

Impact of CO_2 Emission on Climate Change and Global Life Expectancy

Muhammad Umar

Lecturer in Mathematics, Govt. Graduate College Layyah

Email: muhammadumar@ggclayyah.edu.pk

Muhammad Luqman Ilyas

Lecturer in Mathematics, University of Layyah Email: Luqmanilyas@gmail.com

Khalid Hussain

Govt. Graduate College Layyah Email: khalid.123bessnm.go@gmail.com

Khalid Masaud Ansari

Assistant Professor in Mathematics, Govt. Graduate College Layyah

Email: kmansari1215@gmail.com

Abdul Sattar Arshad

Govt. Graduate College Layyah Email: abdulsattararshad12@gmail.com

Muhammad Waqas

Govt. Graduate College Layyah Email: aliw62174@gmail.com

Abdul Hameed

Govt. Graduate College Layyah Email: anokhameedumeedu@gmail.com

Muhammad Zohaib

Govt. Graduate College Layyah Email: Zohaib78446@gmail.com

ABSTRACT

This research article explores the critical intersection of data science, global climate and public health, specifically investigating how escalating **CO_2 emissions** accelerate climate change and subsequently influence **global life expectancy**. Utilizing longitudinal data from the **World Bank Open Data** repository (spanning 1990 to the present), the study employs a combination of **Data Mining as data Wrangling, Exploratory Data Analysis (EDA)** and **machine learning techniques**, to identify patterns between environmental degradation and human health outcomes.

The analysis reveals a dual-threat mechanism: first, the direct impact of air pollution on respiratory and cardiovascular health; and second, the indirect consequences of climate instability—such as extreme weather events, food insecurity, and the spread of vector-borne diseases—which disproportionately lower life expectancy in

developing regions. A significant portion of this study is dedicated to a **comparative analysis of South Asia**, evaluating how its unique industrial trajectory and population density make it more vulnerable to these shifts compared to global averages.

Preliminary findings indicate a strong negative correlation between high per-capita carbon output and the rate of longevity improvement in emerging economies. The project concludes with data-driven **policy recommendations** aimed at South Asian governance, emphasizing the need for decoupled economic growth from carbon intensity to safeguard future public health.

Keywords: *CO₂ Emissions, Climate Change, Life Expectancy, Data Mining, Predictive Modeling.*

Introduction

The global climate crisis is increasingly defined by a measurable paradox: while industrialization has historically driven human progress, the resulting **CO₂ emissions** have created environmental feedback loops that now threaten to truncate human life. This project investigates the quantitative relationship between rising atmospheric carbon levels and **global life expectancy**, utilizing data-driven methodologies to move beyond theoretical speculation. By leveraging the **World Bank's** longitudinal datasets from 1990 to the present, this research treats life expectancy not just as a health metric, but as an indicator of a nation's environmental resilience. The study is grounded in the reality that carbon accumulation drives both direct physiological stressors—such as degraded air quality—and indirect ecological stressors, including food insecurity and the intensification of heat-related mortality.

From a data mining perspective, this investigation employs **Data Mining and Exploratory Data Analysis (EDA)** to identify non-obvious patterns in how carbon intensity impacts different economic tiers. While the global trend suggests a general increase in longevity due to medical advancement, the rate of this improvement shows a significant negative correlation with high per-capita emissions in industrializing zones. A core objective of this study is the **Comparative Analysis of South Asia**, a region characterized by rapid urban growth and high climate vulnerability. By juxtaposing South Asian indicators against global averages, we can determine if the region is experiencing a "longevity penalty"—where the health gains of economic development are being neutralized by the environmental costs of the energy sources powering that development.

The methodology integrates **Data Mining** to analysis longevity trends based on emission trajectories and **Exploratory Data Analysis** to categorize nations into distinct environmental-health profiles. This approach allows for a granular understanding of which countries are successfully decoupling carbon output from public health outcomes and which are falling into "**carbon traps**." Ultimately, the introduction of this project establishes that the fight against climate change is a fundamental necessity for maintaining human longevity. By providing an empirical bridge between environmental data and life expectancy, the following chapters offer a framework for **South Asian** policymakers to transition toward sustainable governance that prioritizes the biological survival of their populations.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Data Mining: Basic Definitions

Data Mining:

Data Mining is a process of discovering hidden patterns, correlations and actionable insights within large datasets of raw data.

Today, we stand at a crossroads. The very industrial processes that extended human life are now altering the Earth's climate so fundamentally that they threaten to shorten it. This chapter introduces the framework for investigating this paradox using the lens of

Data Mining and Environmental Analytics.

The Role of Data Mining in Environmental Science

In an era of "Big Data," the challenge is no longer a lack of information, but the inability to process it into actionable wisdom. Data Mining provides the tools to navigate this complexity.

Defining Data Mining and Knowledge Discovery (KDD)

Data Mining is the process of discovering useful information from large data repositories. It is the "engine" within the broader **Knowledge Discovery in Databases** process.

Core Methodologies for this Project

To analyze the relationship between CO_2 and health, we employ several specialized techniques:

Data Collection: first we get data from the data centers or websites like World Bank Data, Kaggle, Nasa and another such sources that collect the real data.

Data Cleaning: Any raw data have missing values, outliers, duplicate rows. It needs to cleaning process.

Exploratory Data Analysis (EDA): Using statistical summaries and visualizations to find initial patterns.

Data Transformation: For creating best patterns and insights, we transform, normalize and scale the data.

Regression Modeling: Specifically Linear and Non-Linear Regression to quantify how many "years of life" are lost per additional metric ton of CO_2 .

Time-Series Forecasting: Predicting future life expectancy trends based on historical emission growth rates.

Knowledge Presentation: Visualization and knowledge presenting techniques are used to present mined knowledge to users.

These steps are known as data preprocessing in Data Science field. We purify data for mining and creates trends from this preprocessed data.

Moving toward the Information Age

“We are living in the information age” is a popular saying; However, we are actually living the data age.

Data Mining turn a large collection of data into knowledge. A Google search engine collect the hundreds of millions of queries every day. Each query can be viewed as a transaction where the user describes her or his information need.

Data Mining can be viewed as a result of the natural evolution of information technology. The database and data management industry evolved in the development of several critical functionalities

Global Warming & Climate Change

Global warming is the long-term rise in Earth’s average temperature, primarily caused by human-driven emissions of greenhouse gases like carbon dioxide from burning fossil fuels. Climate change is a broader term covering this warming *plus* its side effects, such as melting glaciers, rising sea levels, and changing precipitation patterns.

Climate Change Impacts

Climate change does not affect human health in a single way; its impacts are multi-dimensional:

Thermal Stress: Increased global mean temperatures lead to "Heat Island" effects in cities, causing cardiovascular collapse in elderly populations.

Atmospheric Toxicity: CO_2 often acts as a proxy for other pollutants (PM2.5, Sox, NOx). These particles enter the bloodstream, leading to systemic inflammation and reduced life span.

Glaciers melting and flooding: Glaciers are rapidly melting due to rising the earth temperature and South Asia is the most effected region in the world.

Hydrological Disruptions: Rising sea levels and shifting rainfall patterns lead to water scarcity and crop failure, impacting the nutritional foundations of life expectancy.

The Mechanics of the CO_2 emissions

The warming of our planet by sunlight involves a precise energy exchange between the sun, the Earth’s surface, and atmospheric gases. Once this sunlight hits the Earth, the surface absorbs the energy and warms up. The Earth then attempts to cool itself by radiating energy back toward space. However, because the Earth is much cooler than the sun, it emits this energy as **longwave infrared radiation** (heat).

CO_2 molecules are "tuned" to the frequency of infrared waves. **When a co_2 molecule encounters this outgoing heat, it absorbs the energy, vibrates, and then re-emits it in all directions.** A significant portion of this heat is sent back down to the surface rather than escaping into space. As co_2 concentrations rise, more heat is recycled

downward, creating a thermal blanket that steadily increases global temperatures. This CO_2 emission is a major cause of increasing global earth temperature.

The Mechanics of the Greenhouse Effect

The Earth's temperature is a balance between the energy it receives from the sun and the energy, it radiates back into space. CO_2 acts as a "**thermal blanket**". While essential for life in small amounts, the anthropogenic (human-caused) surge in CO_2 concentration—moving from 280 ppm in the pre-industrial era to over 420 ppm today—has disrupted this equilibrium.

Problem Statement: The South Asian Vulnerability

While climate change is global, its effects are localized. **South Asia** (comprising Pakistan, India, Bangladesh, etc.) is currently the global epicenter of this crisis.

The Problem: South Asian nations are experiencing rapid urbanization with high emission density, but unlike Western nations, they lack the climate-resilient infrastructure to protect their citizens.

The Data Gap: Most global models are trained on Western data. There is an urgent need for a data-mining project that specifically targets the **South Asian** demographic to understand regional life expectancy threats.

Objectives and Scope

Primary Objectives

To curate a multi-source dataset (World Bank, WHO, NASA) spanning 1990–2025.

To perform comparative EDA between developed and developing regions.

To build a predictive algorithm that estimates the "Health Cost of Carbon."

Scope of the Research

The study focuses on 195 countries, with a deep-dive analysis of 8 South Asian nations. It considers CO_2 emissions per capita as the primary independent variable and Life Expectancy at birth as the primary dependent variable.

Data Scraping and Wrangling

Here, we will discuss about data mining about climate change and life expectancy. We will scrape datasets and mine them for making useful trends.

Data Scraping

This project utilizes extensive longitudinal data from the **World Bank Open Data repository**, specifically focusing on CO_2 emissions and Life Expectancy. By mining these global indicators, we analyze the statistical trade-offs between industrial development and human longevity, particularly within the vulnerable South Asian climate.

We opened the World Bank official Website and from home page downloaded the

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

World Development Indicators (WDI) file.

We used our vs code environment. Import the following libraries for reading and working on the dataset of WDI.

```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')

data = pd.read_csv('E:/Data_Mining_project/data/wdi.csv')
data.head()
```

output:

WDI Data Wrangling

Data wrangling for this project involves transforming the raw World Bank CSV/Excel files into a structured **Panel Data** format.

Skill Tools:

Visual Studio Code 2. Jupiter Notebook 3. python Language 4. Library: Pandas 5. Library: Numpy

Wrangling Steps (The Workflow)

Step 1: Filtering and Sub setting

We create a data frame by pandas (pd. Data Frame). This data frame contains with 403256 rows and 69 columns. We take useful indicators for our project.

```
data_collection = data.loc[((data['Indicator Name'] == 'Carbon dioxide (CO2)
emissions (total) excluding LULUCF (% change from 1990)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions (total) excluding LULUCF (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions excluding LULUCF per capita (t CO2e/capita)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Agriculture (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Building (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Fugitive Emissions (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Industrial Combustion (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Industrial Processes (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Power Industry (Energy) (Mt CO2e)'))]
```

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

```
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Transport (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Waste (Mt CO2e)')
| (data ['Indicator Code'] == 'EN.ATM.PM25.MC.M3')
| (data ['Indicator Code'] == 'EN.ATM.PM25.MC.ZS')
| (data ['Indicator Code'] == 'EN.ATM.PM25.MC.T1.ZS')
| (data ['Indicator Code'] == 'EN.ATM.PM25.MC.T2.ZS')
| (data ['Indicator Code'] == 'EN.ATM.PM25.MC.T3.ZS')
| (data ['Indicator Name'] == 'Life expectancy at birth,
female (years)')
| (data ['Indicator Name'] == 'Life expectancy at birth, male
(years)')
| (data ['Indicator Name'] == 'Life expectancy at birth, total
(years)')
| (data ['Indicator Name'] == 'Total greenhouse gas
emissions excluding LULUCF (% change from 1990)')
| (data ['Indicator Name'] == 'Total greenhouse gas
emissions excluding LULUCF per capita (t CO2e/capita)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Industrial Combustion (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Industrial Processes (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Power Industry (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Transport (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Waste (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Agriculture Sector (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Building Sector (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Fugitive Emissions Sector (Energy) (Mt CO2e)')
| (data ['Indicator Name'] == 'Carbon dioxide (CO2)
emissions from Other Sectors of Energy Use'))]
data_collection.shape
```

The raw WDI file contains the data from 1960 to 2023 which is not a complete data. we filter data from 1990 to 2023 and get useful rows.

```
# create a range from 1960 to 1989 with for loop
```

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

```
years = []
for year in range(1960, 1990):
    years.append(str(year))
print(years)

data_collection = data_collection.drop(columns= years)
data_collection = data_collection.drop(['Country Code','Indicator Code', '2024'],
axis=1)
```

We filter the meaningful indicators that we use in our project of “**CO₂ Emission changes the Climate and reduce the Global Life Expectancy**”.

```
data_filteration = data_collection.loc [(data_collection ['Indicator Name'] == 'Carbon
dioxide (CO2) emissions (total) excluding LULUCF (Mt CO2e)')
| (data_collection ['Indicator Name'] == 'Carbon
dioxide (CO2) emissions excluding LULUCF per capita (t CO2e/capita)')
| (data_collection ['Indicator Name'] == 'PM2.5 air
pollution, mean annual exposure (micrograms per cubic meter)')
| (data_collection ['Indicator Name'] == 'Total
greenhouse gas emissions excluding LULUCF (% change from 1990)')
| (data_collection ['Indicator Name'] == 'Total
greenhouse gas emissions excluding LULUCF per capita (t CO2e/capita)')
| (data_collection ['Indicator Name'] == 'Life
expectancy at birth, total (years)')
| (data_collection ['Indicator Name'] == 'Carbon
dioxide (CO2) emissions from Industrial Combustion (Energy) (Mt CO2e)')]
data_filteration.drop(columns=['Unnamed: 0'], inplace=True)
data_filteration.columns
```

Step 2: Handling Missing Values (Imputation)

Environmental data often has gaps (e.g., a country might miss CO₂ reporting for a specific year).

Action: we use Mean () function (filling gaps based on the trend) to handle missing points without losing the entire row.

```
# imputing the missing values with mean value
data_imputed = data_filteration.copy()
data_imputed.iloc[:, 2:] = data_imputed.iloc[:, 2:].apply(lambda x: x.fillna(x.mean()),
```

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

```
axis=1)
data_imputed = data_imputed.dropna(subset=data_imputed.columns[2:],how='all')
data_imputed.shape
```

Now we save this pure dataset in csv file.

```
data_imputed.to_csv('final_data.csv', index=False)
```

Step 3: Reshaping (Wide to Long)

Raw WDI data usually lists years (1990–2023) as individual columns. This is difficult for Analysis and Machine Learning.

- **Action:** Use the melt () function in Pandas to turn these year columns into a single "Year" variable.

```
df = pd.read_csv('E:/Data_Mining_project/data/final_data.csv')
df.head()
```

Melting the data from wide to long format

```
# create dataset with two indicators with different columns of indicator names like
merge them all data is same but indicator name divide into different columns
co2 = df.loc[df['Indicator Name'] == 'Carbon dioxide (CO2) emissions excluding
LULUCF per capita (t CO2e/capita)']
co2_total = df.loc[df['Indicator Name'] == 'Carbon dioxide (CO2) emissions (total)
excluding LULUCF (Mt CO2e)']
life = df.loc[df['Indicator Name'] == 'Life expectancy at birth, total (years)']
pm25 = df.loc[df['Indicator Name'] == 'PM2.5 air pollution, mean annual exposure
(micrograms per cubic meter)']
ghg = df.loc[df['Indicator Name'] == 'Total greenhouse gas emissions excluding
LULUCF per capita (t CO2e/capita)']
ghg_total = df.loc[df['Indicator Name'] == 'Total greenhouse gas emissions excluding
LULUCF (% change from 1990)']
consumption = df.loc[df['Indicator Name'] == 'Carbon dioxide (CO2) emissions from
Industrial Combustion (Energy) (Mt CO2e)']
# melt these three datasets on country name and year
cols_to_melt = [col for col in df.columns if col.startswith('19') or col.startswith('20')]
# adjust the condition for melting above three datasets
co2_melted = co2.melt(id_vars=['Country Name', 'Indicator Name'],
                      value_vars=cols_to_melt,
                      var_name='Year',
```

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

```
        value_name='CO2_value')
co2_total_melted = co2_total.melt(id_vars=['Country Name', 'Indicator Name'],
                                value_vars=cols_to_melt,
                                var_name='Year',
                                value_name='co2_total')
life_melted = life.melt(id_vars=['Country Name', 'Indicator Name'],
                        value_vars=cols_to_melt,
                        var_name='Year',
                        value_name='Life_value')
pm25_melted = pm25.melt(id_vars=['Country Name', 'Indicator Name'],
                        value_vars=cols_to_melt,
                        var_name='Year',
                        value_name='PM2.5_value')
ghg_melted = ghg.melt(id_vars=['Country Name', 'Indicator Name'],
                      value_vars=cols_to_melt,
                      var_name='Year',
                      value_name='GHG_value')
ghg_total = ghg_total.melt(id_vars=['Country Name', 'Indicator Name'],
                           value_vars=cols_to_melt,
                           var_name='Year',
                           value_name='ghg_total')
consumption_melted = consumption.melt(id_vars=['Country Name', 'Indicator
Name'],
                                       value_vars=cols_to_melt,
                                       var_name='Year',
                                       value_name='consumption_value')
# merge these three datasets on country name and year
merged_df = pd.merge(co2_melted, life_melted, on=['Country Name', 'Year'],
                     suffixes=('_co2', '_life'))
merged_df = pd.merge(merged_df, pm25_melted, on=['Country Name', 'Year'],
                     suffixes=('_co2_life', '_pm25'))
merged_df = pd.merge(merged_df, co2_total_melted, on=['Country Name', 'Year'],
                     suffixes=('_co2_life_pm25', '_ghg'))
merged_df = pd.merge(merged_df, ghg_melted, on=['Country Name', 'Year'],
                     suffixes=('_co2_life_pm25_ghg', '_ghg_total'))
merged_df = pd.merge(merged_df, ghg_total, on=['Country Name', 'Year'],
                     suffixes=('_co2_life_pm25_ghg_total', '_consumption'))

merged_df = pd.merge(merged_df, consumption_melted, on=['Country Name',
'Year'], suffixes=('_co2_life_pm25_ghg_total_consumption', '_consumption'))

# rename the columns to country, year, co2_indicator, co2_value, life_indicator,
life_value, pm25_indicator, pm25_value, ghg_indicator, ghg_value,
ghg_total_indicator, ghg_total_value, consumption_indicator, consumption_value
```

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

merged_df.shape

Use the rename method to update the Data Frame

```
# 1. Create a dictionary mapping the old names to the new names
column_mapping = {
    'Indicator Name_co2': 'CO2_Indicator',
    'CO2_value': 'CO2_Value',
    'Indicator Name_life': 'Life_Expectancy_Indicator',
    'Life_value': 'Life_Expectancy_Value',
    'Indicator Name_co2_life_pm25': 'PM25_Indicator',
    'PM2.5_value': 'PM2.5_Value',
    'Indicator Name_ghg': 'Total_CO2_Emissions_Indicator',
    'co2_total': 'CO2_Total',
    'Indicator Name_co2_life_pm25_ghg_total': 'GHG_Indicator',
    'GHG_value': 'GHG_Value',
    'Indicator Name_consumption': 'GHG_Total_Indicator',
    'ghg_total': 'GHG_Total_Consumption',
    'Indicator Name': 'Industrial_Combustion_Indicator',
    'consumption_value': 'Consumption_Value'
}

# 2. Use the rename method to update the DataFrame
# Replace 'df' with the name of your actual DataFrame variable
merged_df.rename(columns=column_mapping, inplace=True)
data = merged_df[['Country Name', 'Year', 'CO2_Indicator', 'CO2_Value',
'Life_Expectancy_Indicator', 'Life_Expectancy_Value', 'PM25_Indicator',
'PM2.5_Value', 'Total_CO2_Emissions_Indicator', 'CO2_Total', 'GHG_Indicator',
'GHG_Value', 'GHG_Total_Indicator', 'GHG_Total_Consumption',
'Industrial_Combustion_Indicator', 'Consumption_Value']]
data.head()
```

Step 4. The "Best" Wrangled Structure (Visual Representation)

A high-standard research dataset should look like a data frame.

Countr y Name	Year	CO2_In dicator	CO2_ Value	Life_Expect ancy_Indica tor	Life_Expectanc y_Val ue	PM25_ Indicat or	PM 2.5_ Val ue	Total_ CO2_E mission s_Indic ator	CO2_Tota l	GHG_ Indi cator	GHG_ Val ue	GHG_ Tota l_Ind icator	GHG_ Total_ Consumption	Industrial_ Combustion_ Indicator	Consumption_Value	
Africa Eastern and Souther n	1990	Carbon dioxide (CO2) emission s excludin g LULUC...	1.232 869	Life expectancy at birth, total (years)	51.09 6326	PM2.5 air pollutio n, mean annual exposur e (mic...	25.1 113 99	Carbon dioxide (CO2) emissio ns (total) exclud e...	Carbon dioxide (CO2) emissions (total) excludi...	378.5 044	Total green house gas emiss ions exclu ding LUL UC...	2 · 6 7 3 4 7 2	Total greenhouse gas emissions excluding LULUC...	42.315217	Carbon dioxide (CO2) emissions from Industrial...	68.4 542

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Country Name	Year	CO2_Indicator	CO2_Value	Life_Expectancy_Indicator	Life_Expectancy_Value	PM25_Indicator	PM2.5_Value	Total_CO2_Emissions_Indicator	CO2_Total	GHG_Indicator	GHG_Value	GHG_Total_Indicator	GHG_Total_Consumption	Industrial_Combustion_Indicator	Consumption_Value
Africa Western and Central	1990	Carbon dioxide (CO2) emissions excluding LULUC...	0.490402	Life expectancy at birth, total (years)	48.459295	PM2.5 air pollution, mean annual exposure (mic...	60.240684	Carbon dioxide (CO2) emissions (total) exclude...	102.7717	Total greenhouse gas emissions excluding LULUC...	2.193276	Total greenhouse gas emissions excluding LULUC...	53.082695	Carbon dioxide (CO2) emissions from Industrial...	5.656

Comparative Exploratory Data Analysis for making trends

This World Development Indicators (WDI) EDA examines global socio-economic trends through **univariate and multivariate analysis**. By cleaning missing values and normalizing scales, we identify correlations between **life expectancy** and **CO₂ emissions**. Visualizations like time-series plots and heatmaps reveal regional disparities and long-term development trajectories.

Exploratory Data Analysis (EDA)

In Exploratory Data Analysis, we make plots for the comparisons and relations between different indicators.

Skill Tools:

Vs code

Jupyter notebook

Pandas

Numpy

Matplotlib

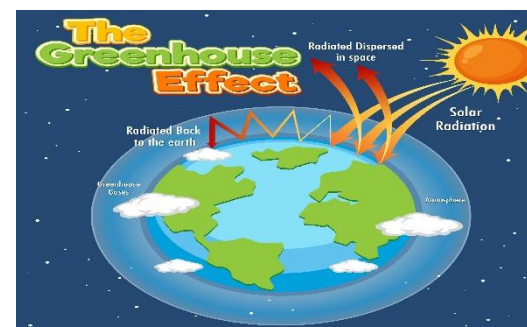
Seaborn

Plotly

Climate Change Main Factors

Climate change is a complex puzzle, but the "main characters" driving it are fairly well-established. It essentially comes down to the **Greenhouse Effect**: certain gases trap heat in our atmosphere that would otherwise escape into space.

Here are the primary factors driving the current shift in our climate:



Greenhouse Gas Emissions (The "Big Three")

Carbon Dioxide (CO₂) 76%: The primary contributor, mostly from burning fossil fuels (coal, oil, and natural gas) for energy and transportation.

Methane (CH₄) 16%: Much more potent than CO₂ at trapping heat in the short term. It comes from livestock (cow burps/manure), landfills, and natural gas leaks.

Nitrous Oxide (N₂O) 6%: Primarily released through agricultural practices, specifically the use of synthetic fertilizers.

Deforestation and Land Use

Industrial Processes & Agriculture

Feedback Loops (Nature's Multipliers)

Green House Gases is the major part that play a vital role in climate change. In Green House Gases, CO₂ emission is **75-76%** which is major part. Here we will discuss CO₂ emissions in detail.

CO₂ Emissions in the World Regions

Carbon dioxide is the primary driver of climate change, accounting for approximately 76% of global greenhouse gas emissions. Primarily released through burning fossil fuels and deforestation, CO₂ traps solar heat in the atmosphere. This "greenhouse effect" destabilizes global temperatures, causing rising sea levels, extreme weather, and long-term environmental shifts.

Nasa Report about CO₂ emission

Ground-based and satellite measurements have measured a sharp increase in atmospheric CO₂ in recent decades. With a more than 50% increase in CO₂ since the start of industrial times in the 18th century, scientific measurements have linked this increase primarily to the burning of coal, oil, and gas. This means that CO₂ levels are now 150% the value they were at in 1750 — this is at least **1.5 times greater** than the natural increase observed at the end of the last ice age **20,000** years ago, which happened over thousands of years.

The animated map below shows the historical changes in global carbon dioxide over time. Note the colors change as the amount of CO₂ rises from **365 parts per million (ppm) in 2002 to over 420 ppm in 2022**. These measurements from the AIRS instrument come from the mid-troposphere, the layer of Earth's atmosphere that is 8 to 12 kilometers (about 5 to 7 miles) above the ground.

Total CO₂ emission:

According to Nasa report, CO₂ concentration reached at **429.61 ppm** in 17th June 2025. This concentration is also rising the global earth temperature and making a highest change in the climate.

From World Bank Dataset wrangling, we create a plot of Total CO₂ emission by regions of the globe.

Global regional rate of Total CO₂ emission:

East Asia & Pacific: This region shows the most increase, surpassing North America around 2000. Emissions climbed from approximately **≈5,000 million tons** in 1990 to over **≈17,000 million tons** by 2023, reflecting rapid industrialization.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

North America: After 1990s, emissions peaked in the mid-2000s (near **≈5,400 million tons**) and have since shown a **gradual decline**, ending 2023 at roughly **≈5,300 million tons**.

South Asia: While starting at a low base, this region has seen consistent, steady growth, nearly quadrupling its output over the 33-year period. Its emission increased from **≈621 million tons** (in 1990s) to **≈3123 million tons** (in 2023).

Latin America & Sub-Saharan Africa: These regions remain the lowest emitters in this dataset. Both show slight upward trends but remain well below the 2,500 million tons mark.

The "2020 Dip": A visible downward notch is present in all regions in the year **2020**, likely corresponding to the global reduction in industrial activity and travel during the COVID-19 pandemic.

Code#

```
import pandas as pd
import plotly.express as px
```

```
fig = px.line(regions, x = 'Year', y = 'CO2_Total', color='Country Name', title='Total
CO2 emissons by World Regions')
fig.update_layout(
    plot_bgcolor='lightblue',
    xaxis_title_font_size=15,
    yaxis_title_font_size=15,
    title_font_size=20
)
fig.update_xaxes(title_text='Year')
fig.update_traces(line=dict(width=3))
fig.update_yaxes(title_text='Total CO2 Emissions (Million Tonnes)')
fig.show()
```

Online ISSN

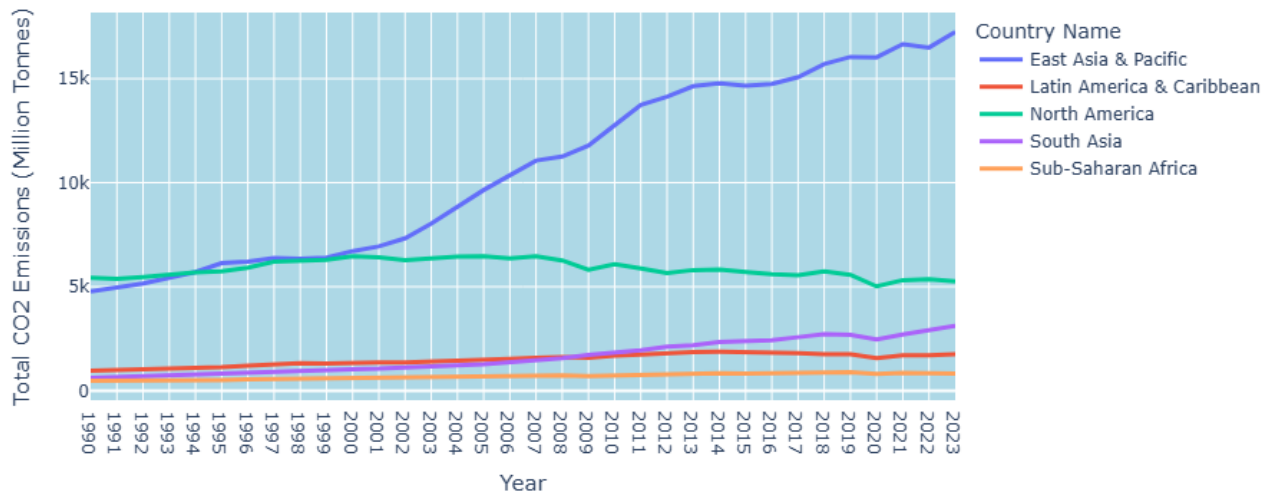
3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Total CO2 emissions by World Regions



CO₂ emission per capita:

North America: Historically the highest emitter, starting at nearly **20 tons per capita**. Despite high levels, it shows a significant downward trend, dropping to approximately **≈14 tons** by 2023.

East Asia & Pacific: Exhibits the fastest growth, more than doubling from **2.5 to over 7 tons per capita**, reflecting rapid industrialization and lifestyle changes.

Latin America & Caribbean: Maintains a relatively stable profile, hovering consistently between **2 and 3 tons per capita** over the three decades.

South Asia: Shows gradual, steady growth but remains low overall, increasing from under **1 tone to nearly 2 tons per capita**.

Sub-Saharan Africa: Remains the lowest emitter throughout the period, with emissions stagnant at **less than 1 tone per capita**.

Code#

```
fig = px.line(regions, x = 'Year', y = 'CO2_Value', color='Country Name', title='Total CO2 emissions by World Regions')
fig.update_layout(
    plot_bgcolor='lightblue',
    xaxis_title_font_size=15,
    yaxis_title_font_size=15,
    title_font_size=20,
)
```

Online ISSN

3007-3197

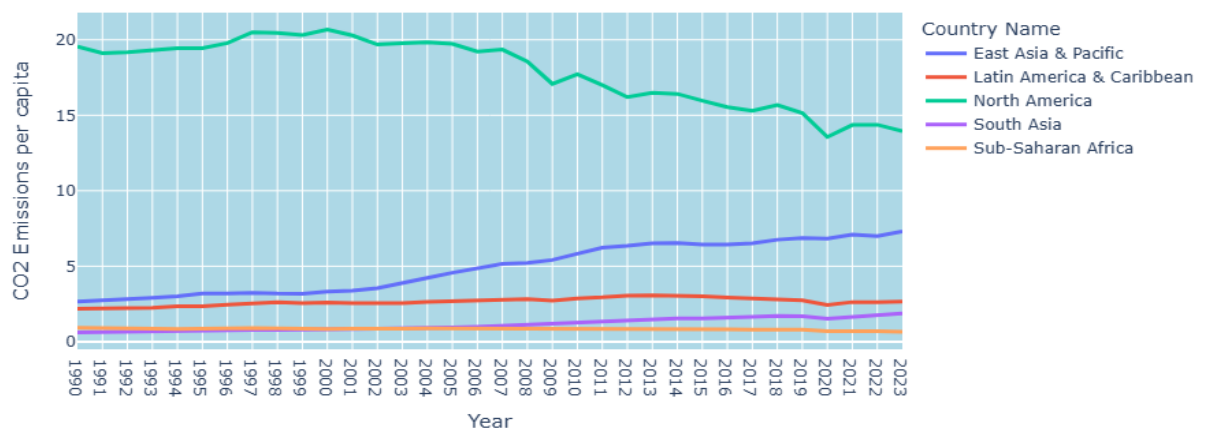
Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

```
# y axis title is total co2 emissions in million tonnes and x axis title is year and y
values in million tonnes
fig.update_xaxes(title_text='Year')
fig.update_yaxes(title_text='CO2 Emissions per capita')
fig.update_traces(line=dict(width=3))
fig.show()
```

Total CO2 emissions by World Regions



High value emission Countries:



China: 13,000 +plus million metric tons in 2023.
Accounts for nearly one-third of global emissions, primarily driven by **coal**, which provides 58% of its energy.



United States: 4,680 million metric tons in 2023.
Major sources include **transportation**, electric power, and industrial manufacturing.



India: ≈3,000 million metric tons.
Rapidly growing emissions tied to **coal-fired power**, which supplies 44% of its energy needs.



Russia: 2,069 million metric tons.
Heavily reliant on its massive **natural gas** deposits for power and export.

Green House Gases

Greenhouse gases, primarily **carbon dioxide**, **methane**, and **nitrous oxide**, trap heat in the atmosphere. This process, driven largely by fossil fuel

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

combustion and deforestation, causes **global warming** and severe climate change. Major production sources include **fossil fuel combustion** for electricity, industrial manufacturing, **deforestation**, and **agricultural practices** like livestock farming and fertilizer use.

East Asia & Pacific (≈200%): The region driver of highest global increases. Rapid industrialization in China and Indonesia and Vietnam.

South Asia (≈190%): The fastest-growing segment in recent decades. Driven by India's infrastructure expansion and energy demand, this region's purple band widens sharply after 2005.

Sub-Saharan Africa (≈80%): Maintains a consistent upward trajectory. While its fossil fuel footprint is lower, agricultural expansion and land-use changes contribute to its steady growth.

Latin America & Caribbean (≈60%): Shows moderate growth. Emissions here are heavily tied to the **AFOLU** sector (Agriculture and Forestry) rather than heavy industry.

North America (≈10%): The smallest relative increase. Despite being a high absolute emitter, a shift from coal to natural gas and renewables has led to a plateauing trend compared to the 1990 baseline.

```
fig = px.area(regions, x = 'Year', y = 'GHG_Total_Consumption',
              color='Country Name',
              title='Total Green House Gasses emissson by World Regions')
fig.update_layout(
    plot_bgcolor='white',
    xaxis_title_font_size=15,
    yaxis_title_font_size=15,
    title_font_size=20
)
# title is total ghg emissions in million tonnes
fig.update_xaxes(title_text='Year')
fig.update_yaxes(title_text='Total Green House Gas Emissions (% change from 1990)')
fig.show()
```

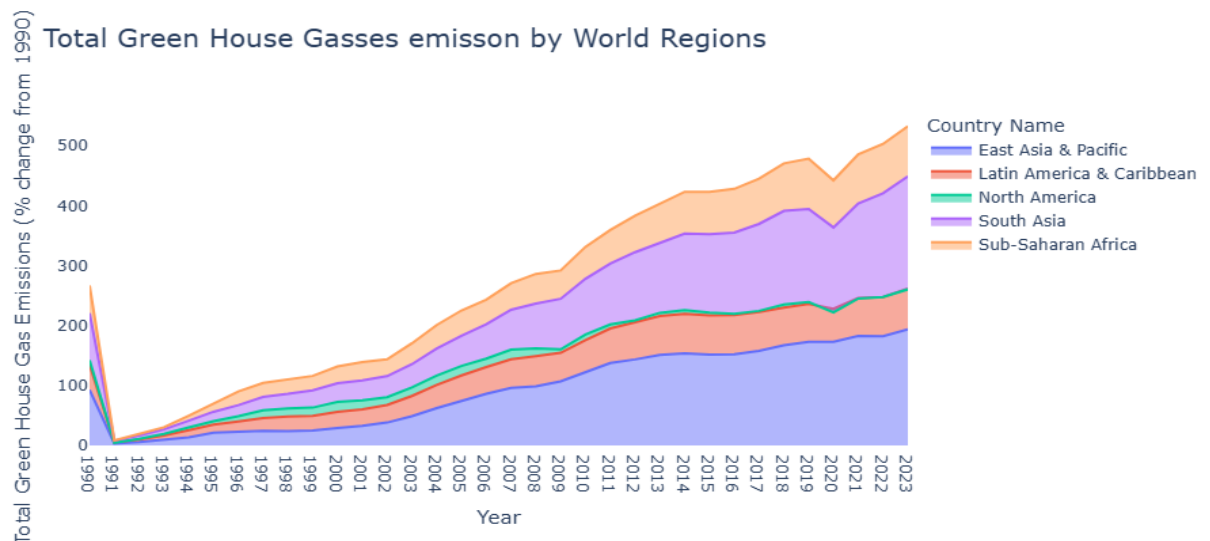
Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>



The South Asian Growth Story:

Rapid Industrialization: This coincides with **India's** rise as a global economic power, where energy demand for manufacturing and urban development exploded.

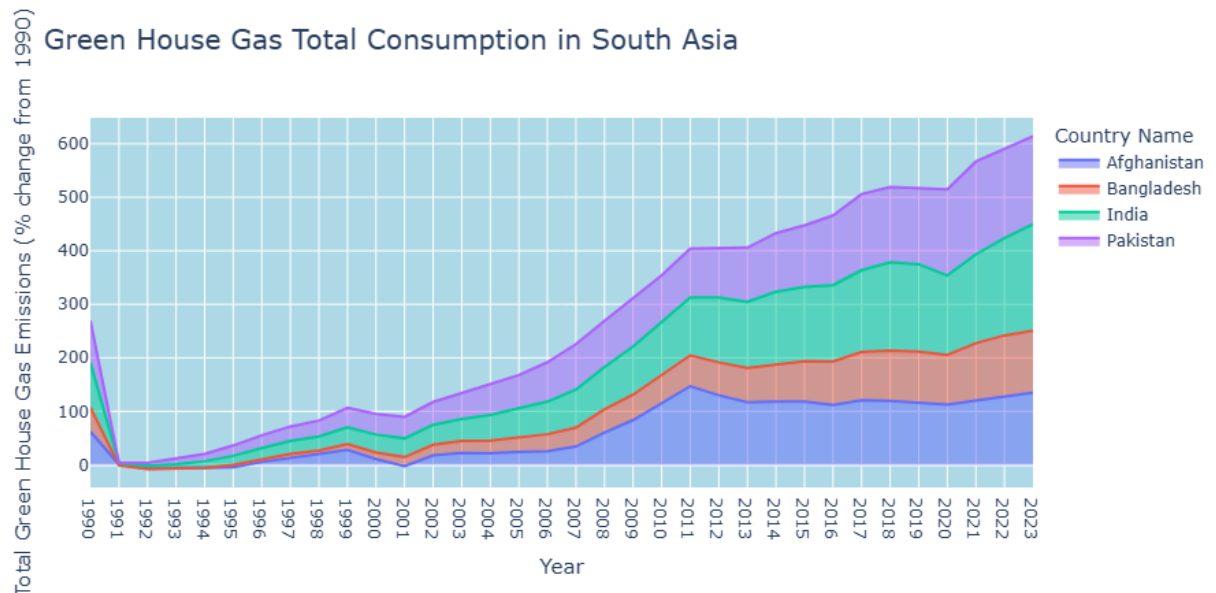
Coal & Power: South Asia remains heavily reliant on **coal**. Coal-fired power plants are the single largest source of CO₂ in this region, providing the cheap energy needed to power its **1.9 billion people**.

Agricultural Footprint: Beyond fossil fuels, South Asia is a global hotspot for **methane** and **nitrous oxide**. This is due to its vast agricultural sector—specifically rice paddy cultivation and the world's largest livestock populations, which emit significant amounts of non-CO₂ greenhouse gases.

GHG emission in 2023, India = 198.86%, Pakistan = 164.21%, Afghanistan = 134.52%, Bangladesh = 116.28% (% change from 1990).

Code#

```
fig = px.area(df_south_asia, x='Year', y='GHG_Total_Consumption', color='Country
Name', title='Green House Gas Total Consumption in South Asia')
fig.update_layout(
    plot_bgcolor='lightblue',
    xaxis_title_font_size=15,
    yaxis_title_font_size=15,
    title_font_size=20
)
fig.update_yaxes(title_text='Total Green House Gas Emissions (% change from
1990)')
fig.show()
```



Global Air Pollution Trends (1990–2023)

World Air Quality Index: Regional Comparison

South Asia (Highest Pollution): This region consistently records the worst air quality. After remaining stable near $60 \mu\text{g}/\text{m}^3$ for 15 years, it saw a massive surge starting in 2010, peaking at nearly $78 \mu\text{g}/\text{m}^3$ around 2014. Despite a sharp dip during the 2020 pandemic lockdowns, it remains the most polluted, ending 2021 at approximately $61 \mu\text{g}/\text{m}^3$.

Sub-Saharan Africa & East Asia: Sub-Saharan Africa's pollution spiked 25% (2014–2016) to $46 \mu\text{g}/\text{m}^3$ before stabilizing near $38 \mu\text{g}/\text{m}^3$. East Asia improved 35% from its 2013 peak ($44 \mu\text{g}/\text{m}^3$) to a 2020 low ($29 \mu\text{g}/\text{m}^3$). Both regions saw a $\approx 30\%$ rebound post-pandemic, ending 2021 at approximately $39 \mu\text{g}/\text{m}^3$.

Latin America & Caribbean: This region has maintained a gradual downward trajectory, improving by roughly 17% from its 1990 levels before a slight uptick in the final years.

North America (Lowest Pollution): Representing the "cleanest" end of the spectrum. It dropped from roughly $14 \mu\text{g}/\text{m}^3$ in 1990 to its lowest point of about $7 \mu\text{g}/\text{m}^3$ in 2020, a total reduction of 50% over three decades. **In 2023, the air in South Asia is roughly 5.5 times more polluted than in North America.**

Code#

```
# line plot of life vs pm25 for each group
fig = px.line(regions , x='Year', y='PM2.5_Value', color='Country Name',
              title='Air Polution (Air Quality Index) of world regions')
```

Online ISSN

3007-3197

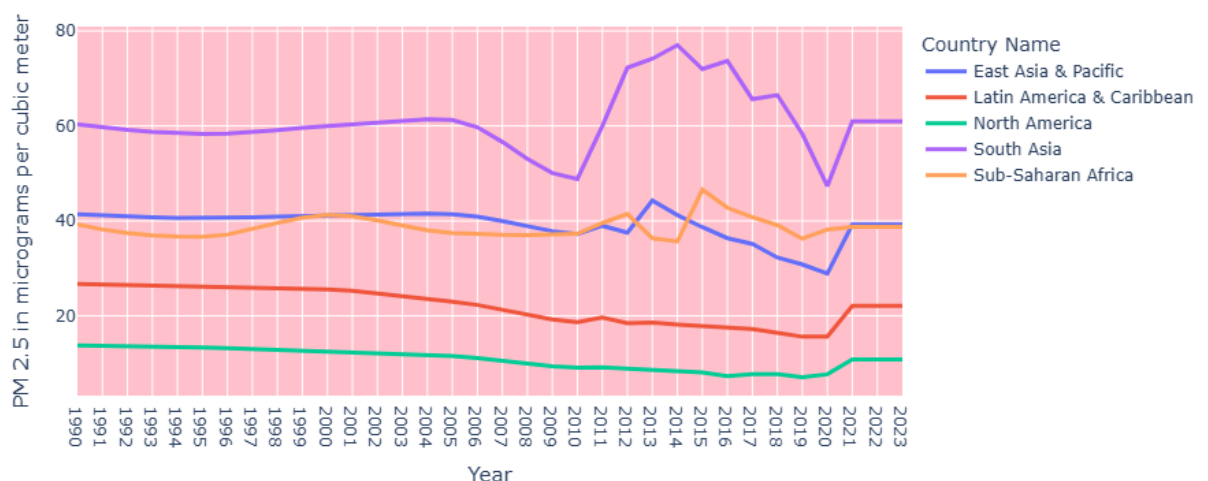
Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

```
# make more beautiful
fig.update_layout(
    plot_bgcolor='pink',
    xaxis_title_font_size=15,
    yaxis_title_font_size=15,
    title_font_size=20
)
fig.update_xaxes(title_text='Year')
fig.update_traces(line=dict(width=3))
fig.update_yaxes(title_text='PM 2.5 in micrograms per cubic meter')
fig.show()
```

Air Pollution (Air Quality Index) of world regions



South Asia Breathless Borders: The Common Air Quality crises in India, Pakistan, Bangladesh, and Afghanistan.

This region is home to approximately **25% of the global population** and suffers from some of the highest pollution burdens worldwide.

Regional Trends and Country Analysis

South Asian air quality is characterized by long-term stability followed by extreme volatility after 2010.

India: Consistently among the most polluted, India peaked at nearly **80 $\mu\text{g}/\text{m}^3$** in 2014.

Pakistan: Shows a trend closely mirrored to India due to shared "airsheds". In 2024, its average concentration stood at **58.55 $\mu\text{g}/\text{m}^3$** , nearly **15 times higher** than WHO safe limits.

Online ISSN

3007-3197

Print ISSN

3007-3189

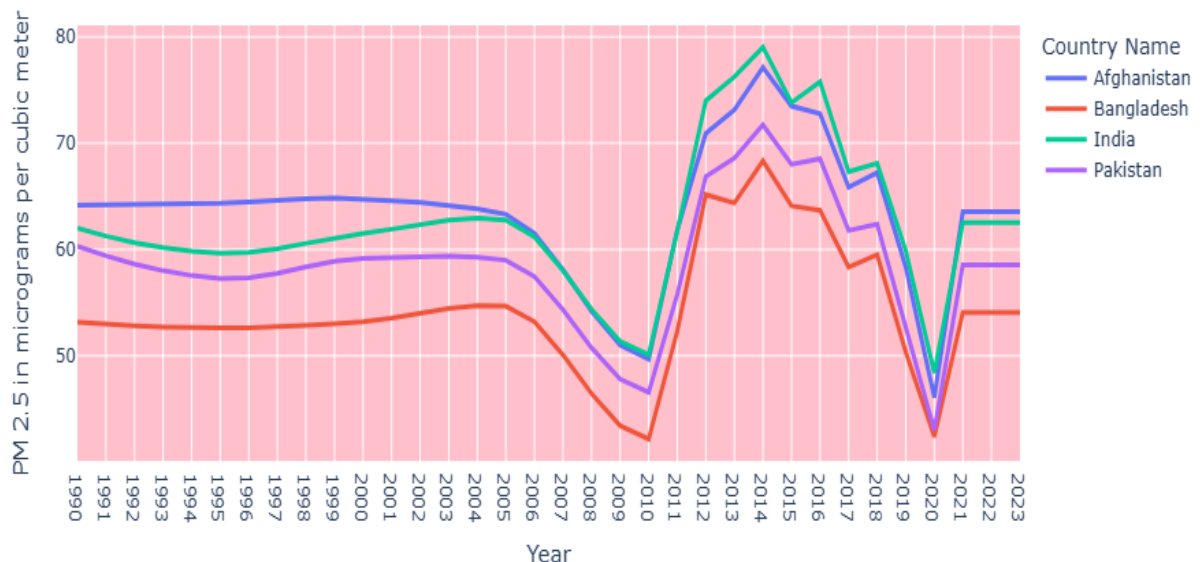
<http://amresearchreview.com/index.php/Journal/about>

Afghanistan: Traditionally recorded the highest levels in the 1990s (around $64 \mu\text{g}/\text{m}^3$). While it saw a decline toward 2010, it spiked alongside its neighbors in the mid-2010s.

Bangladesh: While slightly lower than its neighbors, it recorded no "good" air quality months in 2021 with $58.55 \mu\text{g}/\text{m}^3$.

```
fig = px.line(df_south_asia , x='Year', y='PM2.5_Value', color='Country Name',
              title='Air Polution (Air Quality Index) in South Asia')
fig.update_layout(
    plot_bgcolor='pink',
    xaxis_title_font_size=15,
    yaxis_title_font_size=15,
    title_font_size=20
)
fig.update_xaxes(title_text='Year')
fig.update_traces(line=dict(width=3))
fig.update_yaxes(title_text='PM 2.5 in micrograms per cubic meter')
fig.show()
```

Air Polution (Air Quality Index) in South Asia



In South Asia:

Primary Causes: Key drivers include **residential fuel combustion** (cooking/heating with biomass), unregulated **brick kilns**, and **vehicular emissions** using low-grade fuel. Seasonal **crop residue burning** and stable winter atmospheric conditions further trap pollutants.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Health & Economic Effects: Air pollution causes an estimated **2 million premature deaths** in South Asia annually. Economically, the region loses between **4% and 11% of its GDP** due to healthcare costs and lost productivity.

Conclusion: CO₂ emission with total and per capita, Green House Gas emission and increase in PM2.5 value changing the global Earth Temperature and climate of our planet. That is a global threat for us.

Facts about Global Climate Change and Life Expectancy

Since the mid-19th century, Earth's global surface temperature has risen by approximately **1.1°C to 1.3°C** relative to the pre-industrial average (1850–1900). While early 20th-century warming was gradual, the pace has accelerated dramatically since 1970.

The last decade (2015–2024) has been the warmest on record, with 2024 recently setting a new peak. Since 1982, the warming rate has tripled to roughly **0.20°C per decade**. Driven primarily by human-produced greenhouse gas emissions, this trend has led to shrinking ice sheets, rising sea levels, and more frequent extreme weather events, signaling a profound shift in Earth's energy balance.

Global Earth's surface Temperature:

The Earth's surface temperature has shifted from an average of roughly **27.6°C in 1990** to nearly **28.6°C by 2023**. This represents a significant warming trend of approximately **1.0°C**, over just three decades.

Since 2014, global temperatures have surged, with recent years consistently exceeding **1.1°C** above pre-industrial levels. This rapid **1-degree** shift within your data reflects accelerating greenhouse gases.

```
temp = pd.read_csv('E:/Data_Mining_project/data/temperature_data.csv')
temp.head()
```

```
regional_temp = temp.groupby(['Indicator Name',
'Year'])['Temperature'].mean().reset_index()
regional_temp['Year'] = regional_temp['Year'].astype(int)
fig = px.line(regional_temp, x='Year', y='Temperature', color='Indicator Name',
              title='Average Temperature of World in Degrees Celsius')
fig.update_traces(line=dict(width=3), marker=dict(size=8))
fig.update_layout(legend_title_text = 'Region', plot_bgcolor='white')
fig.show()
```

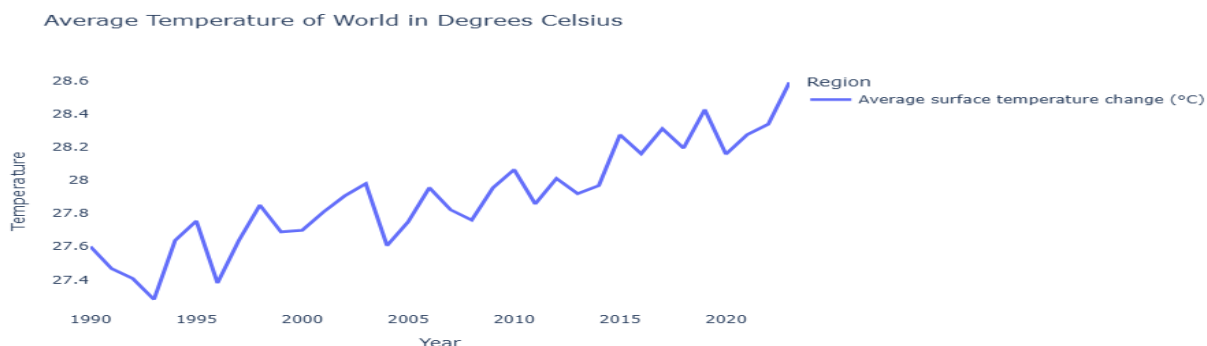
Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>



United Nations Report:

IPCC Sixth Assessment Report (AR6)

The Working Group I report is the first instalment of the IPCC's Sixth Assessment Report (AR6), which will be completed in 2022. The report provides new estimates of the chances of crossing the global warming level of 1.5°C in the next decades, and finds that unless there are immediate, rapid and large-scale reductions in greenhouse gas emissions, limiting warming to close to 1.5°C or even 2°C will be beyond reach.

The report shows that emissions of greenhouse gases from human activities are responsible for approximately 1.1°C of warming since 1850-1900, and finds that averaged over the next 20 years, global temperature is expected to reach or exceed 1.5°C of warming. This assessment is based on improved observational datasets to assess historical warming, as well progress in scientific understanding of the response of the climate system to human-caused greenhouse gas emissions.

Global warming refers specifically to the long-term rise in Earth's average surface temperature, primarily driven by the "greenhouse effect." Human activities—such as burning fossil fuels and deforestation—increase the concentration of gases like CO₂ and CH₄, which trap heat in the atmosphere.

Climate change plays the major role of the "broad consequence." While global warming is the symptom (rising heat). It encompasses shifting weather patterns, rising sea levels, and the increasing frequency of extreme events like droughts and intensified hurricanes. Essentially, global warming acts as the primary engine, while climate change represents the systemic disruptions that follow.

Together, they threaten biodiversity, food security, and human habitats, requiring urgent global shifts toward renewable energy and sustainable practices to mitigate further damage.

The "Climate Change" Effect: A South Asian Perspective

The effect of these emissions is not uniform. South Asia is warming **twice as fast** as the global average, creating a unique set of challenges.

South Asia is one of the world's most climate-vulnerable regions. Rising temperatures are accelerating **Himalayan glacier melt**, threatening water security for billions. Erratic monsoons cause devastating cycles of **extreme flooding** and prolonged droughts, crippling agriculture-dependent economies. Additionally, rising sea levels pose an existential threat to low-lying areas like Bangladesh and the Maldives.

These shifts drive mass **climate-induced migration**, intensify health risks, and could slash regional GDP by nearly 9% by 2100.

More details about South Asia climate change:

The video "Climate change: Impact on South Asia " was produced and published by WION (World Is One News) on December 3, 2023. WION is an Indian News Channel.

Pakistan: A Case of Extreme Vulnerability

Pakistan is consistently ranked among the **top 10 most vulnerable countries** on the Global Climate Risk Index.

Thermal Stress: Cities like Jacobabad and Sibi have recorded temperatures exceeding **50°C**, pushing the human "wet-bulb" temperature to its limit.

Hydrological Extremes: The **2022 floods**, exacerbated by CO₂ induced **glacial melt** and intense **monsoons**, affected 33 million people. Such disasters cause immediate mortality and destroy the healthcare infrastructure necessary to maintain life expectancy.

The "Asian Water Tower" Crisis

The melting of the **Himalayan** and **Hindu Kush glaciers** (often called the Third Pole) threatens the water security of over **1.5 billion people**. As these glaciers recede due to high carbon concentrations, the resulting "boom and bust" cycle of **floods** and droughts directly impacts **nutrition** and **sanitation**—the two pillars of longevity.

Infectious Disease Expansion: In Pakistan, the post-flood periods of 2022-2024 saw unprecedented spikes in **Malaria, Dengue, and Cholera**.

Pakistan Case of Study:

Add detailed sections on the **2022 Floods** impact on infant mortality and the **Smog Crisis** in Punjab.

Pakistan's Crisis:

In cities like Lahore, the convergence of industrial CO₂, crop burning, and vehicular emissions creates a permanent smog layer during winter. This reduces the average life expectancy of a Lahori citizen by roughly **7 years** compared to a citizen in a clean-air city like Zurich a Switzerland city.

Why South Asia is the Target?

South Asia is uniquely vulnerable due to three "Aggravators":

High Population Density: More people are exposed to every ton of CO₂ emitted.

Glacial Proximity: Unlike many other regions, South Asia relies on the melting Himalayas, which are highly sensitive to CO₂-induced warming.

Agricultural Dependence: A slight shift in the Monsoon (caused by climate change) can lead to mass malnutrition, a primary driver of low life expectancy.

Mitigation: Addressing the Root Causes

Transition to Renewables: South Asian nations must accelerate the shift from coal to solar and wind. In Pakistan, the **Renewable Energy Policy 2026** targets a 60% clean energy mix by 2030 to drastically cut CO₂ and PM2.5 levels.

Energy Efficiency: Retrofitting industrial units in cities like **Karachi** and **Faisalabad** with high-efficiency burners to reduce carbon intensity per unit of production.

Methane Management: Implementing "Green Agriculture" to reduce methane and CO₂ from livestock and rice paddies, which are significant in the Indus Basin.

Adaptation: Building Resilience in Pakistan

As a "frontline" state, Pakistan's focus is on survival and recovery.

Nature-Based Solutions: Mangrove restoration in the Indus Delta and the "**Ten Billion Tree Tsunami**" expansion to create natural carbon sinks and cooling zones.

Climate-Resilient Infrastructure: Rebuilding flood-proof hospitals and schools in Sindh and Balochis tan to ensure health services remain uninterrupted during disasters.

Key Focus Points for Control and Implementation

Effective control requires "Precision Governance."

Urban Planning (Sponge Cities): Developing cities that can absorb heavy rainfall to prevent urban flooding and water-borne disease outbreaks.

Air Quality Governance: Implementing the "**Smog Emergency Protocol**" across the Punjab region, including strict vehicular emission testing and industrial "Green Zones."

Water Security: Protecting the Indus River system from CO₂-induced glacial melt through better reservoir management and desalinization projects.

References

World Bank Open Data. (2024). *CO₂ emissions (metric tons per capita)*. Retrieved from <https://datatopics.worldbank.org/world-development-indicators/>

World Bank Open Data. (2024). *Life expectancy at birth, total (years)*. Retrieved from <https://datatopics.worldbank.org/world-development-indicators/>

World Health Organization (WHO). (2023). *Ambient (outdoor) air pollution*. Retrieved from [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)

Air Quality Life Index (AQLI). (2023). *The Invisible Killer: How Air Pollution Shortens Life Expectancy*. Energy Policy Institute at the University of Chicago (EPIC).

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

NASA Global Climate Change. (2024). *Vital Signs of the Planet: Carbon Dioxide.*

Retrieved from: <https://science.nasa.gov/earth/explore/earth-indicators/carbon-dioxide/>

Intergovernmental Panel on Climate Change (IPCC). (2023). *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report.*

<https://www.ipcc.ch/assessment-report/ar6/?hl=en-PK>

United Nations Environment Program (UNEP). (2023). *Emissions Gap Report 2023: Broken Record – Temperatures hit new highs, yet world fails to cut emissions (again).*