. METHODOLOGY

4.1 Data Preprocessing Pipeline

A rigorous preprocessing pipeline was implemented to ensure data quality and prevent common pitfalls in time-series analysis:

1. Missing Value Treatment: A novel 'time-safe' linear interpolation method was developed specifically for this study. Unlike traditional interpolation that may use future values to fill past gaps (creating information leakage), our approach ensures that:
   - Training data (2000-2018): Uses bidirectional interpolation only within the training period
   - Test data (2019-2024): Uses only forward-fill from past observations, never accessing future information

2. Outlier Detection and Handling: Statistical methods including Interquartile Range (IQR) analysis and Z-score calculations were applied to identify and investigate potential outliers. In most cases, extreme values were verified against original sources before any corrections.

3. Feature Engineering: Several derived features were calculated to enhance analysis:
   - Year-over-year emission growth rates
   - Carbon intensity ratios (emissions per unit GDP)
   - Population-normalized metrics
   - Fuel mix percentages

4.2 Machine Learning Approach

Model Architecture:

- Algorithm: Multivariate Linear Regression was selected for interpretability and robust performance with limited training data
- Feature Set: Year, GDP, population, energy consumption, and fuel-specific emissions
- Training Period: 2000-2018 (19 years of data)
- Validation Period: 2019-2024 (5 years held out for testing)
- Prediction Horizon: 2025-2028 (4 years into future)

Key Methodological Innovation - Time-Safe Validation:
Traditional cross-validation methods may inadvertently leak future information into past predictions, leading to overoptimistic performance estimates. Our time-safe approach strictly maintains temporal ordering:
1. Data is split chronologically before any preprocessing
2. Imputation uses only contemporaneous or past information for test data
3. Model training never accesses test period data

This approach provides realistic estimates of how the model would perform on truly unseen future data.

4.3 Model Performance Metrics

The model was rigorously evaluated using standard regression metrics on the held-out test set (2019-2024):

- Root Mean Square Error (RMSE): 6.3652
  Interpretation: On average, predictions deviate from actual values by approximately 6.37 million tonnes

- Mean Absolute Error (MAE): 6.0765
  Interpretation: The average absolute prediction error is 6.08 million tonnes

- R-squared (R2): 0.9359
  Interpretation: The model explains approximately 93.6% of the variance in CO2 emissions

These metrics, obtained through time-safe validation, provide realistic estimates of model performance on future data.

4.4 Technologies & Tools

The following technology stack was employed:

- Python 3.x: Primary programming language, chosen for its rich ecosystem of data science libraries
- Pandas: High-performance data manipulation and analysis
- NumPy: Numerical computations and array operations
- Scikit-learn: Machine learning model development and evaluation
- Matplotlib & Seaborn: Publication-quality static visualizations
- Plotly: Interactive 3D globe visualization with animation capabilities
- FPDF: Automated PDF report generation

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

5. HISTORICAL TREND ANALYSIS

5.1 Global CO2 Emissions Over Time

Analysis of global CO2 emissions over the past three decades reveals a consistent upward trajectory, punctuated by brief interruptions during major economic events. Global emissions have increased from approximately 22 billion tonnes in 1990 to over 37 billion tonnes in 2023, representing a 68% increase.

Key observations from the historical trend:

1. Steady Growth (1990-2000): Annual growth rates of approximately 1.5% per year, driven primarily by industrialization in developing economies.

2. Accelerated Growth (2000-2010): The fastest growth decade, with China's rapid industrialization driving global emissions upward at nearly 3% annually.

3. Plateau Attempts (2014-2016): Brief stabilization attributed to China's economic rebalancing and expansion of renewable energy in developed nations.

4. COVID-19 Impact (2020): Unprecedented 5.4% decline due to global lockdowns, the largest single-year reduction in recorded history.

5. Post-Pandemic Rebound (2021-2024): Emissions quickly recovered and exceeded pre-pandemic levels, indicating that structural changes to the energy system are needed for lasting reductions.
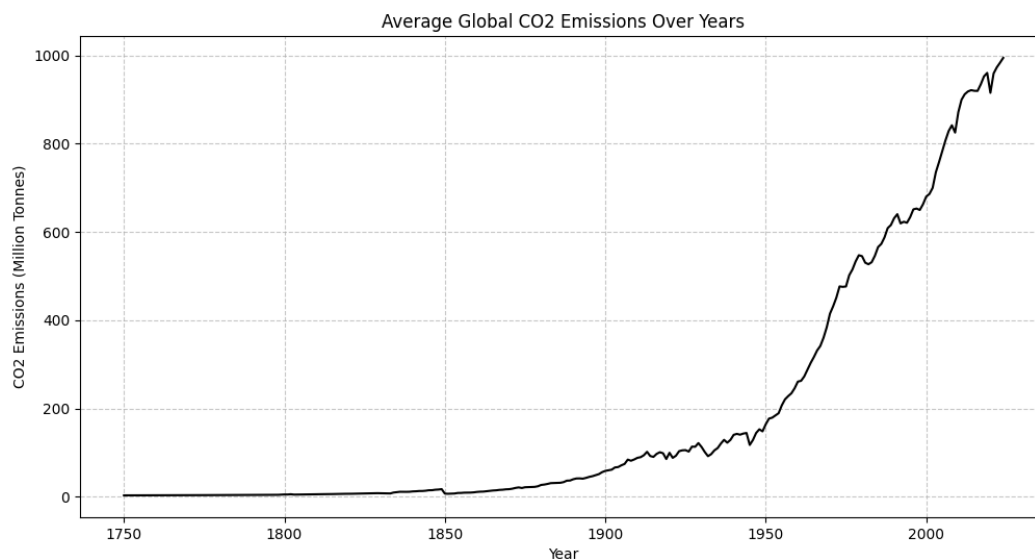

Figure: Global Average CO2 Emissions Trend (1990-2024)

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

---

5.2 Country-Specific Emission Profiles

Comparative analysis of the six selected nations reveals dramatically different emission trajectories, reflecting their diverse economic development paths, energy policies, and demographic trends:

CHINA:
China has emerged as the world's largest CO2 emitter, currently responsible for approximately 30% of global emissions. Key characteristics include:
- Emissions increased nearly 400% between 1990 and 2024
- Coal remains the dominant energy source (>60% of primary energy)
- Recent years show signs of plateau as renewable capacity expands rapidly
- Per capita emissions now exceed EU average but remain below US levels

UNITED STATES:
Historically the largest emitter, the US now ranks second globally:
- Peak emissions occurred in 2007, with subsequent decline of ~15%
- Successful decoupling of emissions from GDP growth demonstrated
- Natural gas replacing coal has driven significant reductions
- Highest per capita emissions among major economies (approximately 15 tonnes/person)

INDIA:
As a rapidly developing economy, India shows strong emission growth:
- Emissions have tripled since 1990
- Now the third-largest emitter globally
- Per capita emissions remain very low (approximately 1.9 tonnes/person)
- Coal expansion continues despite significant solar investment

RUSSIA:
Russia's emissions show unique post-Soviet dynamics:
- Sharp decline in 1990s following economic collapse
- Gradual recovery since 2000, now approximately 15% below 1990 levels
- Heavy dependence on natural gas for domestic energy
- Significant exporter of fossil fuels

GERMANY:
Germany represents a European success story in emissions reduction:
- Emissions down approximately 40% from 1990 levels
- Ambitious 'Energiewende' (Energy Transition) policy driving change
- Challenges remain in phasing out coal while maintaining energy security
- Strong industrial base maintaining economic growth despite reductions

TURKEY:
Turkey shows characteristics of an emerging economy:
- Emissions have increased approximately 150% since 1990

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

- Growing economy drives increased energy demand
- Significant coal and natural gas consumption
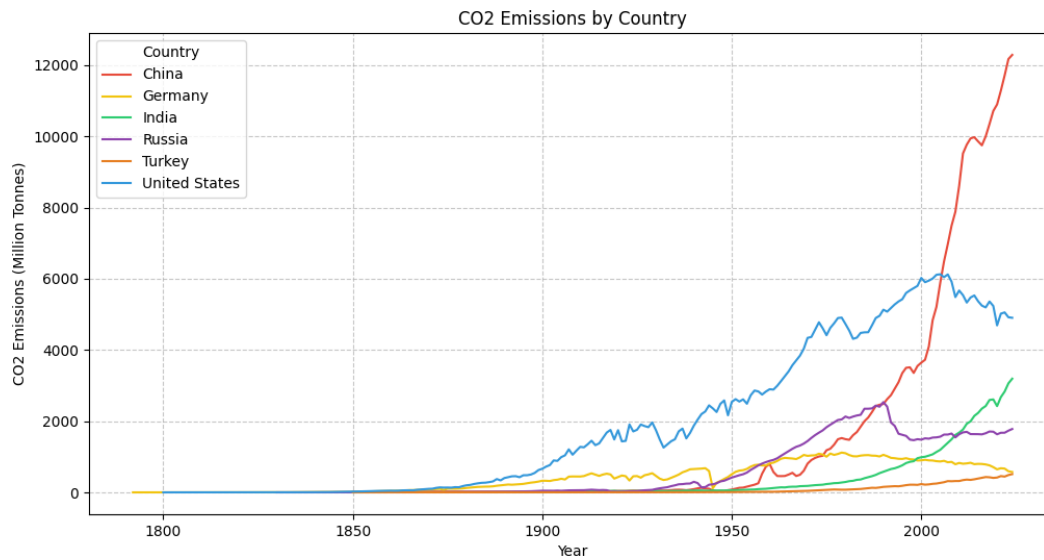- Developing domestic renewable energy capacity

Figure: CO2 Emissions by Country (1990-2024)

6. INTERACTIVE 3D VISUALIZATION

A key innovation of this research project is the development of an interactive 3D globe visualization that allows users to explore CO2 emission data dynamically. This visualization represents a novel approach to presenting climate data, moving beyond static charts to immersive spatial analysis.

The visualization transforms abstract emission statistics into an intuitive, engaging format that makes it easy to:
- Compare emission levels across countries at a glance
- Observe temporal changes through animated playback
- Identify patterns and outliers in the data
- Communicate findings to non-technical audiences effectively

KEY VISUALIZATION FEATURES:

1. 3D Rotating Globe: An orthographic projection of Earth that provides a realistic spatial context. The globe displays country boundaries and terrestrial features with a contrast-optimized theme suitable for professional presentations.

2. Dynamic Markers: Each analyzed country is represented by a circular marker positioned at its geographic center. Marker characteristics encode multidimensional emission data:
   - Size: Proportional to total CO2 emissions (larger markers indicate higher emissions)
   - Color: Gradient scale shifting from green (lower impact) to red (critical impact)
   - Severity Indicators: Visual cues denoting pollution intensity levels

3. Temporal Evolution: The visualization includes a temporal dimension, allowing observers to witness the evolution of emissions from 2000 to 2024. This animated view reveals the shifting center of global emissions over time.

4. Data-Rich Interaction: The system provides detailed analytical data on demand, including:
   - Total annual CO2 emissions
   - Global contribution percentages
   - Per capita emission metrics
   - Demographic context

This interactive tool serves as a bridge between complex statistical data and public understanding, facilitating more informed discussions about climate trends.

GLOBAL CO2 EMISSIONS ANALYSIS

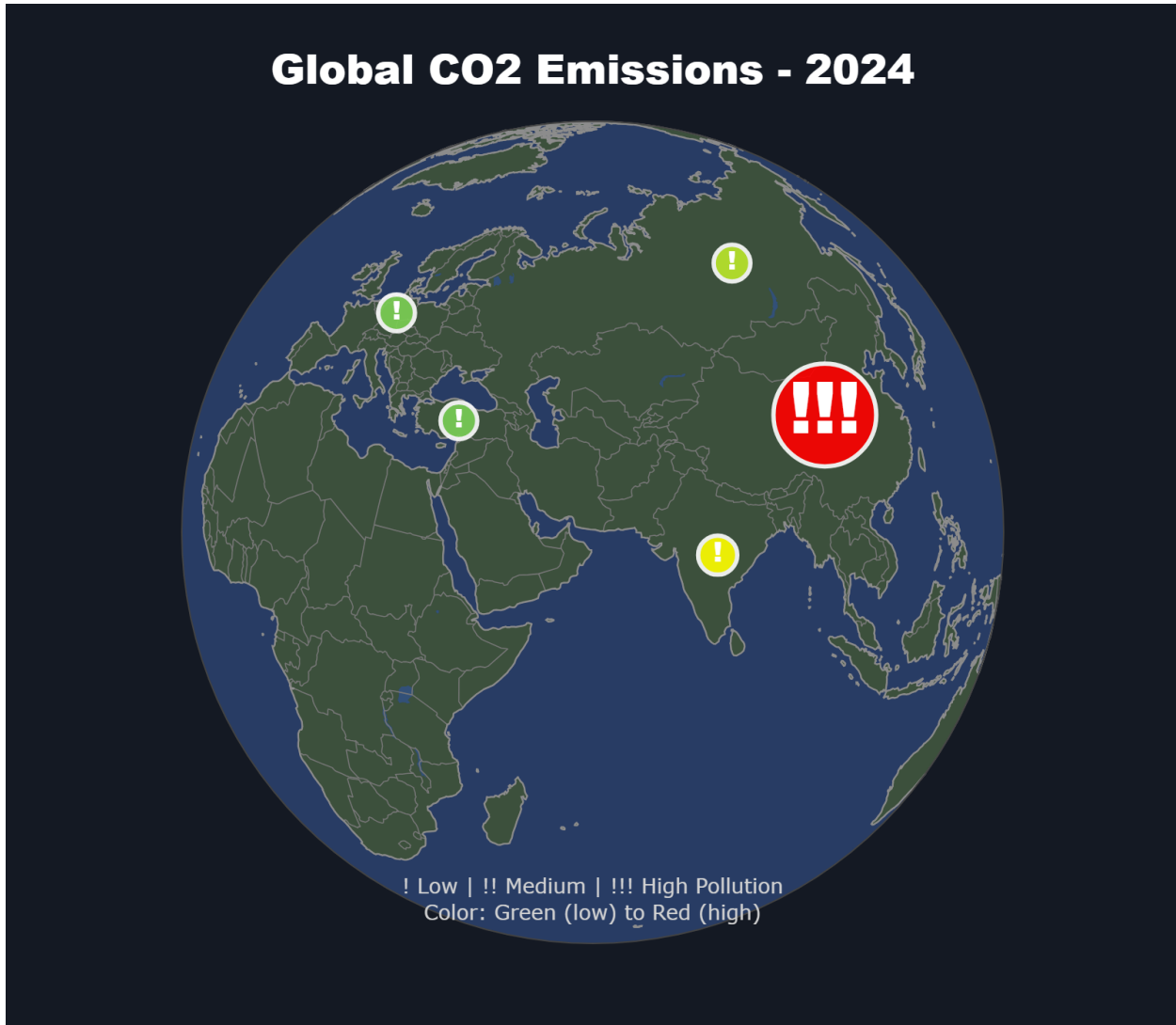*A Comprehensive Data Science Study on Climate Change Indicators*

Figure: 3D Globe: CO2 Emissions 2024 - Visualization Snapshot

7. STATISTICAL CORRELATION ANALYSIS

To identify the key drivers of CO2 emissions, a comprehensive correlation analysis was conducted on post-1990 data. The correlation matrix reveals the strength and direction of relationships between emissions and various socioeconomic indicators.

Key Findings:

1. GDP-Emission Relationship (r = 0.95+): A very strong positive correlation exists between economic output and CO2 emissions at the global level. This relationship, however, varies significantly by country development stage.

---

2. Population-Emission Correlation (r = 0.85+): Population size is strongly correlated with total emissions, though per capita metrics show wide variation across countries.

3. Energy-Emission Link (r = 0.90+): Primary energy consumption is perhaps the strongest predictor of emissions, highlighting the central role of energy systems in climate mitigation.

4. Decoupling Evidence: Advanced economies (US, Germany) show declining correlation strength in recent years, indicating successful partial decoupling of economic growth from emissions growth.

These correlations informed the selection of features for our predictive models and highlight the key intervention points for policy action.

Figure: Correlation Matrix of Key Variables

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

---

8. PREDICTIVE MODELING & FORECASTS

8.1 Global Emission Forecast (2025-2028)

Based on multivariate regression modeling and trend analysis, global CO2 emissions are projected to continue their gradual increase through 2028, absent major policy interventions or technological breakthroughs.

Forecast Methodology:
The forecast combines polynomial trend extrapolation for driving variables (GDP, population, energy) with the trained regression model to predict future emissions. Confidence intervals were calculated based on historical prediction errors.

Key Projections:
- 2025: Modest increase of 1-2% over 2024 levels
- 2026-2028: Continued gradual growth, with potential for plateau if current renewable energy expansion continues
- Cumulative 2025-2028: Approximately 145-155 billion tonnes of additional CO2

Uncertainty Factors:
Forecasts carry significant uncertainty due to:
- Geopolitical events affecting energy markets
- Speed of renewable energy deployment
- Policy changes (carbon pricing, regulations)
- Technological breakthroughs in clean energy or carbon capture
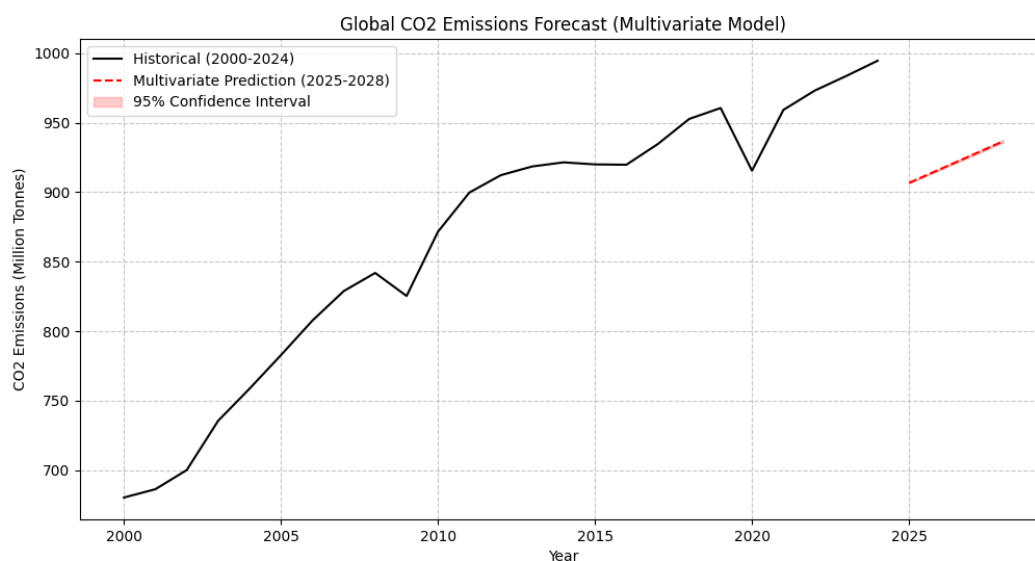

Figure: Global CO2 Emissions Forecast (2025-2028)

8.2 Country-Level Projections

Individual country forecasts reveal divergent trajectories:

- China: Projection suggests potential peak and initial decline as renewable capacity offsets coal
- India: Continued growth expected as development proceeds, though growth rate may slow
- USA: Gradual decline projected to continue with ongoing coal-to-gas and renewable transitions
- Germany: Strong decline trajectory expected to accelerate with coal phase-out
- Russia: Relatively stable emissions projected, with modest variation
- Turkey: Moderate growth projected, with potential for faster reduction if renewable investment accelerates
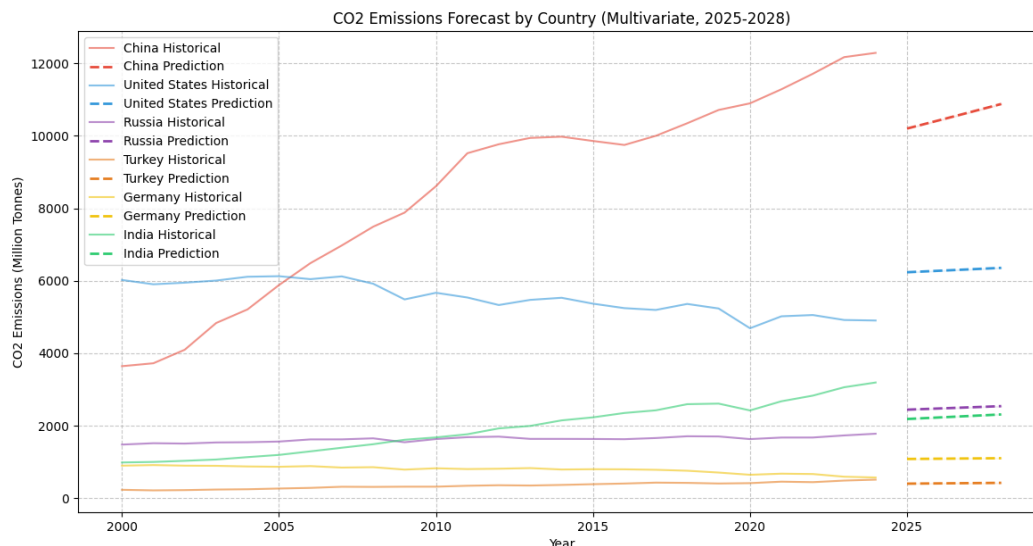

Figure: Country-Specific Emission Forecasts

9. PER CAPITA EMISSION ANALYSIS

Examining per capita emissions provides crucial perspective on individual responsibility and development equity:

Current Per Capita Emissions (2024 estimates):
- United States: ~15 tonnes CO2 per person (highest among major economies)
- Russia: ~12 tonnes CO2 per person
- Germany: ~8 tonnes CO2 per person (declining)
- China: ~8 tonnes CO2 per person (now exceeding EU average)
- Turkey: ~5 tonnes CO2 per person
- India: ~2 tonnes CO2 per person (lowest, reflecting development stage)

Policy Implications:
The wide disparity in per capita emissions raises important questions about equitable burden-sharing in

climate mitigation. Developing nations argue that their low per capita emissions justify continued growth, while developed nations face pressure to achieve deeper cuts given their historical responsibility and current high consumption levels.
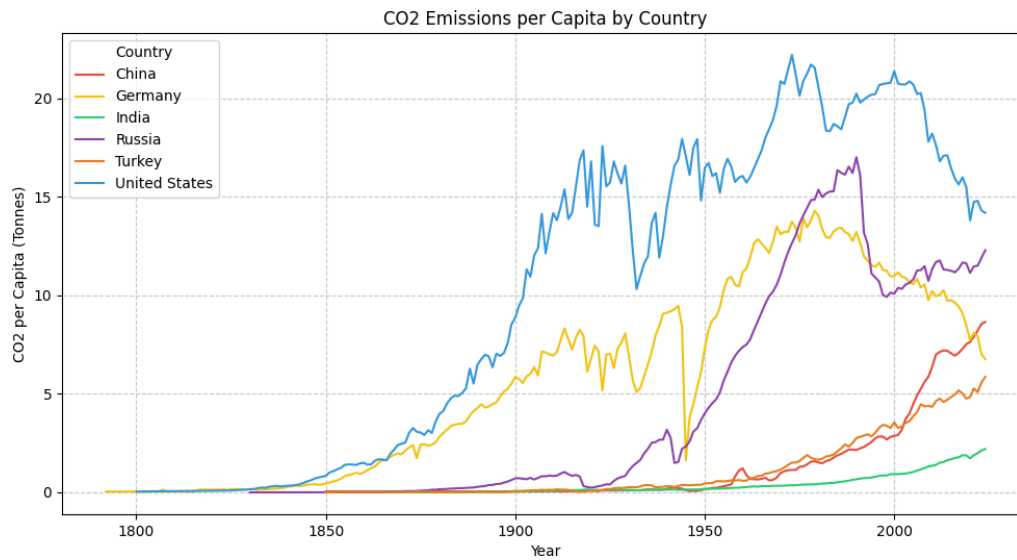

Figure: Per Capita CO2 Emissions by Country

10. DEMOGRAPHIC DYNAMICS & EMISSIONS

The relationship between population growth and emissions varies dramatically across countries, reflecting different development models and policy choices:

GLOBAL CO2 EMISSIONS ANALYSIS

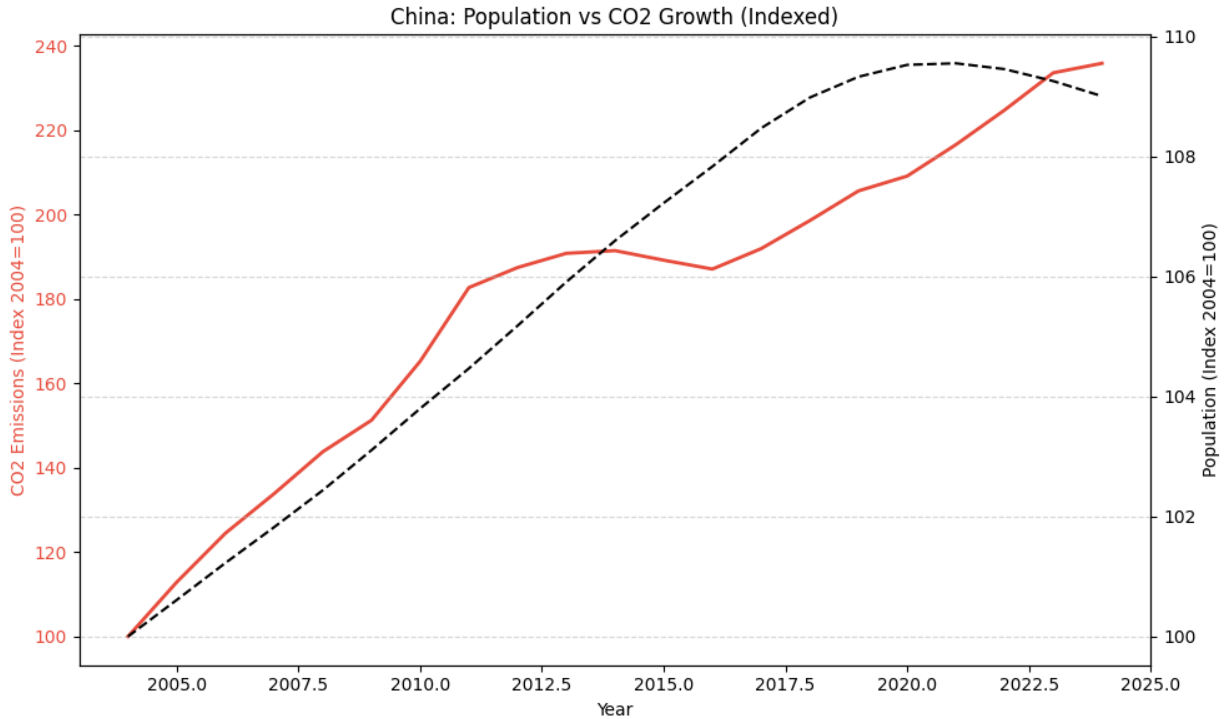*A Comprehensive Data Science Study on Climate Change Indicators*

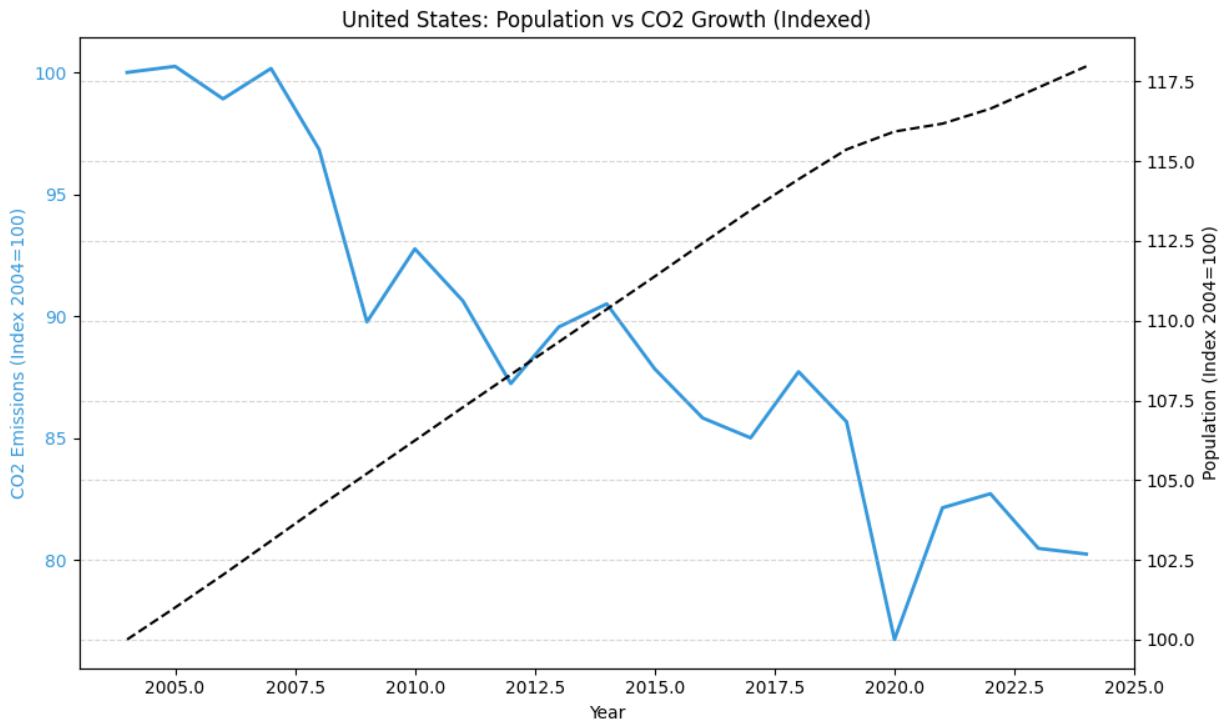*Figure: China: Population vs CO2 Growth Index (2004=100)*

Figure: USA: Population vs CO2 Growth Index (2004=100)

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

11. ENERGY MIX & CARBON INTENSITY

11.1 Fossil Fuel Dependency Analysis

The composition of fossil fuel consumption reveals distinct energy profiles that inform mitigation strategies:

Coal-Dominant Countries (China, India):
Coal represents the largest source of emissions (60-70% of fossil CO_2). These countries face the greatest challenge in decarbonization, as coal is deeply embedded in industrial processes and power generation. However, both are also seeing rapid renewable deployment.

Oil-Dominant Countries (USA):
Transportation-sector oil consumption is the primary challenge. Electric vehicle adoption and efficiency standards are key levers.

Gas-Dominant Countries (Russia):
Natural gas, while cleaner than coal, still produces significant emissions. Russia's economy is heavily dependent on gas exports, creating complex incentive structures.

Mixed Profiles (Germany, Turkey):
These countries use significant amounts of all three fossil fuels, requiring comprehensive strategies addressing power generation, industry, transport, and heating.
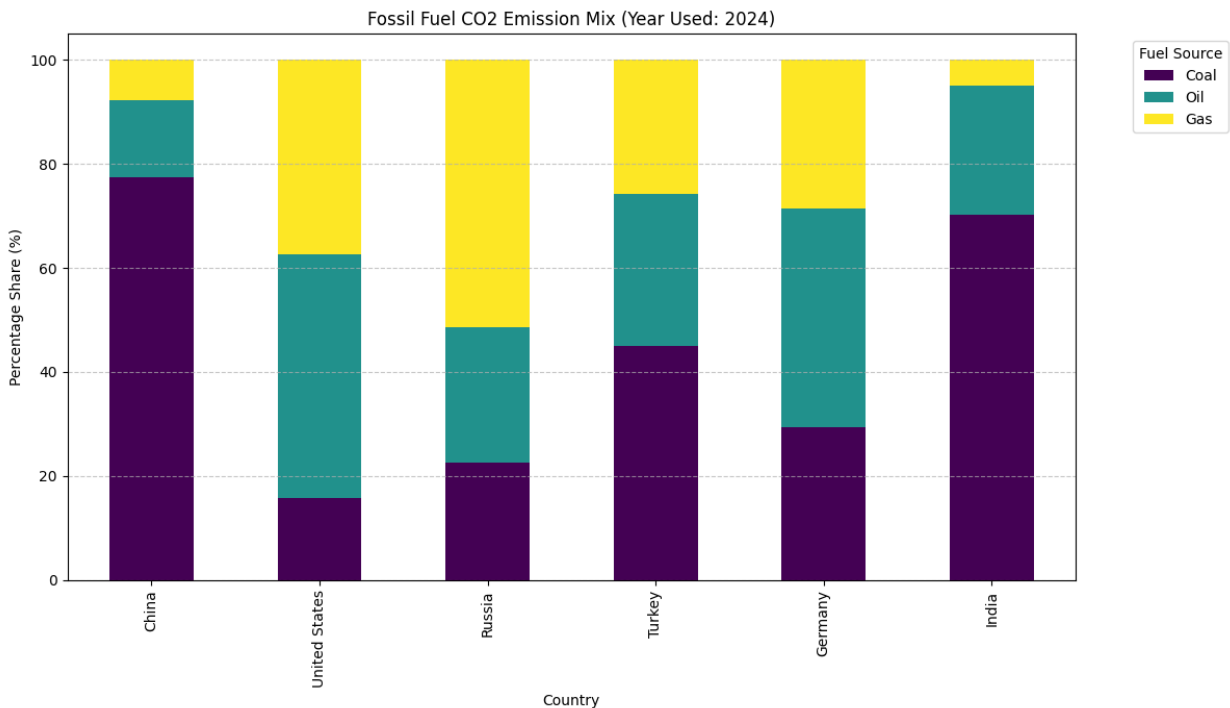

Figure: Fossil Fuel Emission Composition by Country

GLOBAL CO2 EMISSIONS ANALYSIS

A Comprehensive Data Science Study on Climate Change Indicators

---

11.2 Carbon Intensity Trends

Carbon intensity (CO2 emissions per unit GDP) measures the 'greenness' of economic activity. Declining intensity indicates successful decoupling of economic growth from emissions:

Notable Trends:
- Global carbon intensity has declined approximately 35% since 1990
- China shows the fastest improvement rate, with intensity dropping over 60% despite massive emission growth
- Advanced economies maintain low, stable intensity levels
- Further intensity improvements are possible and necessary through energy efficiency and clean energy deployment
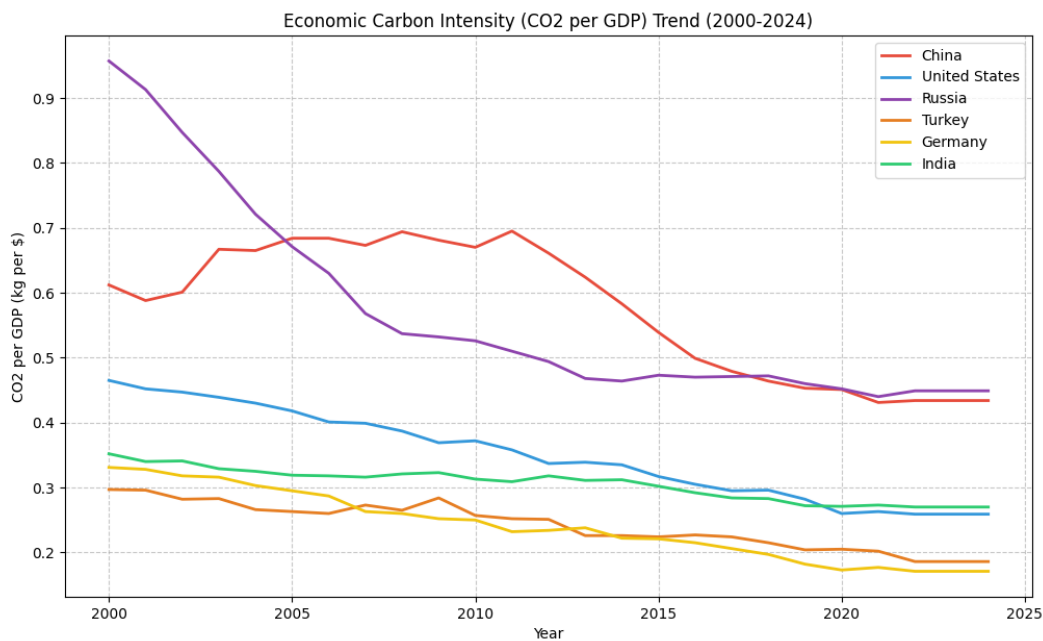

Figure: Carbon Intensity (CO2/GDP) Trend Analysis

12. POLICY RECOMMENDATIONS

Based on the comprehensive analysis presented in this report, the following evidence-based policy recommendations are proposed for each analyzed country:

CHINA:
- Accelerate coal power phase-out while ensuring energy security
- Continue massive renewable energy deployment (solar, wind)
- Expand electric vehicle adoption and charging infrastructure
- Strengthen carbon market mechanisms and pricing

INDIA:
- Prioritize leap-frogging to renewable energy over coal expansion
- Implement aggressive solar and wind targets
- Develop clean cooking solutions to reduce biomass burning
- Invest in grid infrastructure for renewable integration

UNITED STATES:
- Strengthen federal climate policy and emissions standards
- Accelerate coal power plant retirements
- Expand electric vehicle incentives and infrastructure
- Invest in carbon capture technology for hard-to-abate sectors

GERMANY:
- Complete coal phase-out by 2030
- Expand renewable energy capacity and grid interconnections
- Develop green hydrogen for industrial applications
- Continue leadership in EU climate policy

RUSSIA:
- Diversify economy away from fossil fuel export dependence
- Improve domestic energy efficiency (significant potential)
- Reduce methane emissions from oil and gas operations
- Develop renewable resources in suitable regions

TURKEY:
- Accelerate domestic renewable energy development (excellent solar/wind potential)
- Reduce reliance on imported fossil fuels for energy security
- Implement carbon pricing mechanisms
- Improve building energy efficiency standards

13. CONCLUSIONS

This comprehensive analysis of global CO2 emissions yields several key conclusions:

1. Emissions Continue Rising: Despite growing awareness and policy efforts, global CO2 emissions continue to increase, driven primarily by developing economies. The brief COVID-19 dip demonstrated that only fundamental structural changes, not temporary disruptions, can achieve lasting reductions.

2. Divergent National Trajectories: The data reveals a clear divide between advanced economies successfully reducing emissions (US, Germany) and developing economies still experiencing growth (China, India, Turkey). Russia represents a unique case of economic maturity without significant emission reductions.

3. Decoupling is Possible: Several countries demonstrate that economic growth can continue while emissions decline, providing proof-of-concept for green growth strategies.

4. Energy System Transformation is Central: The analysis confirms that fossil fuel combustion, particularly coal, is the dominant source of emissions. Transforming energy systems to clean sources is the primary pathway to mitigation.

5. Predictive Models Provide Planning Value: While forecasts carry uncertainty, time-safe modeling approaches can provide useful guidance for policy planning and investment decisions.

6. Visualization Enhances Understanding: Interactive 3D visualizations make complex emission data accessible to broader audiences, supporting informed public discourse.

The urgency of climate action is underscored throughout this analysis. Without coordinated global efforts to accelerate clean energy deployment, improve efficiency, and phase out fossil fuels, emission trends will continue to exacerbate climate change impacts.

14. REFERENCES & DATA SOURCES

1. Our World in Data - CO2 and Greenhouse Gas Emissions Dataset
   Ritchie, H., Roser, M., & Rosado, P. (2024)
   https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions

2. Global Carbon Project - Annual Carbon Budget Reports
   Friedlingstein et al. (2023)
   https://www.globalcarbonproject.org/

3. IPCC Sixth Assessment Report (AR6)
   Intergovernmental Panel on Climate Change (2021-2023)

4. International Energy Agency - World Energy Outlook
   IEA (2024)

5. World Bank Development Indicators Database
   https://data.worldbank.org/