Advanced Statistics Project
Global Emissions

Anit Upadhyaya, Udhav Shankar, Vishal Paudel

Today's Agenda

- In this presentation, we'll be going over the following:

01
Data & Pre-processing

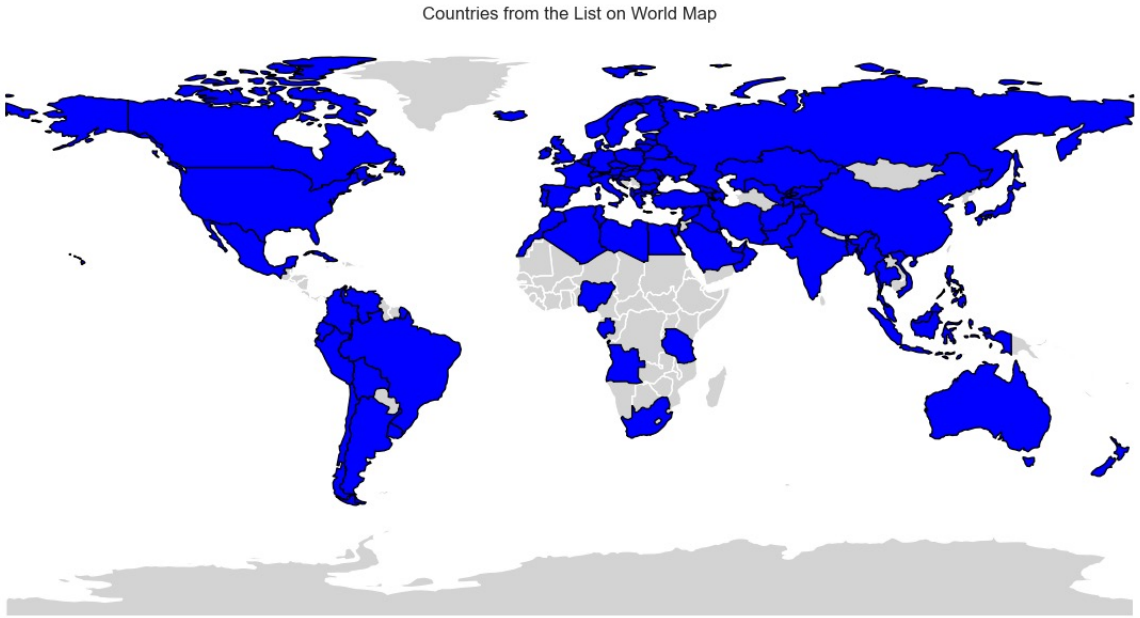
02
Problem Statements

03
Models

04
Results & Conclusion

Understanding dataset

The dataset includes **CO^2 emissions** of **92** different **countries** for a time-span of **27 years** from 1992 to 2018.



92 countries

From the CORGIS Dataset Project. By Sam Donald Version 1.0.0, created 9/23/2022

https://corgis-edu.github.io/corgis/csv/global_emissions/

Understanding dataset

- **19 features**, including greenhouse **gas production in kilotons**.
- Detailed **CO^2 production** data from **sources** like cement, oil, gas, and coal.
- **One Categorical variable**: country name, country code.

```
Data columns (total 20 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Year                                  2484 non-null   int64
 1   Country.Name                          2484 non-null   object
 2   Country.Code                          2484 non-null   object
 3   Country.GDP                           2484 non-null   float64
 4   Country.Population                    2484 non-null   int64
 5   Emissions.Production.CH4              2484 non-null   float64
 6   Emissions.Production.N2O              2484 non-null   float64
 7   Emissions.Production.CO2.Cement       2484 non-null   float64
 8   Emissions.Production.CO2.Coal         2484 non-null   float64
 9   Emissions.Production.CO2.Gas          2484 non-null   float64
 10  Emissions.Production.CO2.Oil          2484 non-null   float64
 11  Emissions.Production.CO2.Flaring      2484 non-null   float64
 12  Emissions.Production.CO2.Other        2484 non-null   float64
 13  Emissions.Production.CO2.Total        2484 non-null   float64
 14  Emissions.Global Share.CO2.Cement     2484 non-null   float64
 15  Emissions.Global Share.CO2.Coal       2484 non-null   float64
 16  Emissions.Global Share.CO2.Gas        2484 non-null   float64
 17  Emissions.Global Share.CO2.Oil        2484 non-null   float64
 18  Emissions.Global Share.CO2.Flaring    2484 non-null   float64
 19  Emissions.Global Share.CO2.Total      2484 non-null   float64
dtypes: float64(16), int64(2), object(2)
memory usage: 388.3+ KB
```

Features in the dataset (Independent Variables)

Data preprocessing

Year wise totals:

- emissions, productions, total GDP and total population.

Data Transformation:

- Applying transformation on columns

Adding New Features

- GDP per capita.
- for trend analysis in time-series data while ignoring the categorical variable:
    - o Continent category,
    - o Income category.

Year	Country.GDP	Country.Population	Emissions.Production.CH4	Emissions.Production.N2O
1992	43100902214144.0	4925506924	6040.16	2029.31
1993	44254710147584.0	4993557961	6055.96	2011.81
1994	45812080437760.0	5060333191	6108.38	2040.12
1995	47587028476928.0	5126526633	6177.81	2090.19
1996	49599196732416.0	5192262944	6188.24	2127.99
1997	51565298985984.0	5257369247	6394.89	2102.07
1998	52817072630784.0	5321912696	6264.77	2115.82
1999	54776791415808.0	5385917834	6218.49	2106.16
2000	57586361456640.0	5449432936	6226.21	2100.92
2001	59280038126592.0	5512511161	6187.16	2114.63
2002	61292084835328.0	5575279315	6307.19	2155.24
2003	63808774490112.0	5637937091	6324.97	2165.57
2004	67375228895232.0	5700721768	6557.05	2248.27

Modified data-frame

Exploratory Data Analysis (EDA)

Summary statistics

	Year	Country.Name	Country.Code	Country.GDP	Country.Population	Emiss
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
2479	False	False	False	False	False	False
2480	False	False	False	False	False	False
2481	False	False	False	False	False	False
2482	False	False	False	False	False	False

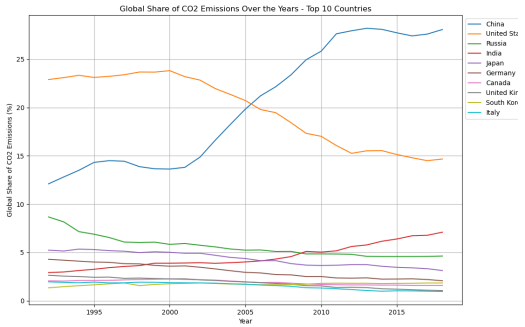
Top Countries based on Mean Production emission:

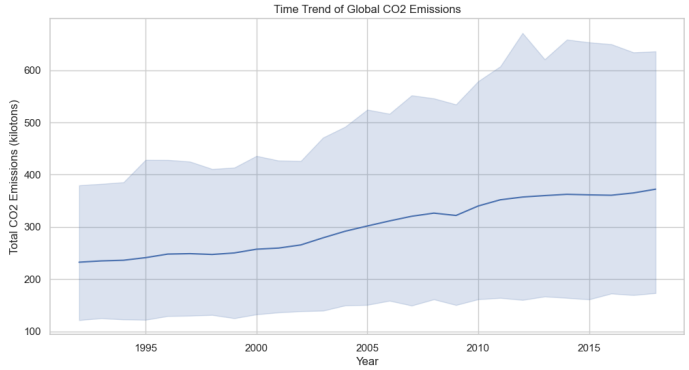
Country.Name	
China	6251.696407
United States	5648.154074
Russia	1604.018481
India	1396.887519
Japan	1240.925519
Germany	869.391630
Canada	549.727926
United Kingdom	524.115741
South Korea	507.217519
Iran	463.024889

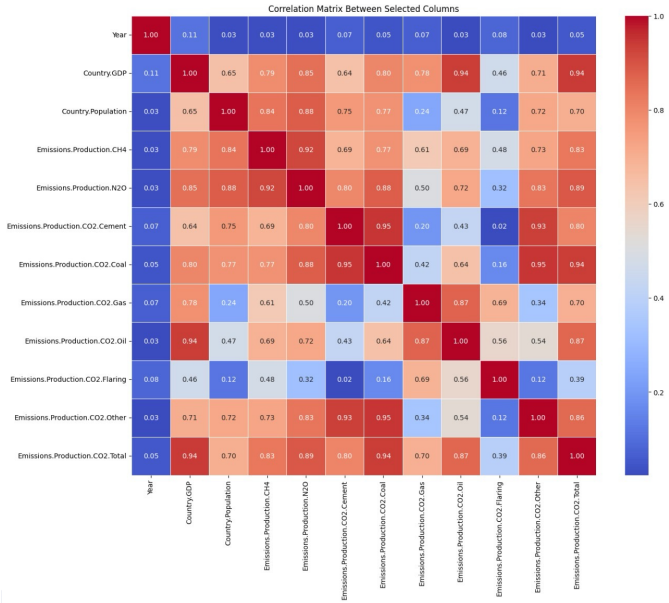
Statistics:

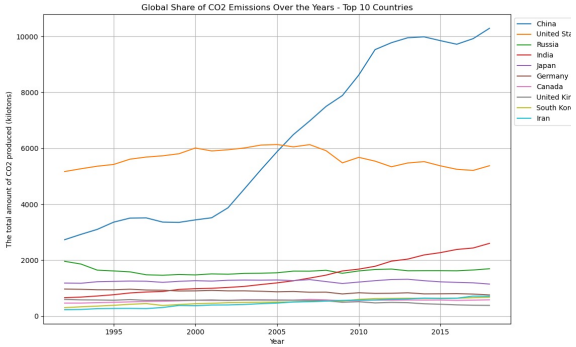
	...	Emissions.Global Share.CO2.Total
2484.000000	...	2484.000000
2005.000000	...	1.020688
7.790449	...	3.089116
1992.000000	...	0.000000
1998.000000	...	0.080000
2005.000000	...	0.220000
2012.000000	...	0.722500
2018.000000	...	28.210000

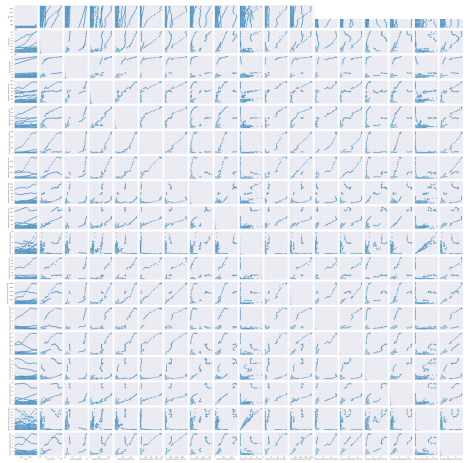
Exploratory Data Analysis (EDA)





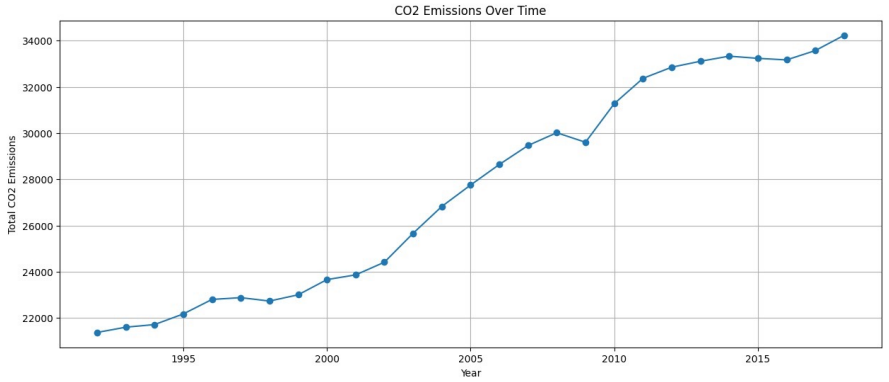


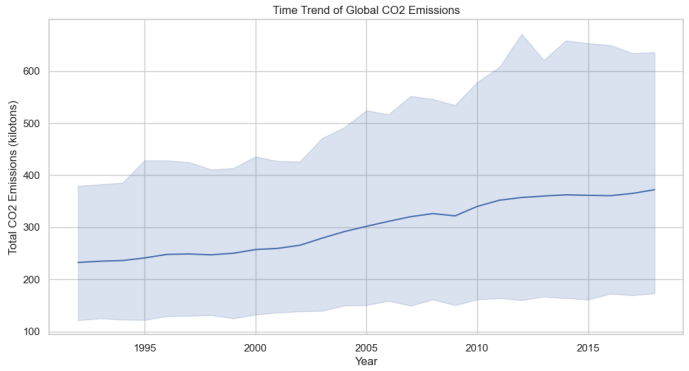


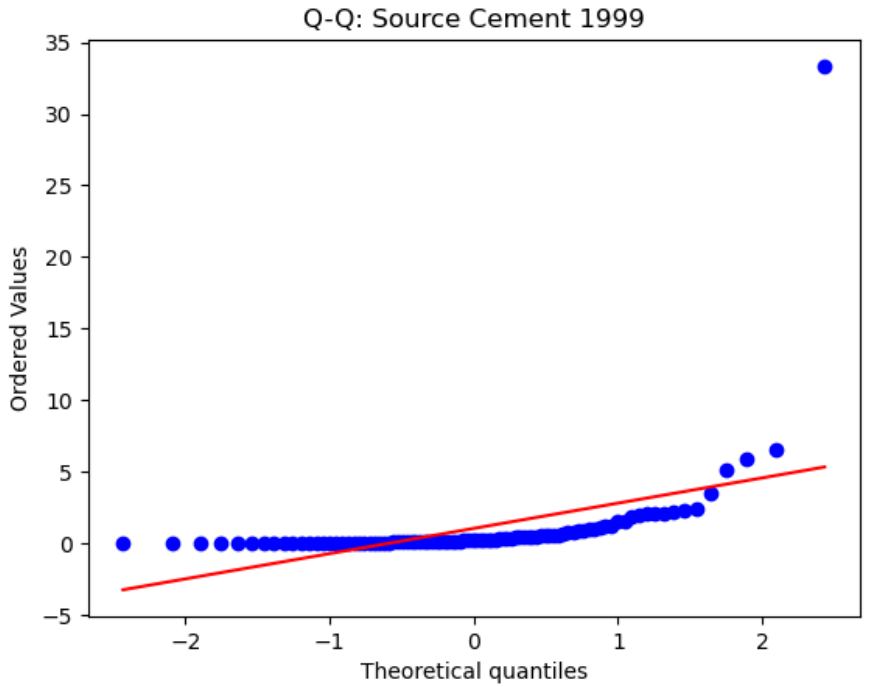


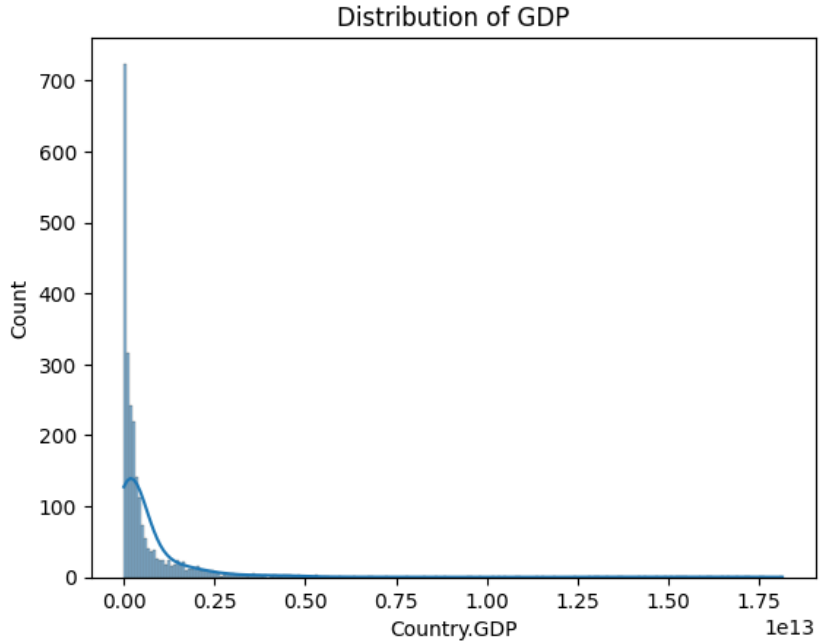
Exploratory Data Analysis (EDA)

- Time Trends
- Distributions
- Checks for Normality









Research Questions

- Analyzing global CO2 emissions with a focus on understanding interconnected factors.
- Investigating the relationship between a country's economic strength, industry composition, and environmental responsibility.
- Examining the equitable distribution of CO2 contributions from different sources.
- Assessing trends in CO2 pollution between high-income and low-income countries, and the effectiveness of global agreements in reducing emissions.

Hypothesis Testing

Analysis of Income Categorization on CO2 Emissions [1]

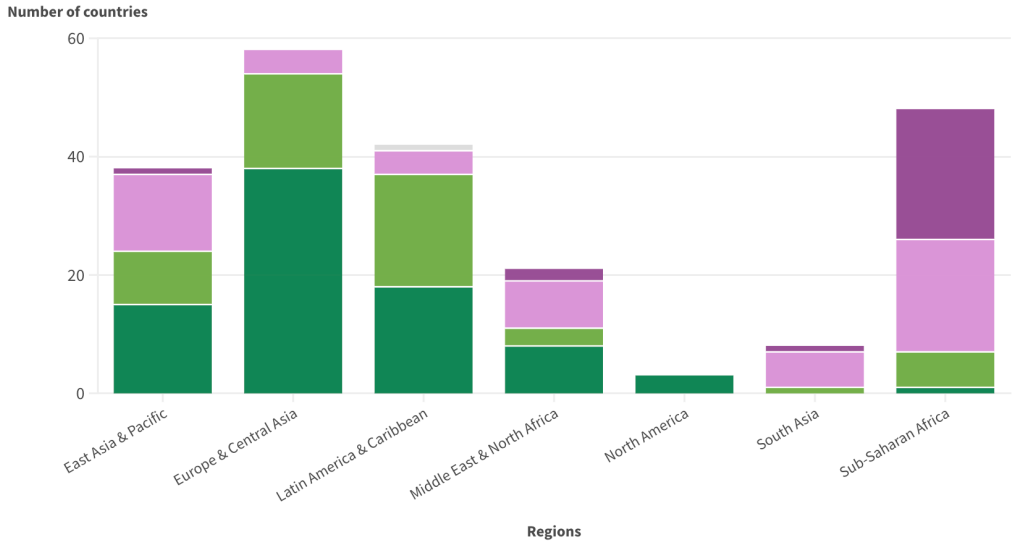
H0: There is no significant difference in CO2 emissions between income categories.

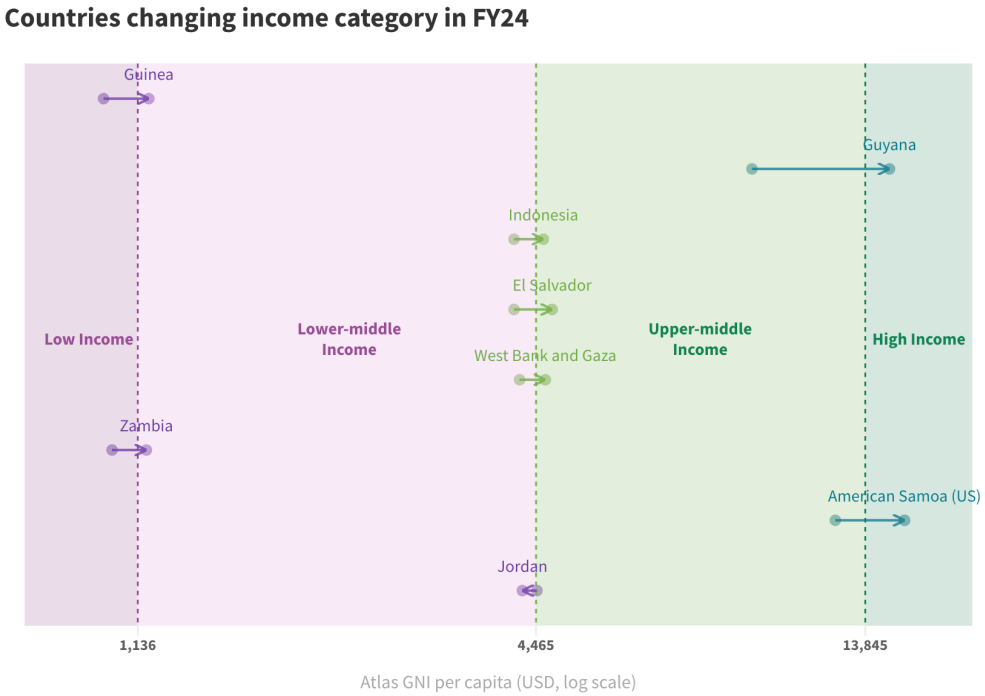
H1: There is a significant difference in CO2 emissions between income categories.

World Bank's Income Categorization

What?

	Low Income	Lower-middle Income	Upper-middle Income	High Income
July 1, 2023 - for FY24 (new)	<= 1,135	1,136 - 4,465	4,466 - 13,845	> 13,845
July 1, 2022 - for FY23 (previous)	<= 1,085	1,086 - 4,255	4,256 - 13,205	> 13,205





Changes every year, but we consider it constant.

https://blogs.worldbank.org/opendata/new-world-bank-group-country-classifications-income-level-fy24

Does income categorization have significant effect on CO2-emsn?

The distribution of total CO2 emission in a particular year?

Seems exponential

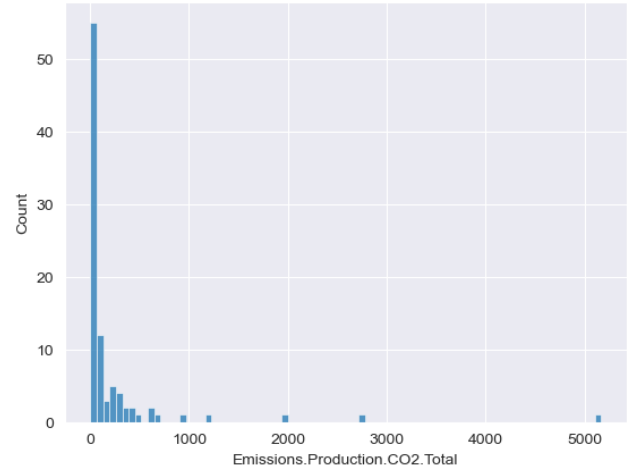
What kind of test can we do to say something about the hypothesis?

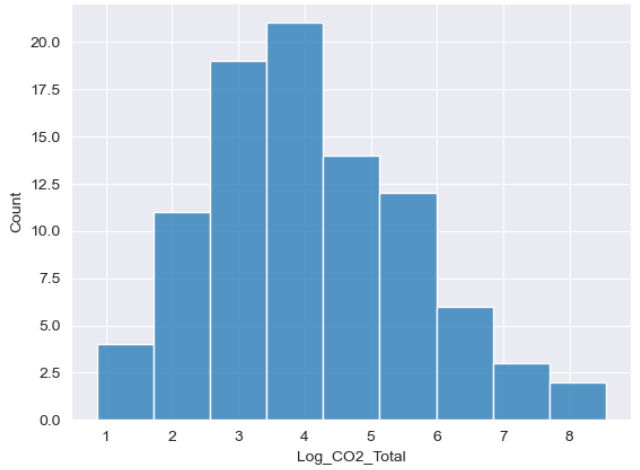
ANOVA

Transformation to make the distribution more symmetric?

Log transformation

The distribution of CO2 total in 1992





Does income categorization have significant effect on CO2-emsn? [1]

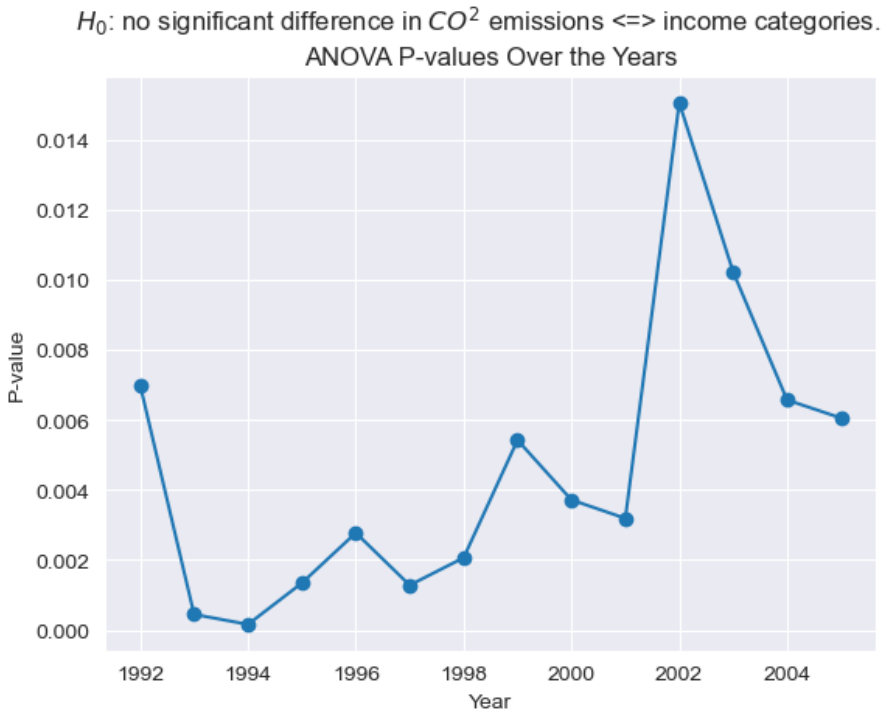
The distribution of CO2 total in 1992

H0: There is no significant difference in CO2 emissions between income categories.

H1: There is a significant difference in CO2 emissions between income categories.

The results of ANOVA:
We successfully reject H0 for all of the years.

So, there is enough evidence to suggest that for any year the income category has significant effect on CO2 emissions.



Impact of greenhouse gases on CO2 emissions: [2]

H0: Production of **N2O** has no significant impact on CO2 emissions.

H1: Production of **N2O** has a significant impact on CO2 emissions.

Impact of greenhouse gases on CO2 emissions:

- We have used Ordinary Least Squares(OLS)regression results.
- Based on the coefficient values and R-squared we were able to reject hypothesis.
- Since, the coefficient of the N2O is significantly large compared to 0. So, we can reject the null hypothesis and were able to establish a significant impact on CO2 emissions.

OLS Regression Results

Dep. Variable:	Emissions.Production.CO2.Total	R-squared:	0.792
Model:	OLS	Adj. R-squared:	0.792
Method:	Least Squares	F-statistic:	4733.
Date:	Sat, 09 Dec 2023	Prob (F-statistic):	0.00
Time:	02:33:47	Log-Likelihood:	-18577.
No. Observations:	2484	AIC:	3.716e+04
Df Residuals:	2481	BIC:	3.718e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-54.4483	9.541	-5.707	0.000	-73.157	-35.739
Emissions.Production.N2O	13.5233	0.383	35.336	0.000	12.773	14.274
Emissions.Production.CH4	0.3125	0.149	2.095	0.036	0.020	0.605

Omnibus:	559.875	Durbin-Watson:	0.083
Prob(Omnibus):	0.000	Jarque-Bera (JB):	19607.984
Skew:	0.297	Prob(JB):	0.00
Kurtosis:	16.751	Cond. No.	197.

Impact of greenhouse gases on CO2 emissions: [3]

H0: Production of **CH4** has no significant impact on CO2 emissions.

H1: Production of **CH4** has a significant impact on CO2 emissions.

Impact of greenhouse gases on CO2 emissions:

- We have used Ordinary Least Squares(OLS)regression results.
- Based on the coefficient values and R-squared we were failing to reject the null hypothesis.
- Since, the coefficient of the CH4 is significantly close compared to 0. So, we are failing to reject the null hypothesis and were able to establish a significant impact on CO2 emissions.

OLS Regression Results

Dep. Variable:	Emissions.Production.CO2.Total	R-squared:	0.792
Model:	OLS	Adj. R-squared:	0.792
Method:	Least Squares	F-statistic:	4733.
Date:	Sat, 09 Dec 2023	Prob (F-statistic):	0.00
Time:	02:33:47	Log-Likelihood:	-18577.
No. Observations:	2484	AIC:	3.716e+04
Df Residuals:	2481	BIC:	3.718e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-54.4483	9.541	-5.707	0.000	-73.157	-35.739
Emissions.Production.N2O	13.5233	0.383	35.336	0.000	12.773	14.274
Emissions.Production.CH4	0.3125	0.149	2.095	0.036	0.020	0.605

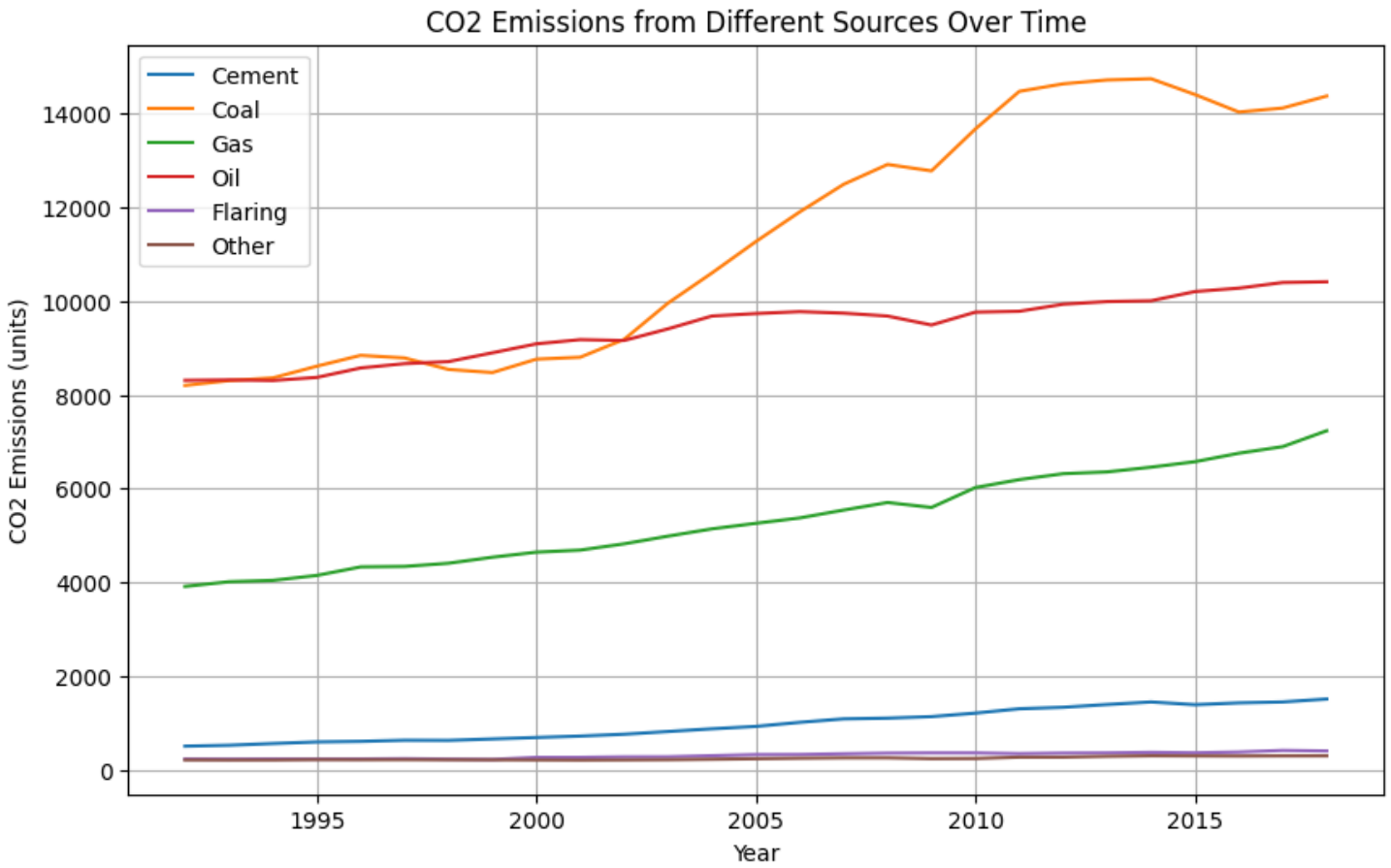
Omnibus:	559.875	Durbin-Watson:	0.083
Prob(Omnibus):	0.000	Jarque-Bera (JB):	19607.984
Skew:	0.297	Prob(JB):	0.00
Kurtosis:	16.751	Cond. No.	197.

Regional/Continental Disparities in CO2 Emissions [5]

H0: Different continents have the same mean total CO2 Emissions

H1: There are significant differences in the mean total CO2 emissions across different continents.

Statistical Models: CO2 prediction from sources



Statistical Models: CO2 prediction from sources

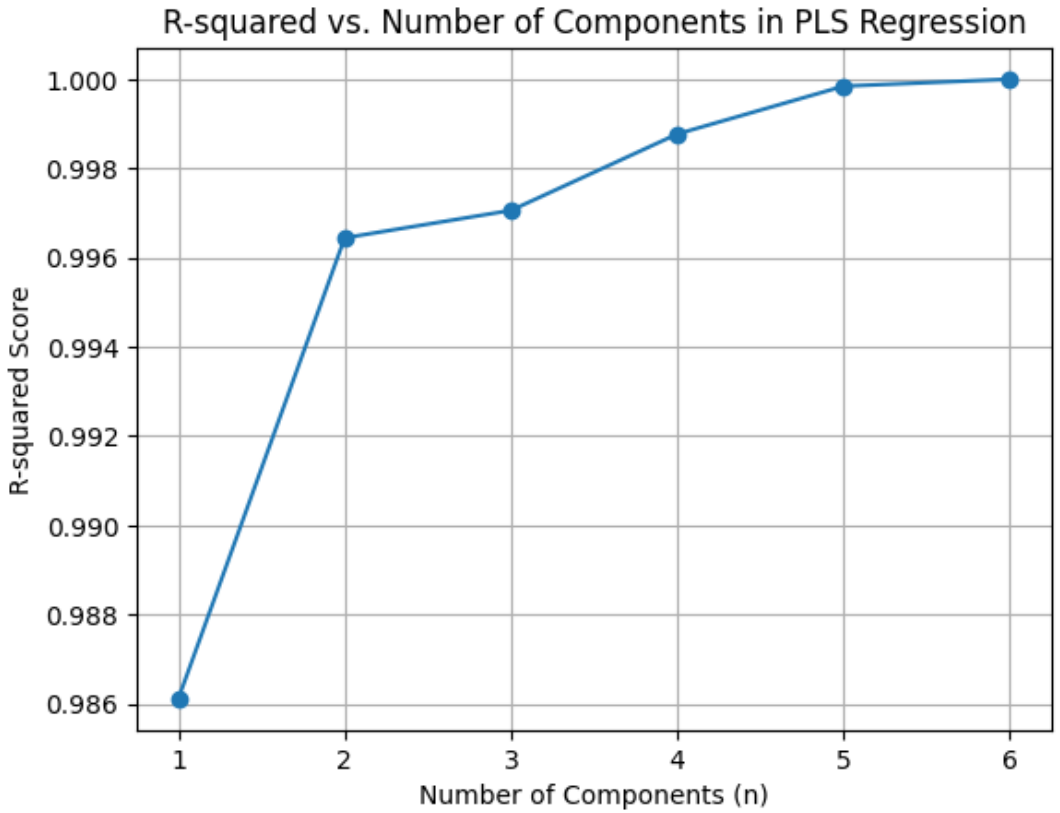
- We have used dimensionality reduction along with regression to calculate the total CO2 emissions of a country from the prominent sources only.

- This has been achieved using Partial Least Squares(PLS) method.

- PLS reduces the input dimensionality to map the output exactly with fewer and lesser input features.

Statistical Models: CO2 prediction from sources

- To apply PLS method, we standardized our data and split it into train & test data with 80:20 ratio.
- To find the optimal no. of components, we ran our code with multiple no. of components and obtained the R-squared score.
- In the R-squared plot, we got a knee at a value of 2.
- Hence, we reduced our no. of components from 6 to 2 model the total emission of CO2.



Statistical Models: CO2 prediction from sources

- The two major components which can be used to determine the total CO2 emission of a country are **coal** and **oil**.

- The equation of the model is:

  Y = B1*x1 + B2*x2

  Where x1 is the CO2 emissions from coal.

  & X2 is the CO2 emissions from oil.

- This model can be effectively used to approximate the total CO2 emissions of any country with just two CO2 sources.

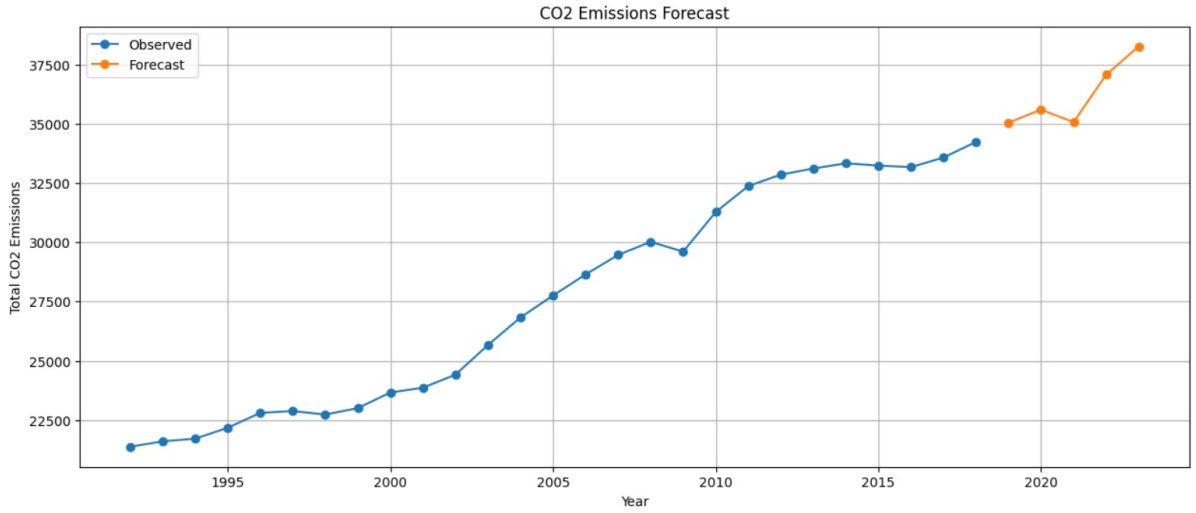
Statistical Models: S-ARIMA

SARIMA Model:

SARIMA stands for Seasonal Autoregressive Integrated Moving Average.

The model consists of three main components: Autoregressive (AR), Integrated (I), and Moving Average (MA).

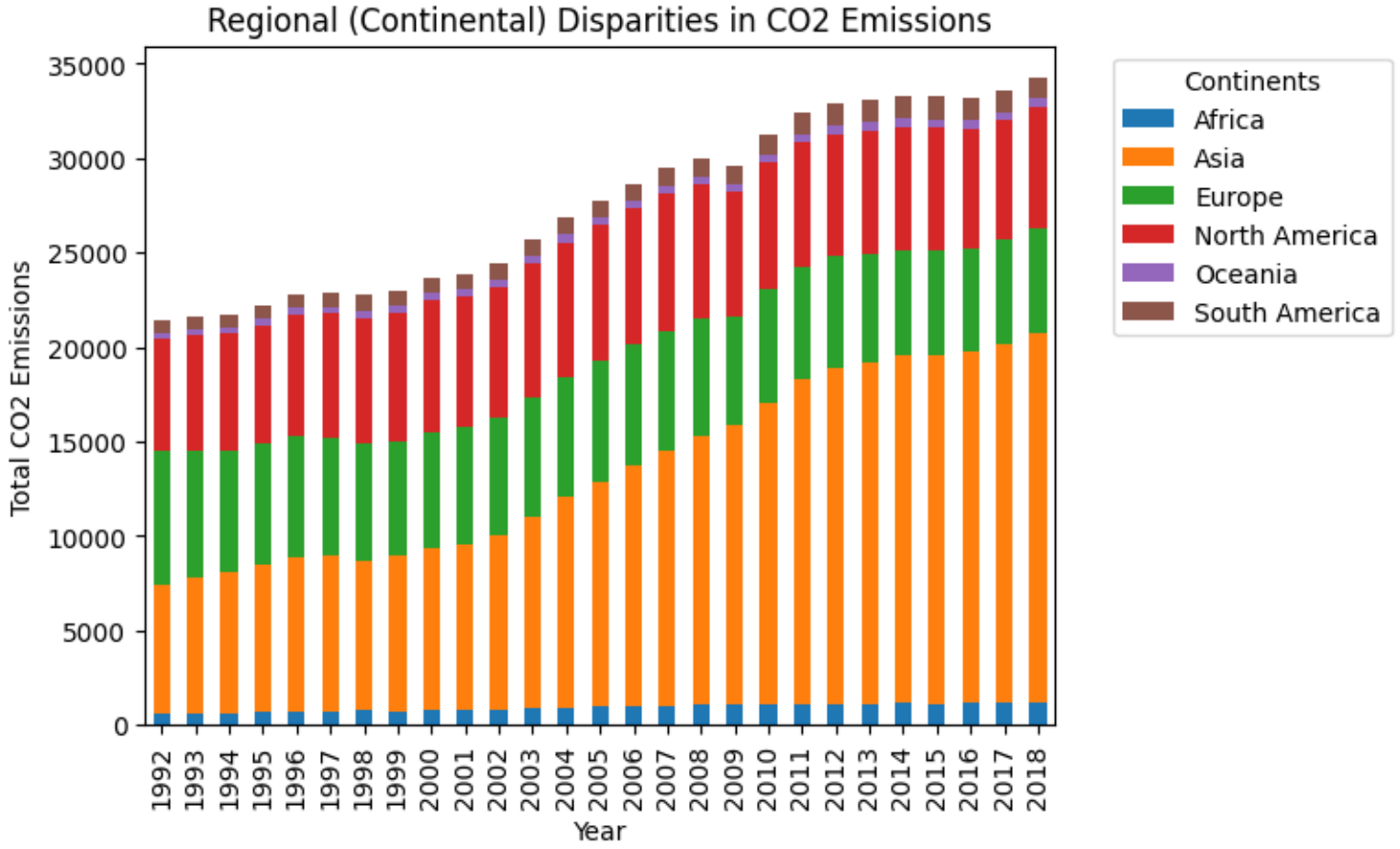
Seasonal components (S) are added to the model to capture periodic patterns.



Statistical Models: S-ARIMA

- The data is grouped by year, and the total CO2 emissions for each year are summed up, creating a time series data.
- Seasonal decomposition of time series is performed using an additive model to identify underlying components such as trend, seasonality, and residuals.
- Seasonal Autoregressive Integrated Moving Average (SARIMA) model is fitted to the time series data.
- The fitted SARIMA model is used to forecast future emissions for a specified number of steps

Statistical Models: ANOVA



Climate Action: Need of the Moment

- **The Kyoto Protocol:** The Kyoto Protocol is an international treaty adopted in 1997 in Kyoto, Japan, as an extension of the United Nations Framework Convention on Climate Change (UNFCCC). It aimed to address global warming and its impacts by committing participating nations to reduce greenhouse gas emissions.
- **5.2% Reduction Relative to 1990:** the agreement by developed countries to collectively reduce their greenhouse gas emissions by an average of 5.2% below their 1990 levels during the commitment period from 2008 to 2012.
- The protocol laid the foundation for subsequent climate agreements and discussions, shaping the ongoing global dialogue on climate action and the reduction of greenhouse gas emissions.

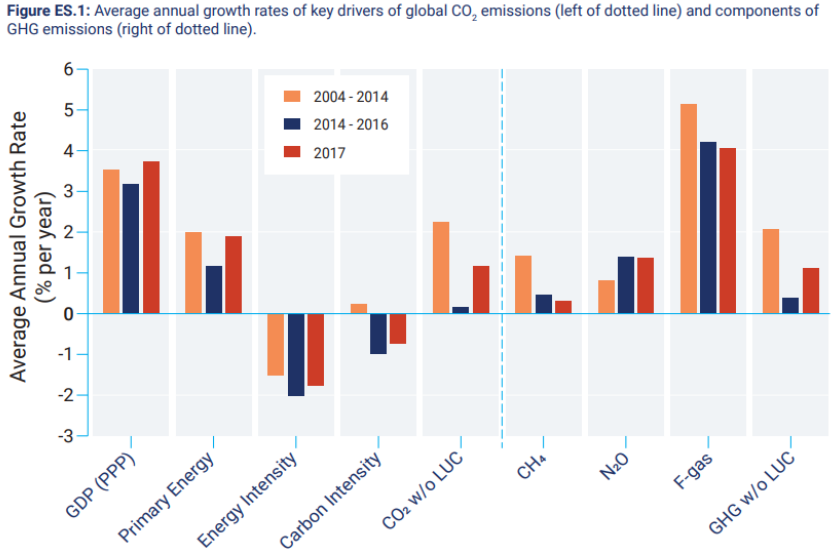
Climate Action: Need of the Moment

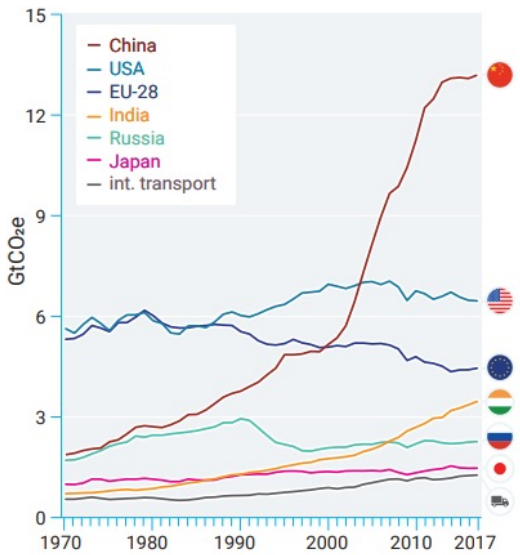
- **The Paris Agreement (2015):** The Paris Agreement is a landmark international treaty negotiated under the United Nations Framework Convention on Climate Change (UNFCCC).
- **Limiting Global Warming:** One of the central objectives of the Paris Agreement is to limit global temperature increase well below 2 degrees Celsius above pre-industrial levels, with efforts to limit the increase to 1.5 degrees Celsius.
- **Nationally Determined Contributions (NDCs):** The agreement emphasizes the concept of Nationally Determined Contributions, wherein each participating country pledges its own voluntary climate action plan.

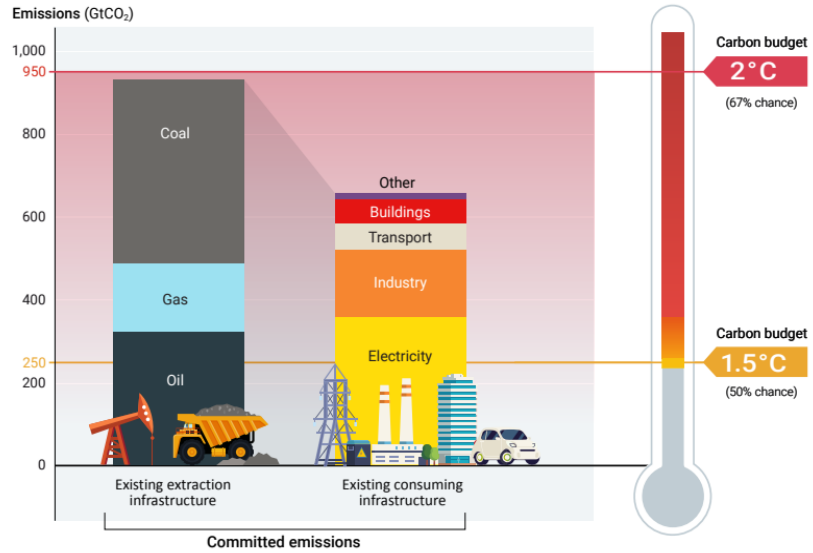
Policy Making for Nations

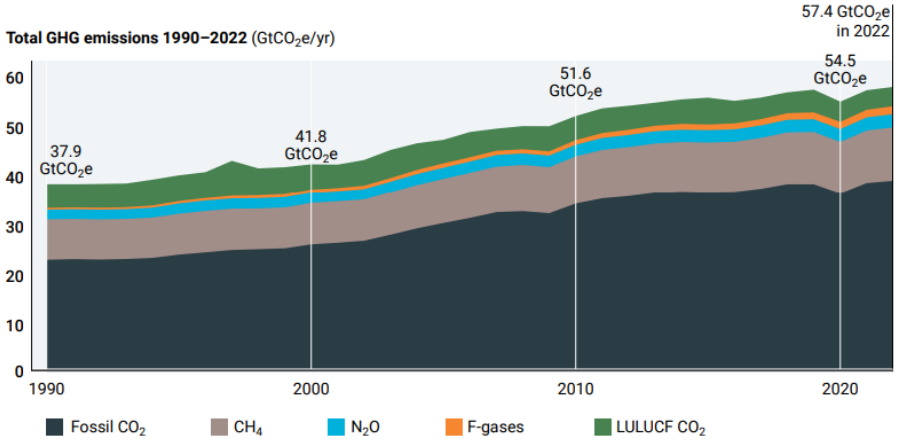
- **Climate Change Impacts:** Our project delves into the complex landscape of global CO2 emissions, identifying key trends and patterns.
- **Informing Policy Decisions:** The findings from our analysis are intended to contribute to informed policy-making. Understanding the trends in CO2 emissions allows policymakers to formulate effective strategies for mitigating climate change, promoting sustainable practices, and meeting international commitments like those outlined in the Kyoto Agreement and Paris Agreement.
- **Identifying Key Sources:** By pinpointing the major sources of CO2 emissions and their temporal patterns, our project aids in prioritizing efforts to reduce carbon footprints. This information is crucial for policymakers and environmental agencies to target specific industries or sectors for regulatory measures, technological innovation, and sustainable practices, thereby fostering a more impactful and targeted approach toward emission reduction.

Verification Of Findings









CO2 emissions from coal rise to all-time high

Coal accounted for over 40% of the overall growth in global CO_2 emissions in 2021. Coal emissions now stand at an all-time high of 15.3 Gt, surpassing their previous peak (seen in 2014) by almost 200 Mt. CO_2 emissions from natural gas also rebounded well above 2019 levels to 7.5 Gt, as demand increased in all sectors. At 10.7 Gt, emissions from oil remained significantly below pre-pandemic levels because of the limited recovery in global transport activity in 2021.

Thank You!