Analyzing the Impact of US Industrial Sectors on Climate Change and Economic Growth (2010-2020)

1. Introduction:

The relationship between industrial growth and climate change is a critical challenge facing the United States, because industries are both essential drivers of economic prosperity and major contributors to greenhouse gas emissions. This project analyzes the dual impact of major US industrial sectors on economic growth and climate change, using statistical analysis of industry-specific GDP data and environmental impact data. The results can give insights into which industries provide the best balance of economic benefits versus environmental costs, helping policymakers and business leaders make informed decisions about industrial development and environmental regulation.

2. Objective:

Analyze the contribution of different industrial sectors in the United States to economic growth and their corresponding environmental impacts, examining the relationships between sectoral GDP contributions and greenhouse gas emissions from 2010 to 2020.

3. About Data

A. Data Sources

For the analysis, we utilized two csv file: gdp_by_industry and ghg_emissions. The dataset is generated from an automated ETL pipeline. gdp_by_industry provides insights how different industries have its contribution in economic growth and ghg_emissions provide insights how different industries are contributing in greenhouse gas emissions. Both datasets are from United States of America from year 2010-2020.

The data pipeline ensures that the dataset is cleaned and transformed as required for the analysis. The final dataset (pipeline output) is complete with no significant missing value and no irrelevant features. Both datasets provide coverage of how major industries are contributing in economy and greenhouse gas emissions.

B. Data Structure

gdp_by_industry csv is structured with columns Year and industries as index.

Description	Line	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Agriculture, forestry, fishing, and hunting	1	145.7	179.9	179.5	215.8	200.6	182.1	167.5	176.8	177.1	164.2	162.9
Forestry, fishing, and related activities	3	28.7	28.8	30.7	31.4	33.5	35.9	37.2	38.1	40.3	41.6	43.9
Construction	9	525.7	525.6	554.9	594.7	649.9	715.3	776.8	840.2	889.1	953.0	957.8
Petroleum and coal products	18	123.8	159.1	159.1	146.6	147.8	131.4	76.0	110.7	154.2	143.4	60.5
Electrical equipment, appliances, and components	26	50.8	48.3	52.2	57.5	53.4	61.4	56.0	57.6	59.8	60.1	57.7
Transportation and warehousing	44	433.5	452.5	473.3	497.4	533.6	583.9	603.0	635.5	677.3	710.0	638.7
Real estate and rental and leasing	57	1997.2	2058.3	2119.7	2215.2	2274.2	2338.0	2419.9	2525.0	2683.4	2805.9	2910.6
Administrative and waste management services	65	438.0	455.0	474.9	490.1	518.3	540.3	559.2	602.9	638.7	670.2	648.9

gdp_by_industry.csv

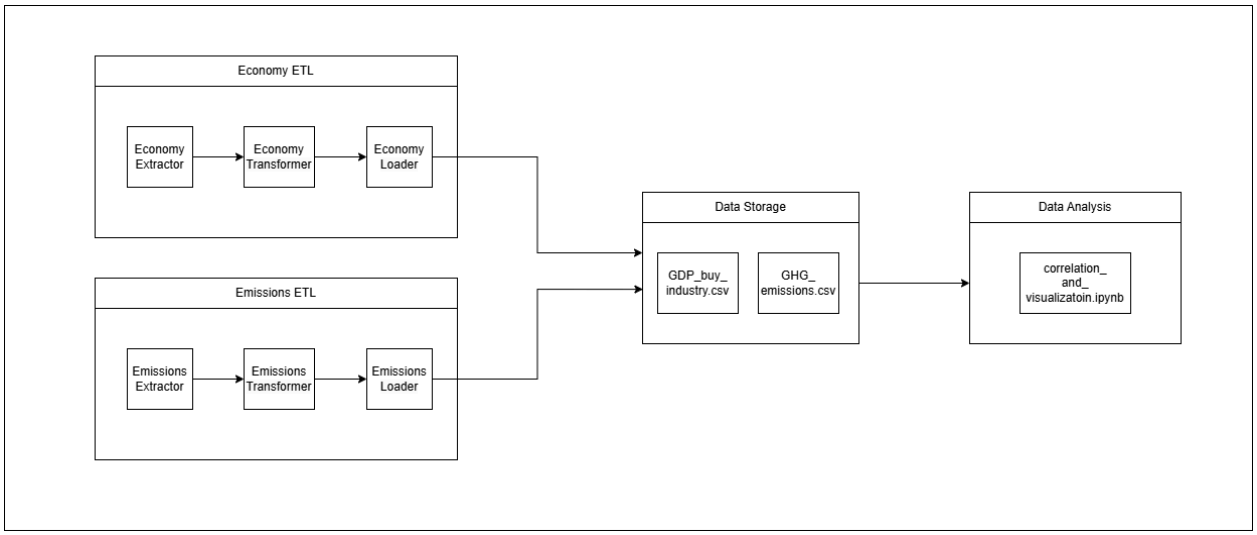
ghg_emmission csv is structured with Year and name of industries.

	Entity	Code	Year	Greenhouse gas emissions from agriculture	Greenhouse gas emissions from land use change and forestry	Greenhouse gas emissions from waste	Greenhouse gas emissions from buildings	Greenhouse gas emissions from manufacturing and construction	Greenhouse gas emissions from transport	Greenhouse gas emissions from electricity and heat	Greenhouse gas emissions from other fuel combustion
0	United States	USA	2010	370829980.0	-390820000.0	148270000.0	552170000.0	469900000.0	1.733120e+09	2.620930e+09	52110000.0
1	United States	USA	2011	369820000.0	-407280000.0	139230000.0	529050000.0	439100000.0	1.686210e+09	2.499530e+09	48700000.0
2	United States	USA	2012	365640000.0	-407560000.0	140310000.0	466350000.0	441920000.0	1.649510e+09	2.370170e+09	46710000.0
3	United States	USA	2013	363920000.0	-407490000.0	136430000.0	535270000.0	438240000.0	1.700900e+09	2.388770e+09	50770000.0
4	United States	USA	2014	363400000.0	-407580000.0	136100000.0	552640000.0	440620000.0	1.693990e+09	2.386490e+09	48780000.0
5	United States	USA	2015	364190000.0	-406300000.0	135200000.0	518710000.0	437190000.0	1.756190e+09	2.243240e+09	48740000.0

ghg_emmissions.csv

C. Data Pipeline

Below is the architecture of the end to end pipeline.



4. Analysis

To understand how different industries contribute to the GPD and Greenhouse gas emissions, we performed several data visualizations techniques (Exploratory Data Analysis) and correlation Analysis on the datasets. We collected some questions and tried to find answers using EDA. By examining these analyses on industries, we tried to find any relation between industries, like if one industry one major of economy does not really emit greenhouse gases. From the beginning it looks like we will find positive corelation but we did not get results as expected

A. Exploratory Data Analysis (EDA):

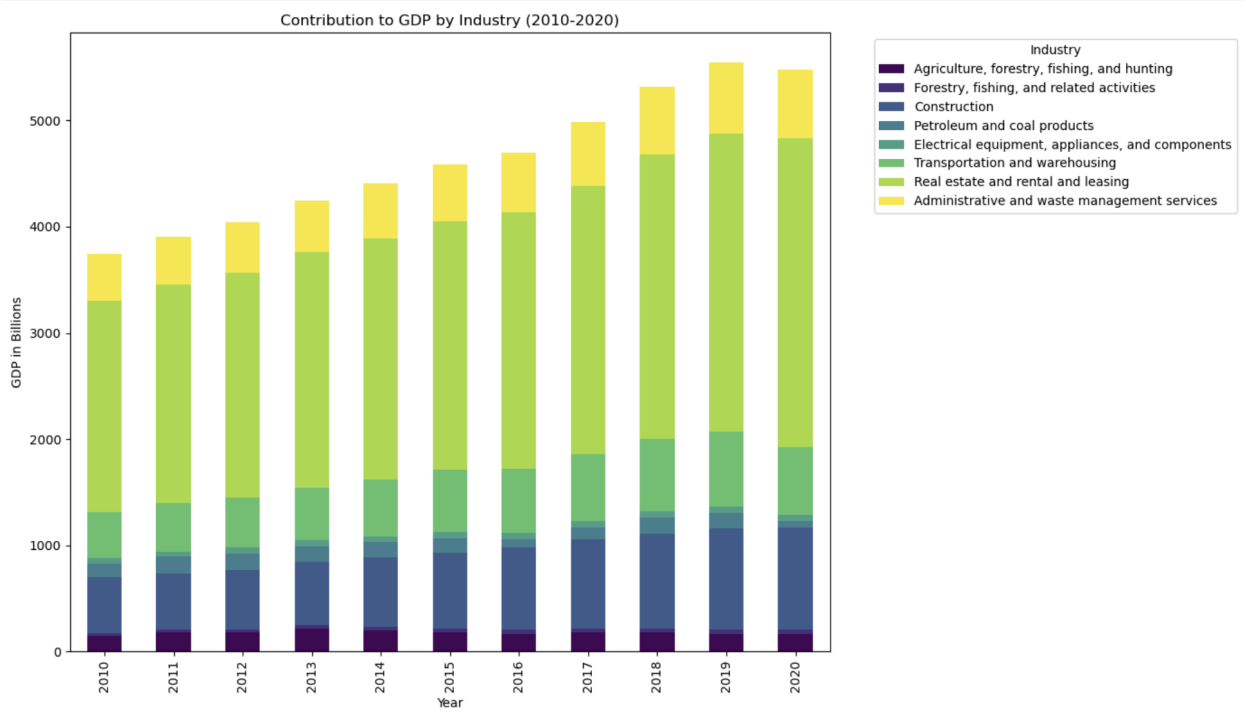
Exploratory Data Analysis (EDA) is crucial for uncovering patterns, detecting anomalies, and gaining insights from the dataset. It enables a better understanding of data, which helps to make

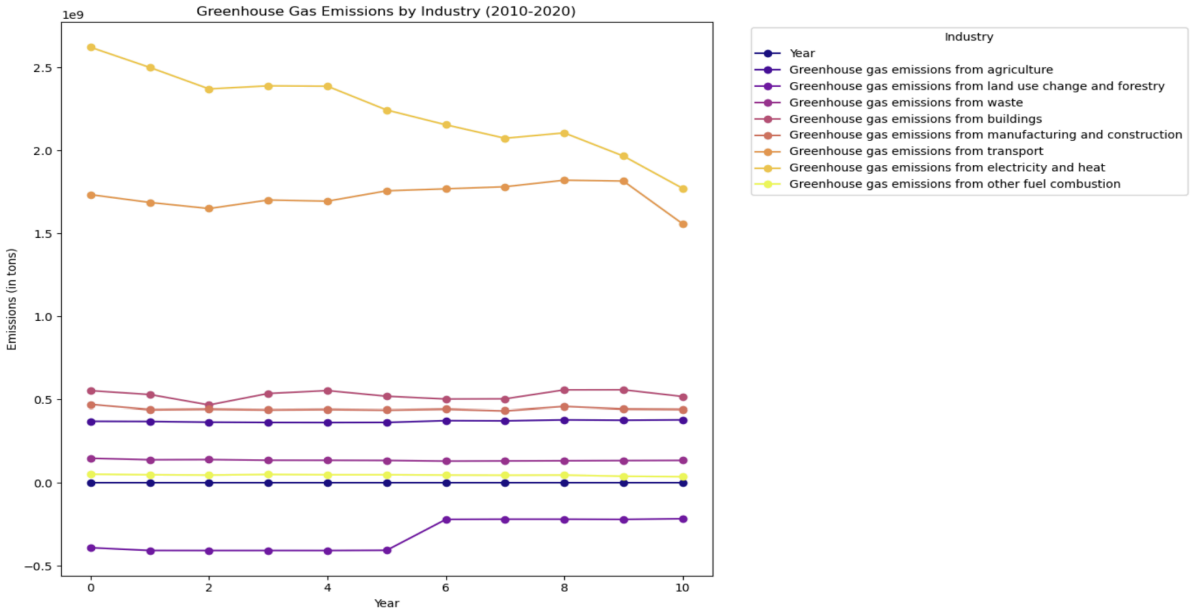
better and efficient decisions. By visualizing these relationships, we can better understand the impact of industries on GDP and gas emissions.

B. Visualization

Visualization simplifies the exploration of complex datasets by uncovering patterns, trends, and outliers that might be missed in raw data. It enhances understanding and communication of insights, aiding in informed decision-making.

Below is the grouped bar chart for both datasets for data profiling.





C. Questions:

1. Find the industry with the highest average contribution to GDP over the period (2010-2020)
2. Which emission category has contributed the most to emissions over the entire period?
3. What is the overall trend of each industry's impact over time?
4. Which industry has experienced the most significant growth in GDP over the time period?
5. Which industry has the most volatile impact on GDP?
6. Which emission category has the most volatile impact?
7. Calculate the growth of each industry from 2010 to 2020
8. Calculate the standard deviation (volatility) of each industry's contribution to GDP

5. Results

The analysis reveals the correlation between GDP and greenhouse gas emissions across various sectors. Key sectors such as Transportation, Electricity, and Construction exhibit strong positive correlations with GDP, indicating that economic growth in these industries is associated with increased emissions. Sectors like Forestry and Petroleum show weaker or negative correlations, suggesting unique dynamics in their contributions.

6. Interpretation

The positive correlation in sectors like Transportation and Electricity highlights their critical role in economic growth and emissions. Conversely, the negative correlation in Forestry suggests efforts in sustainable practices or carbon offsets. These patterns indicate that industrial activities significantly influence emissions, but the impact varies by sector.

7. Conclusion

The analysis answers the question: "How have key industrial sectors contributed to economic growth and greenhouse gas emissions in the US and Brazil over the past two decades?" It reveals that sectors like Transportation, Electricity, and Construction are significant contributors to both economic growth and emissions, suggesting their central role in driving industrial activity. Conversely, sectors like Forestry and Petroleum show different trends, with Forestry indicating potential progress in sustainable practices. These findings align with the hypothesis that industrial activities influence both GDP and emissions but to varying degrees across sectors.

However, the question could not be answered completely due to certain limitations. The analysis focuses only on two countries and does not account for cross-border industrial impacts, policy interventions, or technological advancements over the years. Additionally, the data may have inconsistencies or gaps that affect the accuracy of trends and correlations. Further research could integrate more countries, assess policy effectiveness, and incorporate technological factors to develop a holistic view of the relationship between industrial growth and greenhouse gas emissions.