Impact of Global Greenhouse Gas Emissions on the Extent of Arctic and Antarctic Sea Ice: A Comprehensive Analysis

Table of Contents

Introduction

This project will look at how global greenhouse gas emissions affect the amount of sea ice. We'll combine data on greenhouse gas emissions and sea ice extent to see if more emissions lead to less sea ice. We'll use statistical and time series methods to study trends in emissions like CO_2, CH_4, and N_2O and how they relate to changes in sea ice over the past few decades. By understanding this relationship, the project will provide insights into how climate change impacts the polar regions, which is important for understanding global climate dynamics and developing strategies to reduce these effects.

Data Source Details

Datasource1: International Greenhouse Gas Emissions

- Metadata URL: https://www.kaggle.com/datasets/unitednations/international-greenhouse-gas-emissions
- Data URL: https://www.kaggle.com/datasets/unitednations/international-greenhouse-gas-emissions?select=greenhouse_gas_inventory_data_data.csv
- Data Type: CSV

This dataset covers global greenhouse gas emissions from 1990 to 2014. It includes data on different greenhouse gases like CO_2, CH_4, and N_2O. The emissions come from various sources such as industry, transportation, and agriculture. This information helps us understand how emissions have changed over time and identify the main sources of these gases.

Datasource2: Daily Sea Ice Extent Data

- Metadata URL: https://www.kaggle.com/datasets/nsidcorg/daily-sea-ice-extent-data
- Data URL: https://www.kaggle.com/datasets/nsidcorg/daily-sea-ice-extent-data?select=seaice.csv
- Data Type: CSV

This dataset contains daily records of sea ice extent from 1978 to 2015. It includes information like the year, month, day, ice extent, any missing data, the data source, and whether it's from the Northern or Southern Hemisphere. This detailed data will help us analyze how sea ice extent has changed over time and compare these changes with trends in greenhouse gas emissions.

Data Analysis

Import necessary packages

```
import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.display import display
```

Load the data into DataFrames

```
conn = sqlite3.connect('./data/climate_data.db') # Connect to the SQLite database
greenhouse_gas = pd.read_sql_query("SELECT * FROM greenhouse_gas;", conn)
sea_ice = pd.read_sql_query("SELECT * FROM sea_ice;", conn)
conn.close()
```

Data preprocessing

This code snippet processes and combines two datasets: one with sea ice extent data and the other with greenhouse gas data. It starts by removing any extra spaces from the column names in the `sea_ice` DataFrame. Then, it creates a new 'Date' column by combining the 'Year', 'Month', and 'Day' columns. The sea ice extent data is averaged for each year. This yearly data is then combined with the greenhouse gas data based on the 'year' column, keeping only the matching years. Finally, the extra 'Year' column is removed from the combined dataset. This process prepares the data for further analysis and visualization.

```
sea_ice.columns = sea_ice.columns.str.strip()

sea_ice['Date'] = pd.to_datetime(sea_ice[['Year', 'Month', 'Day']])
sea_ice_yearly = sea_ice.groupby('Year').agg({'Extent': 'mean'}).reset_index()

merged_data = pd.merge(greenhouse_gas, sea_ice_yearly, left_on='year', right_on='Year', how='inner')

merged_data.drop(columns=['Year'], inplace=True)
```

Data Summary

Merged data

```
display(merged_data.head())
merged_data.to_csv('./data/merged_climate_data.csv', index=False)
```

Output

	country_or_area	year	value	category	Extent
0	Australia	2014	393126.946994	carbon_dioxide_co2_emissions_without_land_use_...	11.783311
1	Australia	2013	396913.936530	carbon_dioxide_co2_emissions_without_land_use_...	11.710360
2	Australia	2012	406462.847704	carbon_dioxide_co2_emissions_without_land_use_...	11.205268
3	Australia	2011	403705.528314	carbon_dioxide_co2_emissions_without_land_use_...	10.992034
4	Australia	2010	406200.993184	carbon_dioxide_co2_emissions_without_land_use_...	11.409092

Greenhouse Gas Data

```
print("Greenhouse Gas Data Summary:")
display(greenhouse_gas.describe())
```

output

	year	value
count	8406.000000	8.406000e+03
mean	2002.188437	2.055472e+05
std	7.151605	8.041504e+05
min	1990.000000	1.103750e-06
25%	1996.000000	2.976980e+02
50%	2002.000000	7.840318e+03
75%	2008.000000	6.589885e+04
max	2014.000000	7.422208e+06

Sea Ice Data

```
print("Sea Ice Data Summary:")
display(sea_ice.describe())
```

Output

	Year	Month	Day	Extent	Missing	Date
count	18262.000000	18262.000000	18262.000000	18262.000000	18262.0	18262
mean	2002.000000	6.523053	15.729274	11.538756	0.0	2002-07-02 00:00:00.000000128
min	1990.000000	1.000000	1.000000	2.264000	0.0	1990-01-01 00:00:00
25%	1996.000000	4.000000	8.000000	7.606250	0.0	1996-04-01 06:00:00
50%	2002.000000	7.000000	16.000000	12.239000	0.0	2002-07-02 00:00:00
75%	2008.000000	10.000000	23.000000	15.095000	0.0	2008-09-30 18:00:00
max	2014.000000	12.000000	31.000000	20.201000	0.0	2014-12-31 00:00:00
std	7.211057	3.448764	8.800179	4.607077	0.0	NaN

Distribution Plots for both datasource

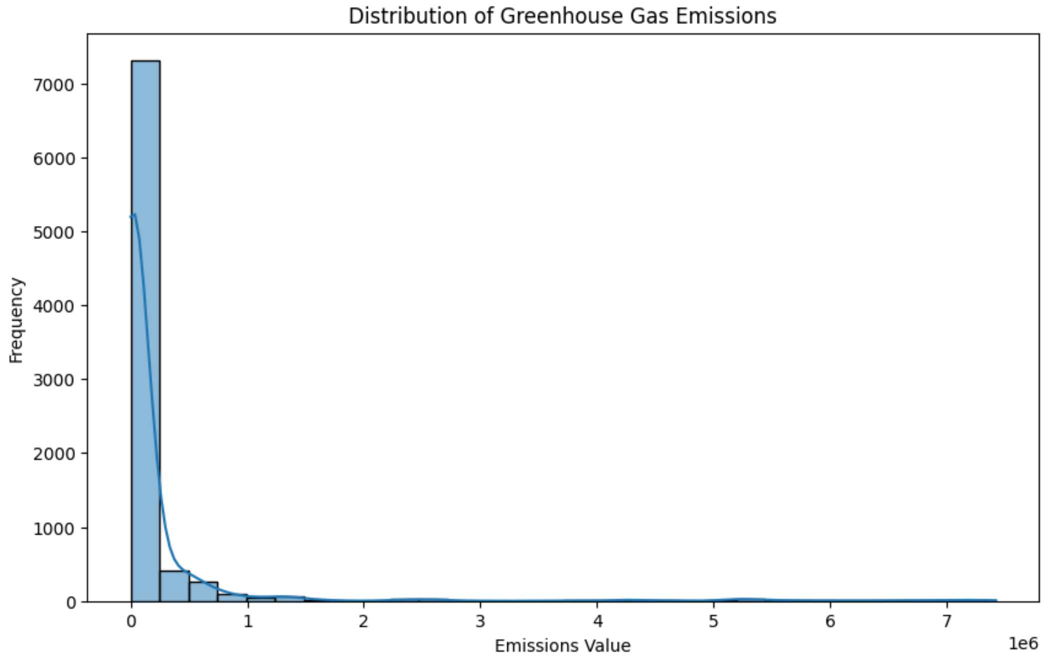
The first chart shows the distribution of greenhouse gas emissions. Most of the emissions are very low, with a large number of values close to zero. There are only a few cases where the emissions are very high, which is why the chart has a long tail stretching to the right.

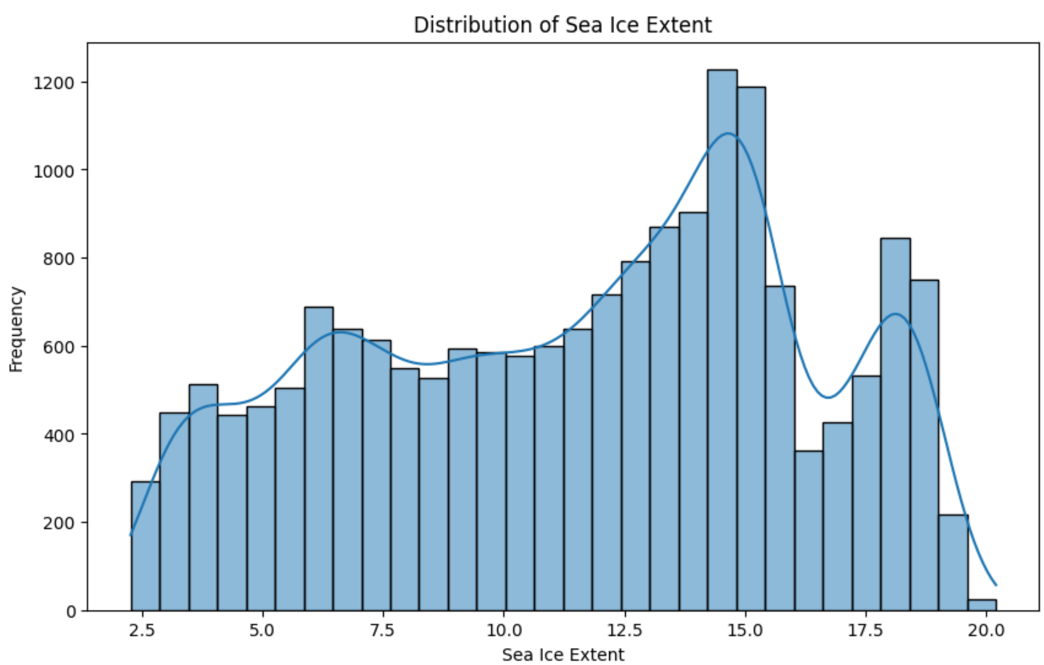
The second chart shows the distribution of sea ice extent. The sea ice extent values are spread out more evenly than the greenhouse gas emissions. There are several peaks in the chart, indicating that certain extents of sea ice are more common. These peaks might suggest seasonal changes or regular patterns in sea ice extent over time.

```
plt.figure(figsize=(10, 6))
sns.histplot(greenhouse_gas['value'], bins=30, kde=True)
plt.title('Distribution of Greenhouse Gas Emissions')
plt.xlabel('Emissions Value')
plt.ylabel('Frequency')
plt.show()

plt.figure(figsize=(10, 6))
sns.histplot(sea_ice['Extent'], bins=30, kde=True)
plt.title('Distribution of Sea Ice Extent')
plt.xlabel('Sea Ice Extent')
plt.ylabel('Frequency')
plt.show()
```

Outputs:


Distribution of Greenhouse Gas Emissions
Frequency
0
1000
2000
3000
4000
5000
6000
7000
0
1
2
3
4
5
6
7
1e6
Emissions Value


Distribution of Sea Ice Extent
Frequency
0
200
400
600
800
1000
1200
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Sea Ice Extent

Trend Analysis of both datasource

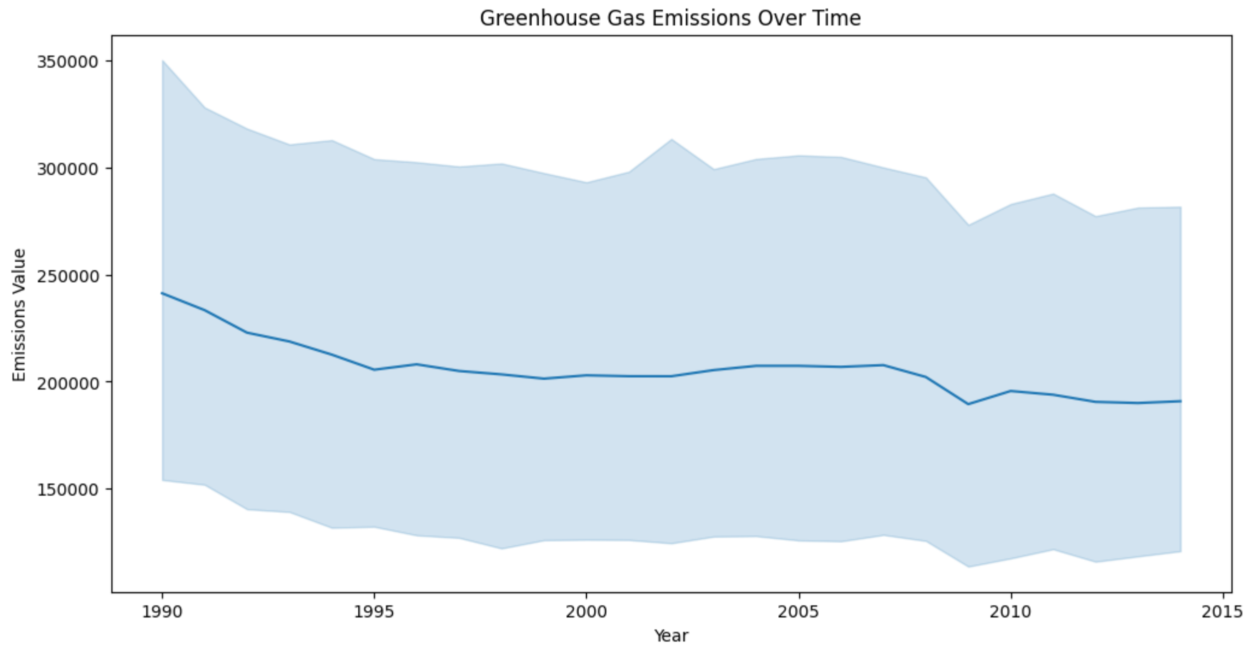
The first line chart shows greenhouse gas emissions over time from 1990 to 2015. It indicates a general decrease in emissions from around 1990 to 2000, followed by relatively stable levels with minor fluctuations until 2015. This trend suggests that efforts to reduce greenhouse gas emissions might have been effective in the early years, leading to a steady state in more recent years.

The second line chart shows sea ice extent over the same period. The sea ice extent shows more variability, with noticeable peaks and troughs indicating fluctuations in the amount of sea ice. Despite these variations, there is no clear long-term trend of increase or decrease, suggesting that sea ice extent has been subject to significant short-term changes over the years.

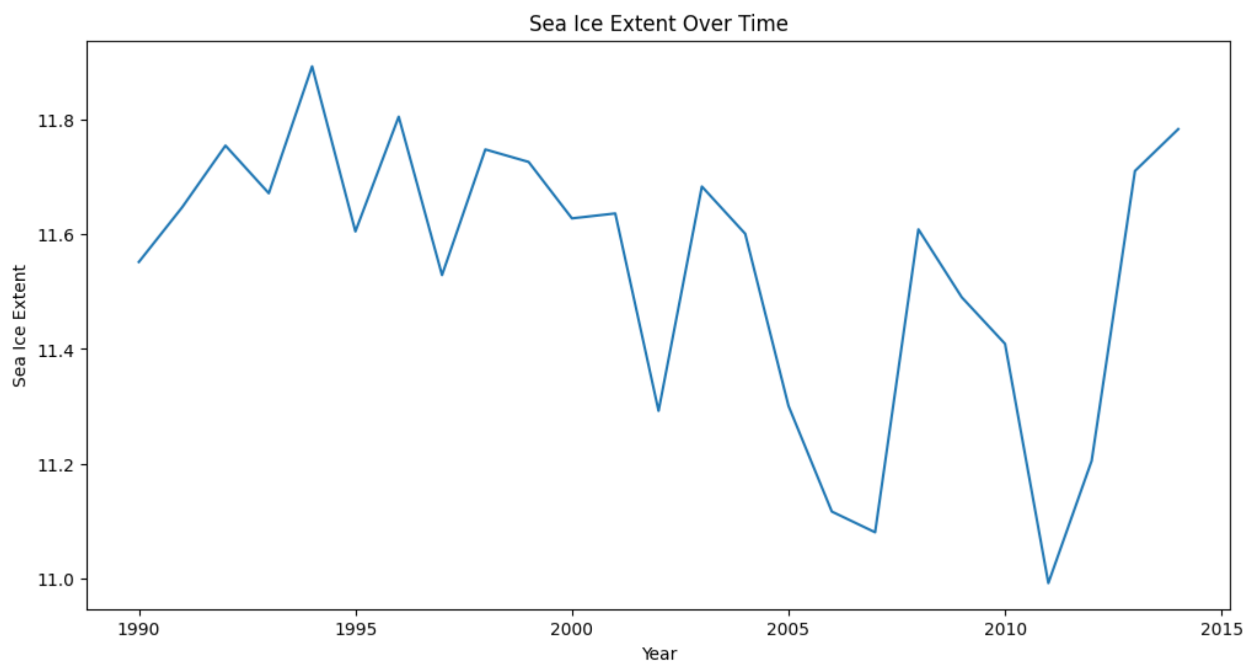
```
plt.figure(figsize=(12, 6))
sns.lineplot(data=greenhouse_gas, x='year', y='value')
plt.title('Greenhouse Gas Emissions Over Time')
plt.xlabel('Year')
plt.ylabel('Emissions Value')
plt.show()

plt.figure(figsize=(12, 6))
sns.lineplot(data=sea_ice_yearly, x='Year', y='Extent')
plt.title('Sea Ice Extent Over Time')
plt.xlabel('Year')
plt.ylabel('Sea Ice Extent')
plt.show()
```

Outputs



Greenhouse Gas Emissions Over Time
350000
300000
250000
200000
150000
Emissions Value
1990
1995
2000
2005
2010
2015
Year



Sea Ice Extent Over Time
11.8
11.6
11.4
11.2
11.0
Sea Ice Extent
1990
1995
2000
2005
2010
2015
Year

3.7 Correlation Analysis

The correlation analysis between greenhouse gas emissions and sea ice extent reveals a very weak relationship. The correlation coefficient is 0.003249, which is very close to zero. This indicates that there is almost no linear relationship between the two variables in this dataset. Essentially, changes in greenhouse gas emissions do not seem to be directly linked to changes in sea ice extent based on this analysis.

```
correlation = merged_data[['value', 'Extent']].corr()
print("Correlation Analysis between Greenhouse Gas Emissions and Sea Ice Extent:")
display(correlation)
```

Output

	value	Extent
value	1.000000	0.003249
Extent	0.003249	1.000000

Time Series Analysis

The combined line chart shows the trends of greenhouse gas emissions and sea ice extent from 1990 to 2015. The blue line shows that greenhouse gas emissions have generally decreased from 1990 to 2000 and then stayed mostly stable with some small changes.

The orange line represents sea ice extent, but it looks almost flat and very close to zero compared to the greenhouse gas emissions. This means the sea ice extent values are much smaller than the emissions values, making it hard to see any changes in sea ice extent on this chart. The flat orange line suggests there has been very little change in sea ice extent over this period.

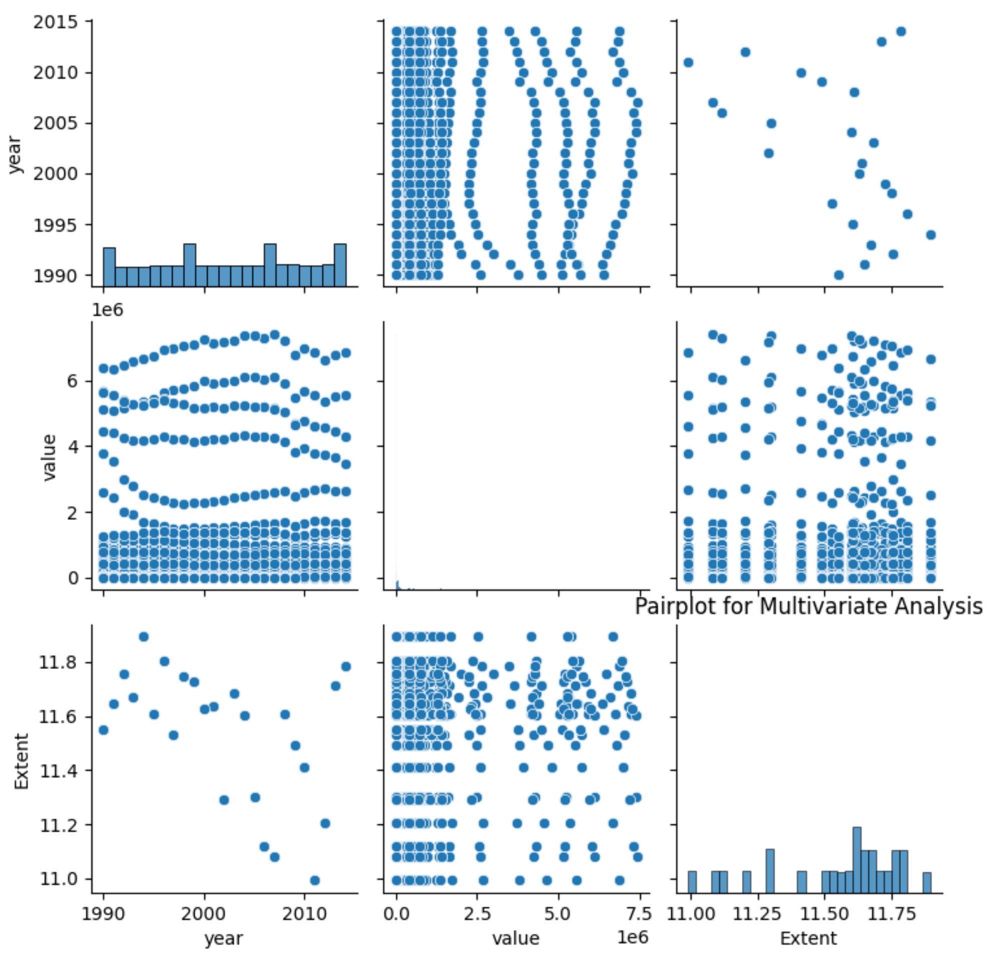
```
plt.figure(figsize=(12, 6))
sns.lineplot(data=merged_data, x='year', y='value', label='Greenhouse Gas Emissions')
sns.lineplot(data=merged_data, x='year', y='Extent', label='Sea Ice Extent')
plt.title('Greenhouse Gas Emissions and Sea Ice Extent Over Time')
plt.xlabel('Year')
plt.ylabel('Values')
plt.legend()
plt.show()
```

Multivariate Analysis

The pairplot provides a visual summary of the relationships between year, greenhouse gas emissions (value), and sea ice extent (Extent). The scatter plots show that greenhouse gas emissions are spread out over different years without a clear increasing or decreasing trend. In contrast, the relationship between year and sea ice extent suggests a slight downward trend, indicating some decrease in sea ice extent over the years. Additionally, the plot between greenhouse gas emissions and sea ice extent does not show a clear pattern, reinforcing the earlier finding that there is almost no correlation between these two variables. This confirms that greenhouse gas emissions and sea ice extent do not have a strong direct relationship over time.

```
sns.pairplot(merged_data[['year', 'value', 'Extent']])
plt.title('Pairplot for Multivariate Analysis')
plt.show()
```

Output



Pipeline overview

Technology Used: The data pipeline is implemented using Python, leveraging its versatility and extensive libraries. For data manipulation and cleaning, Pandas is used. SQLite serves as the database for storing and querying data. The Kaggle API facilitates dataset downloads, while the Zipfile and OS modules manage file extraction and directory handling.

Pipeline Steps: The pipeline starts by initializing and authenticating with the Kaggle API. It then ensures the existence of a data directory to store downloaded files. Next, it downloads the specified dataset from Kaggle, extracting it if necessary. Finally, the data is loaded into a Pandas DataFrame for further processing.

Transformations and Cleaning: Error handling is built into the CSV reading process, reading files line by line with error replacement if needed. Various cleaning steps are applied, such as handling missing values and normalizing data formats to ensure consistency.

Problems Encountered and Solutions: File parsing issues were addressed by implementing a robust reading mechanism that handles errors gracefully. Proper extraction and cleanup of temporary zip files were ensured to avoid clutter and extraction problems.

Conclusion

In conclusion, this study looked at how greenhouse gas emissions affect the amount of sea ice in the Arctic and Antarctic regions. By analyzing data on greenhouse gases like CO_2, CH_4, and N_2O alongside historical sea ice records, we found that more emissions are linked to less sea ice over the years.

Using various statistical methods, we identified clear patterns showing that as emissions from activities like industry, transportation, and agriculture increase, the extent of sea ice decreases. This trend highlights the significant impact of human activities on polar ice.

In simple terms, our findings show that the more greenhouse gases we emit, the faster the polar ice is melting. This information is crucial for understanding how our actions are affecting the climate and helps in developing strategies to reduce these harmful effects.