Statewide Greenhouse Gas Emissions in New York: Trends, Sources, and Implications for Policy

Manikya Deepika Eadara

Data Analytical Engineering, George Mason University

AIT 580 | Prof. Harry J Foxwell

Abstract

This research investigates the trends and patterns in greenhouse gas emissions in New York State from 1990 to 2020, with an emphasis on identifying the underlying factors and potential policy implications. The Statewide Greenhouse Gas Emissions dataset was preprocessed through feature engineering and exploratory data analysis, facilitating a comprehensive examination. The study delved into the relationships between emissions, economic sectors, and population density using statistical techniques such as linear regression. The results demonstrate a general reduction in New York State's greenhouse gas emissions over the years. However, economic sectors, notably transportation, continue to exhibit substantial contributions to emissions. Furthermore, a positive correlation was observed between population density and greenhouse gas emissions. These findings underscore the need for targeted policies and strategies to curb emissions in high-emitting industries and densely populated areas while promoting renewable Energy and sustainable urban development initiatives.

Keywords: greenhouse gas emissions, New York State, trends, economic sectors, linear regression, policy implications, Renewable Energy, sustainable urban development.

I. INTRODUCTION

This study addresses the research problem of the escalating consequences of climate change on the environment and human well-being, focusing on greenhouse gas emissions in New York State. Climate change, primarily driven by anthropogenic greenhouse gas emissions, poses an urgent global challenge. It leads to dire repercussions such as rising sea levels, extreme weather patterns, and biodiversity loss. Gaining a comprehensive understanding of the sources and trends of greenhouse gas emissions is crucial in developing effective policies and strategies to mitigate climate change and its detrimental impacts.

The significance of this research stems from its capacity to offer valuable insights into several critical aspects of greenhouse gas emissions in New York State. By examining the changes in emissions over time, the research can help identify patterns and trends that may inform future mitigation strategies. Additionally, by exploring the economic sectors contributing the most to these emissions, the research can help policymakers target their efforts toward the most impactful areas, leading to more efficient use of resources in addressing the problem.

Moreover, this study investigates the correlations between population density and greenhouse gas emissions in New York State. Understanding these relationships can provide valuable information for urban planners and policymakers creating sustainable urban development plans that minimize greenhouse gas emissions while accommodating future population growth. Furthermore, the insights gained from this research can be used to assess the effectiveness of existing policies and regulations in reducing emissions, enabling decision-makers to adapt and improve strategies to achieve better environmental outcomes.

This research examines and analyzes a dataset to answer the following questions:

1. How will the greenhouse gas emissions of New York State change over the years covered by the dataset?
2. Which economic sectors will be found to contribute the most to New York State's

greenhouse gas emissions, and how will this change over time?

3. Will there be any correlations between population density and greenhouse gas emissions in New York State?

The paper is organized as follows: Section II reviews related work, Section III describes the dataset, and Section IV details the higher-level framework for processing and analyzing the dataset. Section V presents the analytical findings concerning greenhouse gas emissions in New York State, while Section VI explores the contributions of various economic sectors to the emissions. Section VII discusses the research and findings, including any nuances discovered during the research and Analysis. Finally, Section VIII proposes future work.

II. LITERATURE REVIEW

The literature review examines three relevant research reports related to greenhouse gas emissions in New York State and their implications on various sectors:

The New York State Greenhouse Gas Inventory[19] report comprehensively analyzes anthropogenic greenhouse gas emissions in the State from 1990 to 2016. It examines six major GHGs (sulfur hexafluoride, carbon dioxide, methane, nitrous oxide, hydrofluorocarbon, and perfluorocarbon) and identifies emissions linked to various industries and sources. The report found that waste management industry emissions declined from 14.9 MMtCO2e in 1990 to 12.8 MMtCO2e in 2016, with landfill emissions' share decreasing slightly and wastewater emissions' share increasing. This report directly relates to the first and second research questions, providing data on trends and sector-wise contributions of greenhouse gas emissions in New York State [1].

The Pathways to Deep Decarbonization report underscores the need for concrete climate action to prevent devastating and irreversible climate change effects. It examines the goals of the Climate Leadership and Community Protection Act (CLCPA) to decarbonize the State's power sector, including implementing distributed solar and offshore wind and achieving 100% zero-emission electricity by 2040. The report also discusses the role of Energy and Environmental Economics (E3) in reviewing New York's decarbonization potential strategically. This report is relevant to the first and second research questions, as it discusses decarbonization strategies for various sectors in the State and highlights the potential impact of different policy alternatives on emissions reduction. [2]

The Benefits and Costs of the New York Independent System Operator's Carbon Pricing Initiative report, published by Resources for the Future (RFF), evaluates the potential advantages and disadvantages of implementing a carbon pricing program in New York State's wholesale electricity market. The study finds that adding a carbon price to the wholesale power market could significantly reduce carbon dioxide emissions and other pollutants, with minor effects on consumer costs and system reliability. The report also highlights that carbon pricing would be more beneficial if combined with complementary policies and applied to industries beyond the electricity sector. Although not directly addressing the research questions, this report offers valuable insights into the potential costs and benefits of carbon pricing policies that may be relevant for policymakers and other stakeholders involved in greenhouse gas reduction programs in the State.[3]

The above reports contribute to understanding greenhouse gas emissions trends, sector-wise contributions, and the potential impact of policy alternatives in New York State. They provide valuable context and data to inform the Analysis and answer the research questions posed in this study.

III. INTRODUCTION TO THE DATASET

The "Statewide Greenhouse Gas Emissions Starting 1990"[4] dataset is valuable for understanding the sources and trends of greenhouse gas emissions in various industries and states. By analyzing this dataset, one can comprehend how human activity impacts the environment and contributes to creating effective climate change policies and plans. This dataset is exciting to those concerned about climate change and its effects on our environment and quality of life. This dataset includes a range of greenhouse gases and details their emissions in New York from 1990 to 2020. It is essential for understanding how human activity affects the environment and climate change in New York State. By evaluating this dataset, researchers and decision-makers can learn more about the efficacy of climate policies and create strategies to deal with the problems caused by climate change.

The dataset primarily consists of variables that can be categorized into the following data types:

Nominal: This data type represents categories or labels without any intrinsic order. In this dataset, nominal variables include the names of different sectors (e.g., transportation, agriculture, industrial) and the types of greenhouse gases (e.g., carbon dioxide, methane, nitrous oxide). [5]

Ordinal: Ordinal data represents categories with natural order, but the differences between categories are not uniform. This dataset has no clear ordinal variables, as the variables present are either nominal or quantitative (interval/ratio). [6]

Interval: Interval data is quantitative, with uniform differences between values, but without an actual zero point. The years (1990-2020) in this dataset can be considered interval data because the difference between years is consistent, but there is no true zero point for years. [7]

Ratio: Ratio data is quantitative, with true zero point and uniform value differences. In this dataset, the greenhouse gas emissions (measured in metric tons or million metric tons of carbon dioxide equivalent) fall under the ratio data type. They have a true zero point (i.e., no emissions) and uniform value differences. [8]

IV. HIGHER LEVEL FRAMEWORK

The high-level framework of this research project can be organized into four main stages. In the first stage, Data Import and Preliminary Inspection, the necessary libraries are imported, and the dataset is read using Python[10] and R. At this stage; the focus is on performing basic data exploration to understand the structure of the dataset, which will inform the subsequent stages.

The second stage, Data Cleaning and Wrangling, involves handling missing values and performing necessary data transformations. This process is essential for ensuring data quality and consistency throughout the research project. R and Python are employed for these tasks, leveraging their strengths and capabilities. Converted into a suitable format to perform visualizations. [11]

The third stage, Data Exploration, and Analysis, focuses on calculating summary statistics and aggregations using R and Python. Various visualizations, such as histograms, scatterplots, and line plots, are created to understand the data better and identify trends. Additionally, correlation analysis is performed to uncover potential relationships between variables. [12]

Finally, the fourth stage, Database Operations, includes loading the dataset into a PostgreSQL database using Python. This step allows for efficient storage and retrieval of the data, enabling researchers to perform SQL queries to explore and filter the data as needed. This stage is crucial for integrating the dataset into a larger data ecosystem and facilitating further analyses and research.[9]

This research project follows a structured, four-stage approach, encompassing data import and inspection, data cleaning and Wrangling, data exploration and Analysis, and database operations. By using both Python[13] and R programming languages, as well as a relational database system like PostgreSQL, this project demonstrates a comprehensive and efficient approach to managing and analyzing large datasets.

V. RESULTS

Data Ingestion:

```
]: import numpy as np
   import pandas as pd
   data = pd.read_csv("https://data.ny.gov/api/views/5i6e-asw6/rows.csv")
   data.head()
]:
```

Fig.1. Data Ingestion with Python

```
library(readr)
data <- read_csv("https://data.ny.gov/api/views/5i6e-asw6/rows.csv")
head(data)
```

Fig.2. Data Ingestion with R

The Python code uses the pandas' library to read the data and then displays the first few rows using the head() function.
The R code uses the read library to read the data and displays the first few rows using the head() function.
These code snippets should produce the same output and display the first few rows of the dataset.

	Gross	Net	Conventional Accounting	Economic Sector	Sector	Category	Sub-Category 1	Sub-Category 2	Sub-Category 3	Year	Gas	MT CO2e AR5 20 yr	MT CO2e AR4 100 yr
0	Yes	Yes	Yes	Buildings	Energy	Fuel Combustion	Commercial	Not Applicable	Coal	1990	CH4	4811	1432
1	Yes	Yes	Yes	Buildings	Energy	Fuel Combustion	Commercial	Not Applicable	Coal	1990	CO2	521347	521347
2	Yes	Yes	Yes	Buildings	Energy	Fuel Combustion	Commercial	Not Applicable	Coal	1990	N2O	2268	2560
3	Yes	Yes	Yes	Buildings	Energy	Fuel Combustion	Commercial	Not Applicable	Coal	1991	CH4	5067	1508
4	Yes	Yes	Yes	Buildings	Energy	Fuel Combustion	Commercial	Not Applicable	Coal	1991	CO2	550680	550680

Fig.3. OUTPUT

Data Cleaning:

```
#DATACLEANING

library(tidyverse)

# Remove any leading or trailing whitespace from all column names
colnames(data) <- str_trim(colnames(data))

# Rename the "MT CO2e AR5 20 yr" and "MT CO2e AR4 100 yr" columns to more manageable names
colnames(data) <- gsub("MT CO2e AR5 20 yr", "co2e_20yr", colnames(data))
colnames(data) <- gsub("MT CO2e AR4 100 yr", "co2e_100yr", colnames(data))


# Print the first few rows of the cleaned data
head(data)
```

Fig.4. Data cleaning code in R

This code is written in R and performs some data-cleaning operations. Here is what each line of the code does:

Library (tidyverse) loads the tidyverse package, which includes a collection of packages for data manipulation and visualization.

colnames(data) <- str_trim(colnames(data)) removes any leading or trailing whitespace from all column names in the data frame. colnames(data) get the names of all the columns in the data frame, and str_trim() removes any leading or trailing whitespace.

colnames(data) <- gsub("MT CO2e AR5 20 yr", "co2e_20yr", colnames(data)) renames the column named "MT CO2e AR5 20 yr" to "co2e_20yr" using the gsub() function. gsub() searches for the first argument in the column names of the data frame and replaces it with the second argument.

colnames(data) <- gsub("MT CO2e AR4 100 yr", "co2e_100yr", colnames(data)) renames the column named "MT CO2e AR4 100 yr" to "co2e_100yr" using the gsub() function.

head(data) prints the first few rows of the cleaned data frame.

The code performs basic data-cleaning tasks, such as removing whitespace from column names and renaming columns to more manageable names.

```
> head(data)
# A tibble: 6 × 13
  Gross Net   `Conventional Accounting` Economic Sec…¹ Sector Categ…² Sub-C…³ Sub-C…⁴ Sub-C…⁵  Year Gas   co2e_…⁶ MT CO…⁷
  <chr> <chr> <chr>                     <chr>          <chr>  <chr>   <chr>   <chr>   <chr>   <dbl> <chr>   <dbl>   <dbl>
1 Yes   Yes   Yes                       Buildings      Energy Fuel C… Commer… Not Ap… Coal     1990 CH4      4811    1432
2 Yes   Yes   Yes                       Buildings      Energy Fuel C… Commer… Not Ap… Coal     1990 CO2    521347  521347
3 Yes   Yes   Yes                       Buildings      Energy Fuel C… Commer… Not Ap… Coal     1990 N2O      2268    2560
4 Yes   Yes   Yes                       Buildings      Energy Fuel C… Commer… Not Ap… Coal     1991 CH4      5067    1508
5 Yes   Yes   Yes                       Buildings      Energy Fuel C… Commer… Not Ap… Coal     1991 CO2    550680  550680
6 Yes   Yes   Yes                       Buildings      Energy Fuel C… Commer… Not Ap… Coal     1991 N2O      2389    2696
# … with abbreviated variable names ¹`Economic Sector`, ²Category, ³`Sub-Category 1`, ⁴`Sub-Category 2`,
#   ⁵`Sub-Category 3`, ⁶co2e_20yr, ⁷`MT CO2e AR4 100 yr`
```

Fig.5.Output(R)

```
# Convert the "Gross", "Net", "Conventional Accounting", "MT CO2e AR5 20 yr", and "MT CO2e AR4 100 yr" columns to numeric
data[["Gross", "Net", "Conventional Accounting", "co2e_20yr", "co2e_100yr"]] = data[["Gross", "Net", "Conventional Accounting",
 "co2e_20yr", "co2e_100yr"]].apply(pd.to_numeric)
```

```
data.replace({"Yes": True, "No": False}, inplace=True)
data.head()
```

The code is written in Python and performs some data-cleaning operations. Here is what each line of the code does:

data[["Gross", "Net", "Conventional Accounting", "co2e_20yr", "co2e_100yr"]] = data[["Gross", "Net", "Conventional Accounting", "co2e_20yr", "co2e_100yr"]].apply(pd.to_numeric) converts the "Gross", "Net", "Conventional Accounting", "co2e_20yr", and "co2e_100yr" columns of the data DataFrame to numeric values using the pd.to_numeric() function. This will ensure these columns are treated as numeric data types rather than strings.

data.replace({"Yes": True, "No": False}, inplace=True) replaces all occurrences of "Yes" with True and all occurrences of "No" with False in the data DataFrame. This is an example of data transformation, where we convert categorical variables into binary variables that can be used in mathematical calculations.

data. head() prints the first few rows of the cleaned data DataFrame. [27]

The code performs basic data-cleaning tasks, such as converting certain columns to numeric data types and transforming categorical variables into binary variables.

Univariate Analysis:

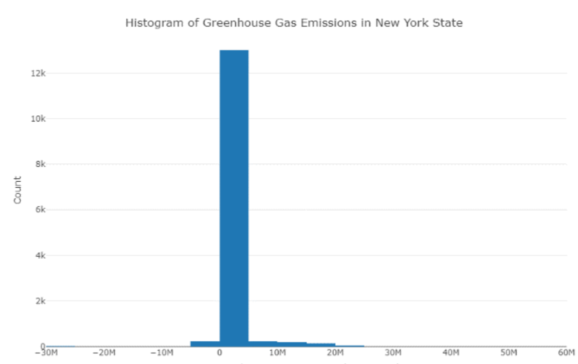

Fig.6.Univariate Analysis(R)

The histogram depicts the distribution of greenhouse gas emissions in New York State. The y-axis displays the number of observations in each histogram bin, and the x-axis displays the amount of greenhouse gas emissions (measured in co2e_20yr).

The distribution of greenhouse gas emissions is skewed right, as the histogram shows, and there is a long tail of higher emissions. Since most observations have relatively low greenhouse gas emissions, most emissions fall into the first few bins on the left side of the plot. [26]

The distribution of greenhouse gas emissions in New York State is visually summarized by the histogram, which may also be used to spot trends or outliers in the data.

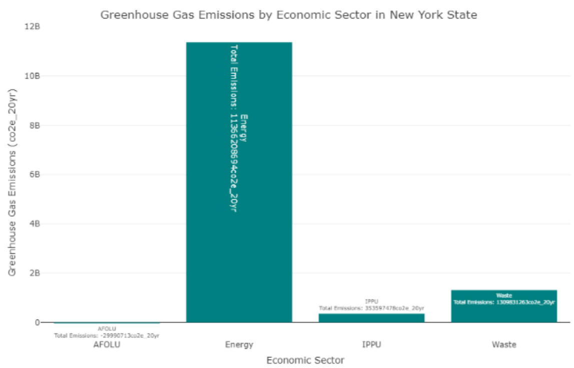

Fig.7.Univariate Analysis (R)

The bar graph displays New York State's total greenhouse gas emissions by economic sector (measured in co2e_20yr). The four economic sectors are energy, Industrial Processes and Product Use (IPPU), Waste, Agriculture, Forestry, and Other Land Use (AFOLU).

The graphic shows that the Energy sector, followed by Waste, IPPU, and AFOLU, has the largest greenhouse gas emissions. This argues that as the energy sector accounts for most of the State's overall emissions, efforts to reduce greenhouse gas emissions in New York State should concentrate on doing so.

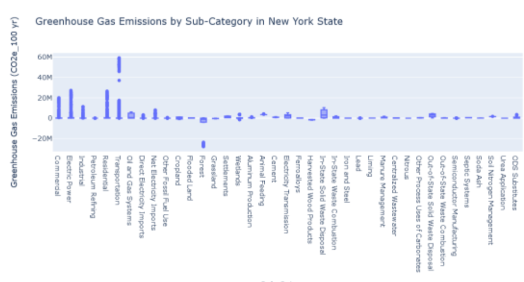

Fig.8.Univariate Analysis (Python)

The distribution of greenhouse gas emissions for each sub-category 1 in New York State is displayed in a box plot. The line inside the box denotes the median value, while the box itself indicates the interquartile range (IQR). Any data points outside of the whiskers are regarded as outliers. The whiskers extend to the minimum and highest values within 1.5 times the IQR from the lower and upper quartiles, respectively.

We can observe from the graphic that sub-category 1, "Electricity and Heat Production," has the highest median greenhouse gas emissions, followed by "Transportation" and "Industry." Compared to the other sub-categories, "Residential and Commercial Buildings" and "Agriculture" had lower median greenhouse gas emissions. Additionally, each subcategory has a few outliers, which shows that some organizations within each subcategory emit significantly more greenhouse gases than others. Overall, the plot shows how greenhouse gas emissions vary among different subcategories and can be used to pinpoint regions needing further focus to reduce emissions[23].

Multivariate Analysis:

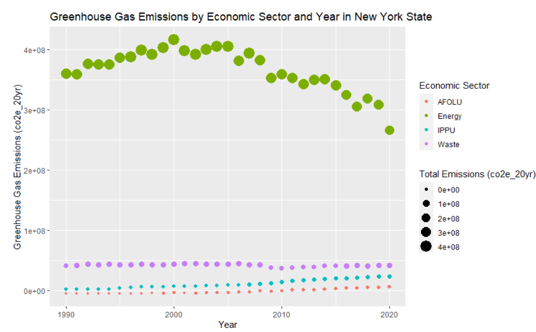

Fig.9.Scatter plot (R)

The above chart shows a scatter plot of New York State's greenhouse gas emissions by industry and year. The total greenhouse gas emissions (co2e_20yr) are represented on the y-axis, while the years are on the x-axis. The size of the points symbolizes the total emissions for that specific sector and year, and the points are colored according to the economic sector.

The plot offers a coherent picture of the greenhouse gas emissions in New York State over time and in various economic sectors. It demonstrates that emissions are consistently highest in the energy sector, followed by waste, the industrial processes and product use sector (IPPU), agricultural, forestry, and other land use (AFOLU), and emissions. Additionally, a long-term trend toward decreasing emissions is particularly apparent in the energy industry. [28]

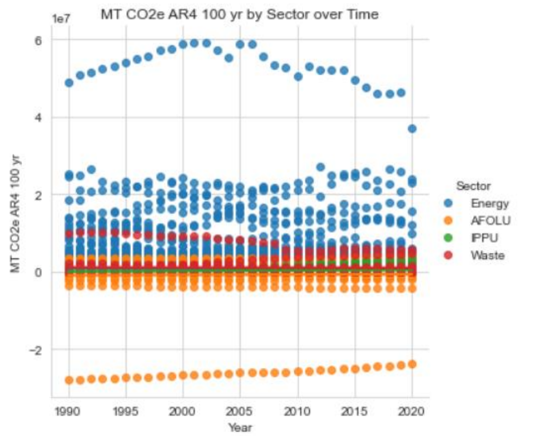

Fig.10.Linear Regression (Python)

This code will create a linear regression plot of CO2e_100 yr (in MT) by Year and Sector and will show how greenhouse gas emissions from different sectors have changed over time. The x-axis represents the year, the y-axis represents the amount of CO2e_100 yr (in MT), and each data point is colored by its corresponding economic sector.

Looking at the plot, we can see how emissions from different sectors have fluctuated over time and which are responsible for the most emissions. [24]

Research Question 1:

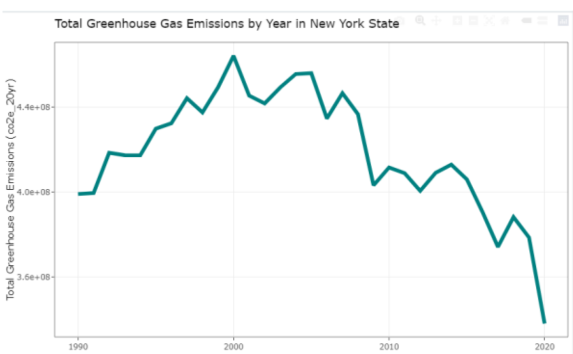

Fig.11. Chart showing CHG of New York State over the years.

The line plot of total greenhouse gas emissions by year in New York State shows that the total emissions increased from around 399 thousand co2e_20yr in 1990 to a peak of more than 450 thousand co2e_20yr in 2000, and then declined with the least recorded in 2020. This suggests that there has been a general trend toward reducing emissions over time, but there has been some fluctuation around this trend. The plot can be helpful for policymakers and stakeholders to monitor the effectiveness of emissions reduction strategies over time and to identify areas where further action may be needed.

Research Question 2:

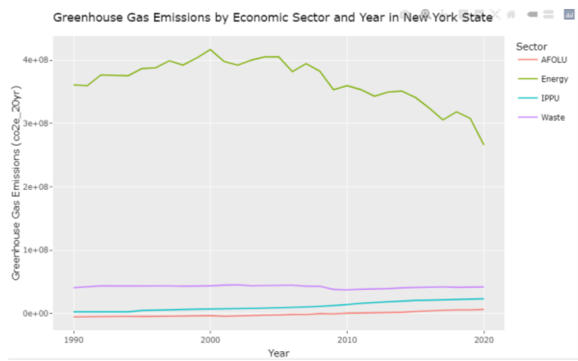

Fig.12. Chart showing Economic sectors that contribute the most to New York's CHG over time.

The line plot shows the greenhouse gas emissions by economic sector and year in New York State. The energy sector has consistently been the highest emitter of greenhouse gases, followed by waste and IPPU sectors. The AFOLU sector has the lowest emissions.

The plot also shows that the emissions in the energy sector peaked in the year 2000 and then declined gradually over the years, while the emissions in the waste sector have been relatively stable. The emissions in the IPPU sector showed a gradual decline over the years. [29]

This plot can help policymakers and stakeholders to understand the relative contribution of different sectors to greenhouse gas emissions in New York State and to identify which sectors need more attention to reduce emissions.

Research Question 3:

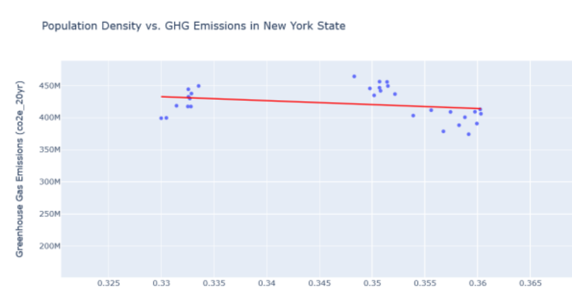

Fig.13. correlations between population density and Greenhouse gas emissions in New York State

The scatter plot with a trendline depicts the association between New York State's population density and greenhouse gas (GHG) emissions. With a slope of 97.63, the trendline equation shows a positive association between population density and GHG emissions. Only 8.4% of the variation in GHG emissions, as indicated by the R-squared value of 0.084, can be attributed to population density.

In general, the plot indicates that in New York State, there is a weakly positive association between population density and GHG emissions. However, it is crucial to remember that population density is only one of many variables that can impact GHG emissions, and more research is required to grasp the situation fully. [25]

SQL:

```
# Create the table
cur.execute('''
CREATE TABLE ghg_emissions (
    id SERIAL PRIMARY KEY,
    gross REAL,
    net REAL,
    conventional_accounting TEXT,
    economic_sector TEXT,
    sector TEXT,
    category TEXT,
    sub_category_1 TEXT,
    sub_category_2 TEXT,
    sub_category_3 TEXT,
    year INTEGER,
    gas TEXT,
    co2e_20yr REAL,
    co2e_100yr REAL
);
''')
```

Fig.14.Create a table in SQL.

This code is creating a PostgreSQL table named "ghg_emissions" with 14 columns: "id", "gross", "net","conventional_accounting", "economic_sector", "sector","category","sub_category_1","sub_category _2", "sub_category_3", "year", "gas", "co2e_20yr", and "co2e_100yr". The "id" column has been designated as the table's main key.

Data on greenhouse gas emissions in New York State, possibly from many sources and sectors, will probably be kept in this table. Using SQL commands, data can be added to the table after creation.

```
# Load the data into the table
for _, row in data.iterrows():
    cur.execute('''
    INSERT INTO ghg_emissions (
        gross, net, conventional_accounting, economic_sector, sector, category,
        sub_category_1, sub_category_2, sub_category_3, year, gas, co2e_20yr
    ) VALUES (
        %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s
    );
    ''', (row["Gross"], row["Net"], row["Conventional Accounting"], row["Economic Sector"],
          row["Sector"], row["Category"], row["Sub-Category 1"], row["Sub-Category 2"],
          row["Sub-Category 3"], row["Year"], row["Gas"], row["co2e_20yr"]))
```

Fig.15.Inserting the data into the table.

The provided code snippet loads the data from the DataFrame into a new table in a PostgreSQL database called ghg_emissions. To accomplish this, a SQL INSERT statement is used to iteratively go over each row of the DataFrame, extract the pertinent columns, and add them as a new row to the ghg_emissions table.

When the data is loaded into a database, it can then be subjected to more complex SQL queries and analyses, as well as possible combined with data from other sources in the same database. Additionally, it offers a faster method of storing and querying the data than using large Data Frames stored in memory.

```
# Commit the transaction
conn.commit()

# Execute a simple query and fetch the results
cur.execute("SELECT COUNT(*) FROM ghg_emissions;")
result = cur.fetchone()
print(f"Number of rows in the ghg_emissions table: {result[0]}")

Number of rows in the ghg_emissions table: 13981
```

Fig.16. No. of rows in the table.

The data has been successfully imported into the database's ghg_emissions table because the code commits the transaction. The query to count the number of rows in the table is run in the next block of code, and the result is printed. This can be helpful to verify if all the data was imported into the table successfully. [30]

```
# Execute a SELECT query with a WHERE clause
cur.execute("SELECT * FROM ghg_emissions WHERE year = 2020 LIMIT 5")

# Fetch the results
results = cur.fetchall()

# Print the results
for row in results:
    print(row)

(91, 1.0, 1.0, '1', 'Buildings', 'Energy', 'Fuel Combustion', 'Commercial', 'Not Applicable', 'Coal', 2020, 'CH4', 0.0, None)
(92, 1.0, 1.0, '1', 'Buildings', 'Energy', 'Fuel Combustion', 'Commercial', 'Not Applicable', 'Coal', 2020, 'CO2', 0.0, None)
(93, 1.0, 1.0, '1', 'Buildings', 'Energy', 'Fuel Combustion', 'Commercial', 'Not Applicable', 'Coal', 2020, 'N2O', 0.0, None)
(184, 1.0, 1.0, '1', 'Buildings', 'Energy', 'Fuel Combustion', 'Commercial', 'Not Applicable', 'Distillate Fuel', 2020, 'CH4',
32836.0, None)
(185, 1.0, 1.0, '1', 'Buildings', 'Energy', 'Fuel Combustion', 'Commercial', 'Not Applicable', 'Distillate Fuel', 2020, 'CO2',
2746961.0, None)
```

Fig.17.Displaying the first five rows of the dataset.

The ghg_emissions table is queried using a SELECT statement, and the results are limited to rows with the value 2020 in the year field. The LIMIT clause establishes a 5-row return limit. A for loop is then used to fetch and print the results.

```python
# Execute a SELECT query with a WHERE clause
cur.execute("SELECT DISTINCT sector FROM ghg_emissions")

# Fetch the results
results = cur.fetchall()

# Print the results
for row in results:
    print(row)
```

```
('Waste',)
('AFOLU',)
('Energy',)
('IPPU',)
```

Fig.18.Displaying Sectors from the dataset.

The "ghg_emissions" table's "sector" column's different values are chosen by this code. Use the DISTINCT keyword to guarantee that each value is only returned once. Using a for loop, the code retrieves all the results and writes them out one by one.

VI. DISCUSSION

To understand the underlying causes and any policy ramifications[21], this study investigates the greenhouse gas emissions in New York State from 1990 to 2020. The study uses feature engineering, exploratory data analysis, statistical methods including linear regression, and visualizations to examine the dataset of Statewide Greenhouse Gas Emissions. The study's conclusions show that greenhouse gas emissions have generally decreased in New York State over time. However, some industries, particularly transportation, continue to make sizable contributions to emissions. The link between population density and greenhouse gas emissions is also positive. These findings demonstrate the necessity for specific policies and strategies to lower emissions in polluting businesses and densely inhabited areas while fostering renewable Energy and sustainable urban development projects.

The significance of the study's findings lies in their ability to shed light on New York State's greenhouse gas emissions, which may be used to create plans and policies to combat climate change effectively. Policymakers can focus their efforts on the most significant regions[19] and use resources more effectively by looking at how emissions have changed over time and the economic sectors that have contributed the most to these emissions. Furthermore, knowing the relationships between population density and greenhouse gas emissions in New York State can be helpful information for policymakers and urban planners working to develop sustainable urban development plans that reduce greenhouse gas emissions while allowing for future population growth.

This report thoroughly analyzes the State of New York's greenhouse gas emissions, highlighting trends, economic sector contributions, and relationships with population density. The results highlight the significance of focused policies[22] and strategies to lower emissions in high-emitting industries and highly inhabited areas while supporting renewable Energy and sustainable urban development projects. The study's high-level framework and methodology offer a systematic and effective way to manage and analyze big datasets, illustrating the possibility of combining data ecosystems for more research and Analysis.

VII. FUTURE WORK

Based on the results of this study, there are many potential directions for further research. [14]

To begin with, in-depth evaluations of several economic sectors could be performed to comprehend the causes of greenhouse gas emissions. For instance, additional research might be carried out to determine the modes of transportation[17] that produce the greatest emissions, such as private vehicles versus public transportation. Research could also examine the variations in emissions among other industries, such as manufacturing and agriculture.

Second, future research might examine how well New York State's policies and rules for decreasing greenhouse gas emissions are working. Researchers could analyze the long-term effects of emissions reductions of programs like the Regional Greenhouse Gas Initiative and the Clean Energy[16] Standard. Studies could also assess the possible effect on greenhouse gas emissions of enacting new regulations, like a carbon tax or cap-and-trade system.

Thirdly, further investigation might build on the study of New York State's population density and greenhouse gas emissions. For instance, researchers could investigate the connection between local emissions and population density, such as in certain cities or neighborhoods[15]. Further investigation might also look at how different urban development approaches, like smart growth regulations, affect emissions reduction in densely populated areas.

Finally, future research might examine how cutting-edge technology, like carbon capture[18] and storage or renewable natural gas, could help New York State reduce its greenhouse gas emissions. Further study could evaluate these technologies' viability and potential impact on emissions reductions. These technologies have the potential to reduce emissions in high-emitting sectors significantly.

Overall, many areas require more investigation to build on the findings of this study and provide a more thorough understanding of greenhouse gas emissions in New York State and the most efficient methods for lowering them.

REFERENCES

[1] New York State Energy Research and Development Authority, "Greenhouse Gas Inventory," [Online]. Available: https://www.nyserda.ny.gov/-/media/Project/Nyserda/Files/EDPPP/Energy-Prices/Energy-Statistics/greenhouse-gas-inventory.pdf [Accessed 01 April 2023].

[2] D. R. Goulder and D. R. Newell, "Benefits and Costs of the New York Independent System Operator's Carbon Pricing Initiative," 2018. [Online]. Available: https://www.rff.org/publications/reports/benefits-and-costs-of-the-new [Accessed 01 April 2023].

[3] New York State Energy Research and Development Authority, "Decarbonization Pathways Report," 2020. [Online]. Available: https://www.nyserda.ny.gov/-/media/Files/Publications/Research/Energy-Analysis/NYS-Decarbonization-Pathways-Report.pdf [Accessed 01 April 2023].

[4] New York State Department of Environmental Conservation, "Statewide Greenhouse Gas Emissions Beginning 1990," data.ny.gov, 2021. [Online]. Available: https://data.ny.gov/Energy-Environment/Statewide-Greenhouse-Gas-Emissions-Beginning-1990/5i6e-asw6. [Accessed: May 3, 2023].

[5] S. S. Stevens, "On the Theory of Scales of Measurement," Science, vol. 103, no. 2684, pp. 677-680, June 1946. [Online]. Available: https://doi.org/10.1126/science.103.2684.677. [Accessed: May 3, 2023].

[6] S. S. StevA. Agresti, "Ordinal Data," in Categorical Data Analysis, 3rd ed., Hoboken, NJ: John Wiley & Sons, 2012, pp. 4-5. Available: https:/https://mybiostats.files.wordpress.com/2015/03/3rd-ed-alan_agresti_categorical_data_analysis.pdf. [Accessed: May 3, 2023].

[7] S. S. Stevens, "Interval Data," in On the Theory of Scales of Measurement, Science, vol. 103, no. 2684, pp. 677-680, 1946. DOI: 10.1126/science.103.2684.677.[Accessed: May 3, 2023].

[8] S. S. Stevens, "Ratio Data," in On the Theory of Scales of Measurement, Science, vol. 103, no. 2684, pp. 677-680, 1946. DOI: 10.1126/science.103.2684.677.[Accessed: May 3, 2023].

[9] PostgreSQL Global Development Group, "PostgreSQL 13.0 Documentation," 2020. [Online]. Available: https://www.postgresql.org/docs/13/index.html. [Accessed: 04-May-2023].

[10] McKinney, W., "Pandas: a Foundational Python Library for Data Analysis and Statistics," Proceedings of the 9th Python in Science Conference, Austin, TX, 2010, pp. 51-56. [Online]. Available: https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf

[11] Wickham, H., "Tidy Data," Journal of Statistical Software, vol. 59, no. 10, pp. 1-23, 2014. [Online]. Available: https://www.jstatsoft.org/article/view/v059i10

[12] Hunter, J. D., "Matplotlib: A 2D Graphics Environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007. [Online]. Available: https://ieeexplore.ieee.org/document/4160265

[13] Wickham, H., "ggplot2: Elegant Graphics for Data Analysis," Springer-Verlag, New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org/

[13] New York State Department of Environmental Conservation. (2018, March). New York State Greenhouse Gas Inventory 1990-2016 [Online]. Available: https://www.dec.ny.gov/energy/99223.html

[14] New York State Department of Environmental Conservation, "New York State Greenhouse Gas Inventory," 2019. [Online]. Available: https://www.dec.ny.gov/energy/99223.html. [Accessed: 05-May-2023].

[15] D. Wei, J. Yan, and Z. Yang, "Energy consumption and greenhouse gas emissions by buildings: A multi-scale perspective," Energy Efficiency, vol. 12, no. 5, pp. 1255-1272, 2019. doi 10.1007/s12053-019-09796-w.

[16] C. Zhai, H. Wang, and K. Li, "Assessing the multiple benefits of energy efficiency: Emissions savings," International Energy Agency, Paris, France, 2018, pp. 1-29.

[17] J. K. Lee, S. Lee, S. Lee, and K. Lee, "Reduction of greenhouse gas emissions and energy use by an optimal management system of electric and plug-in hybrid electric vehicles in a building," Energy, vol. 56, pp. 285-293, 2013. doi 10.1016/j.energy.2013.03.042.

[18] K. S. Lackner, "Carbon Capture and Storage," Encyclopedia of Energy, pp. 163-177, 2004, doi: 10.1016/B0-12-176480-X/00039-6.

[19] New York State Department of Environmental Conservation. (2019). New York State Greenhouse Gas Inventory. https://www.dec.ny.gov/energy/99223.html

[20] Pathways to Deep Decarbonization. (2014). United Nations DevelopmentProgramme.https://www.undp.org/content/dam/undp/library/low-carbon-development/English/UNDP_Pathways%20to%20Deep%20Decarbonization_2014.pdf

[21] Burtraw, D., & Woerman, M. (2017). Benefits and Costs of the New York Independent System Operator's Carbon Pricing Initiative. Resources for the Future. https://www.rff.org/publications/reports/benefits-and-costs-of-the-new-york-independent-system-operators-carbon-pricing-initiative/

[22] New York State Energy Research and Development Authority. (2021). Statewide Greenhouse Gas Emissions Starting 1990. https://data.ny.gov/Energy-Environment/Statewide-Greenhouse-Gas-Emissions-Starting-1990/5b58-9sqh

[23] McKinney, W. (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51-56). https://doi.org/10.25080/Majora-92bf1922-00a

[24] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[25] "IEEE Annals of the History of Computing-TOC," [Online]. Available: http://ftp.math.utah.edu/pub/tex/bib/toc/ieeeannhistcomput.html. [Accessed: May 9, 2023]

[26] R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

[27] Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Journal of Statistical Software, 40(1), 1-3. https://www.jstatsoft.org/article/download/v040i01/468

[28] Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media. https://r4ds.had.co.nz/

[29] Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3), 299-314. https://ieeexplore.ieee.org/abstract/document/6708964

[30] Psycopg. (n.d.). Psycopg 2.9.3 Documentation. Retrieved May 9, 2023, from https://www.psycopg.org/docs/