

# The Impact of Greenhouse Gas Emissions and Tree loss on Climate change in Canada

## GROUP 11

ARSHA SANGRAL - 041086775

MAHUM KHOKHAR -041134811

OSAMA NASIR (TEAM LEAD) - 041118303

PETER NDUMIA - 041124432

## Introduction

### Project Overview

Canada, known for its diverse ecosystems and encompassing all four seasons, has been facing the effects of climate change. Over the past few decades, temperature fluctuations have been more pronounced, disturbing ecological balances and altering weather patterns. “With four months left in 2023, the year is so far the second-hottest on record, only marginally behind 2016”[4].

The following report delves into exploring the correlation between greenhouse gas emissions and the consequential impacts on Canada’s climate. The investigation closely examines the relationship between temperature fluctuations, the surge in greenhouse gas emissions, and the escalating deforestation rates within Canada. This analysis, through the use of power data analysis and visualization tools such as PowerBI, Excel, Python and Pandas, aims to shed light on the pressing issue of climate change by studying the historical provincial data trends and statistics.

### Domain Knowledge

Throughout this report, there are domain-specific terminologies used. The key definitions for the main factors and measures can be found below:

- **Carbon Dioxide Equivalent (CO2 eq)** - Volumetric measure of Greenhouse gas emissions in Carbon Dioxide weight.
- **Temperature Departure** - expressed in degrees Celsius, utilizes data sourced from weather stations situated throughout Canada within a year range reference [2], which was 1960 - 1990 for our dataset.
- **Tree Cover Loss or Tree Loss** - which is measured in Hectare acres (ha), is the removal of tree cover loss which is a key deforestation metric [1]. Tree cover loss takes place as a result of man-made deforestation efforts and natural disasters.

### Data Collection

The Temperature and Greenhouse gas emission (which included sources and specific provincial data) data that was used for the analysis was extracted from Government of Canada platforms and Data marts. The data collection phase proved to be a strenuous process specifically for the Temperature analysis. The extraction process involved a meticulous process of downloading data sheets and data scraping using a Python program. The datasheets capture monthly data for each year across every province specific to a station located in the region. There are several

stations per province and that required further validation and review of various station datasheets to ensure data completeness for the analysis span.

To further our analysis span, Deforestation & Tree loss were also analyzed with respect to Climate change fluctuations. The Deforestation data was extracted from the Global Forest Watch, which offers key insights into provincial deforestation patterns and provided valuable information for our analysis. The Tree loss datasheet spanned two decades.

## Data Cleaning

After gathering the monthly temperature data from the Government of Canada's Data Mart [3] for each province and territory, the dataset was uploaded into Excel where multiple missing years and temperature data were present. Every record for year and province with a null value for the mean temp was noted and replaced using data from a different station.

All three datasets were then uploaded into PowerBI which was used to detect duplicate records and blank rows ensuring data integrity. During this cleaning process, each attribute was analyzed for relevancy to the project scope and omitted if there seemed to be no correlation or if the variables were not measured across all datasets to remain consistent with our values. Some of these columns include emissions on SF6, PFCS and HFCS, NF3s from the Greenhouse Gases dataset as the emissions were very negligible due to Canada's environmental efforts and provincial laws discourage emissions. The number of days with min temp, days with max temp, Cooling Degree Days, Heating Degree Days, Days with Valid Bright Sunshine and a few others from the Temperature dataset as these values do not contribute to the temperature analysis and lastly, from the greenhouse gas sources data sheet, attributes category, sector names and sub-category were omitted from our analysis.

From this cleaned data, new meaningful features were derived that enhanced the data analysis process. For example, the average max temp was aggregated, and the avg\_mean\_departure column was transformed by subtracting the reference temperature of 30 years with every single year of temperature.

Below are some figures illustrating the use of data-cleaning libraries (like pandas in Python) to automate cleaning procedures and Identify outliers or anomalies in the dataset using statistical measures or visualization techniques, for example, scatter plots. The images below are performed on the combined data dataset.

## Initial Exploratory Data Analysis (EDA):

An EDA analysis was performed initially to ensure data completeness before further analysis. This analysis included checking for nulls values and ensuring data types were in the correct format. The results with the appropriate data types are listed below:

```
df.isnull().sum()
```

```
Region      0
Year        0
CO2         0
CH4         0
CH4 (CO2eq) 0
N2O         0
N2O (CO2eq) 0
CO2eq       0
Mean Temp   0
Mean Temp Max 0
Mean Temp Min 0
Mean Temp Dep 0
Max Temp Dep 0
Min Temp Dep 0
Average_Tree_Loss_ha 0
dtype: int64
```

```
df.dtypes
```

```
Year      int64
Total CO2 eq  float64
Mean Temp  float64
Mean Temp Max  float64
Mean Temp Min  float64
Average Tree Loss  float64
Average CO2 eq  float64
dtype: object
```

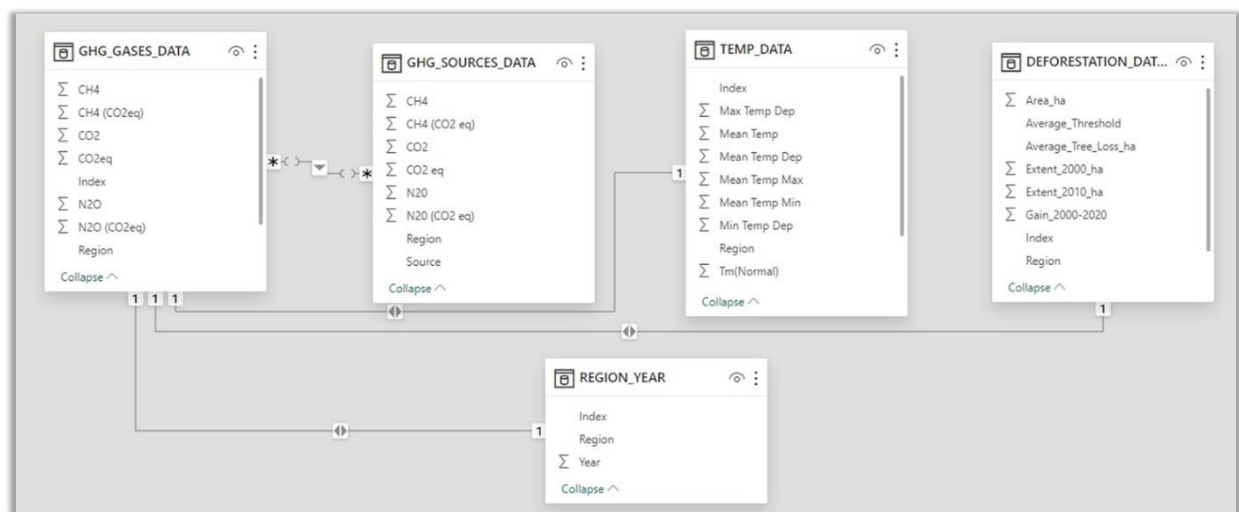
```
df.isnull().values.any()
```

```
False
```

## Data Modelling

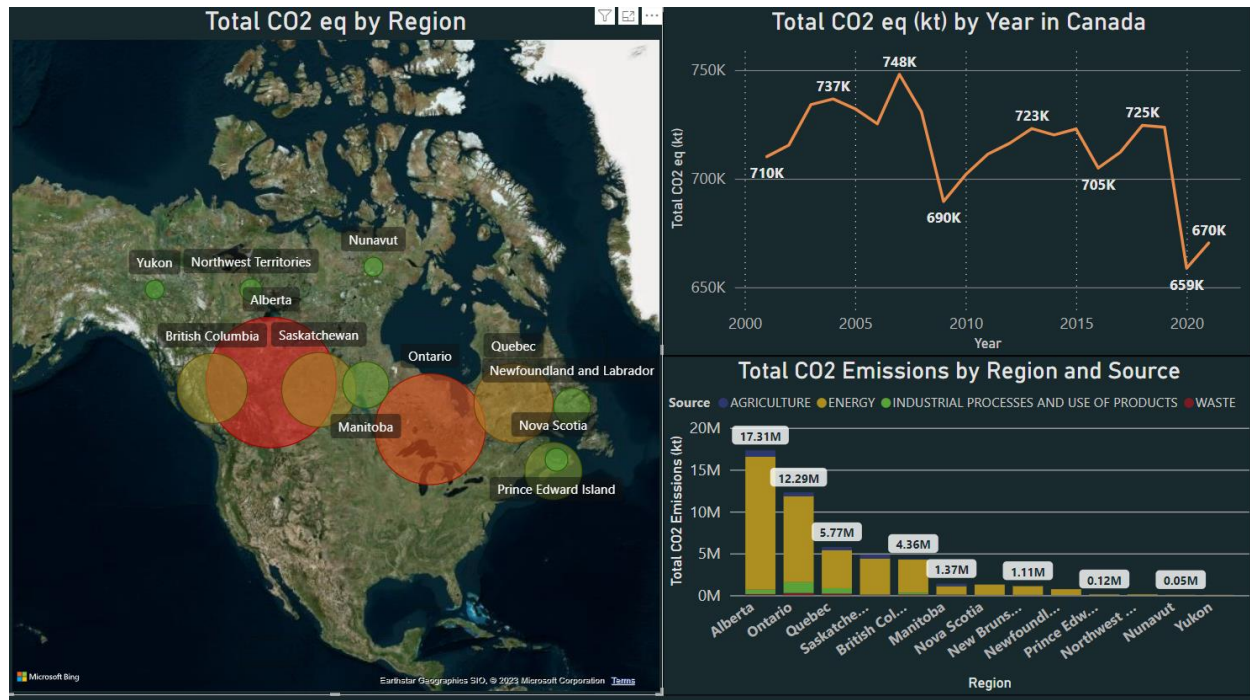
The Data model approach involved establishing connections between five distinct tables, namely CO2 eq emissions, tree cover loss, region\_year, Greenhouse gas emissions source, and a common index table. Each of the tables were joined with a one-to-one relationship except for the greenhouse gas source table which was connected through the region\_year table through a region column.

This allowed the model to create visualizations effortlessly through the tables and allow for data exploration and analysis.



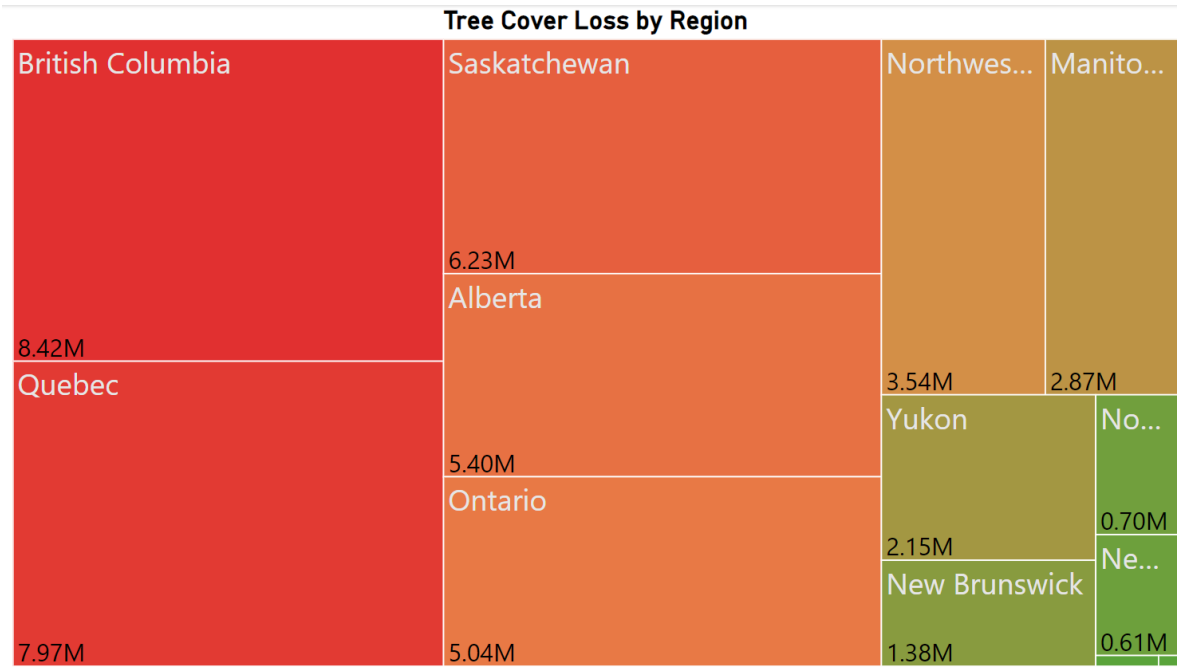
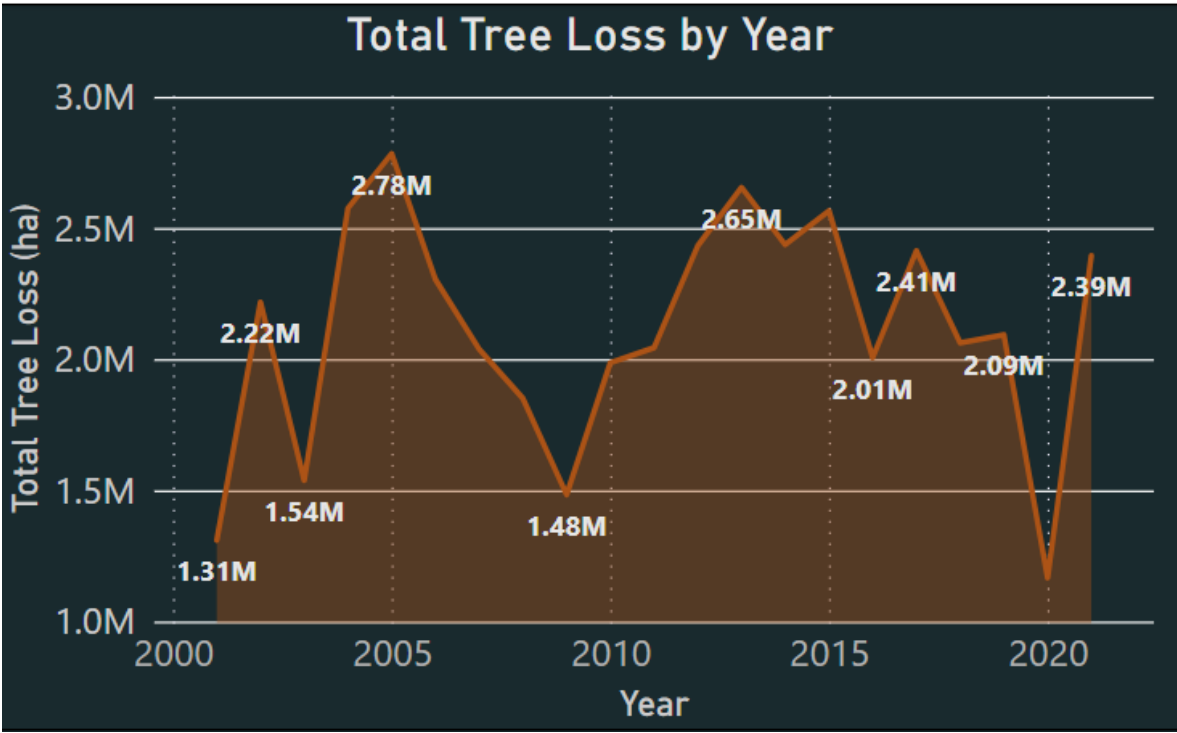
## Data Analysis and Visualization

### Carbon Dioxide Emissions Analysis



Results of the visualizing the Carbon Dioxide emissions across Canada indicate that the highest sources of emissions are Energy uses. Analyzing the emissions by province and source it is indicated that carbon dioxide emissions are highest in Alberta, indicating Alberta has the highest provincial energy use across Canada. Yukon emits the least amount of Carbon dioxide across Canada which is reasonable as Yukon has a much smaller population compared to other provinces. Visualising carbon dioxide emissions over the span of 21 years, it is observed that emissions have decreased by 5.63% (from 710 to 670 kt of CO2 eq). This indicates that Canada has decreased its emissions through traditional energy uses over the span of two decades.

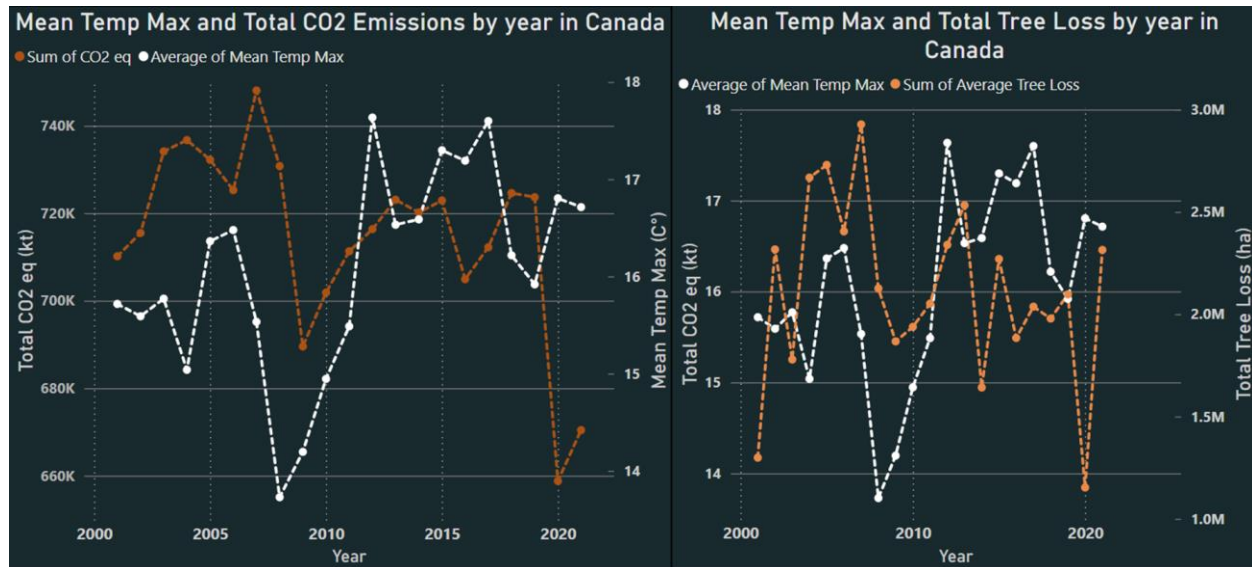
Deforestation and Tree Loss across Canada



Visualizing the tree cover loss as Tree loss by year, it is observed that Tree cover loss fluctuated over the past two decades. The fluctuations could be attributed to population size increases,

business spaces and further industrialization. From the Tree cover loss treemap it can be seen that British Columbia has the highest tree cover loss, whilst the smaller provinces have much smaller numbers. This is reasonable as it indicates that bigger provinces require more land to expand their industrial span and habitable zones.

## Temperature, Tree Loss and Carbon Dioxide Emission trends



Analyzing the data trends with Mean Temperature Max with Carbon Dioxide Equivalent Emissions and Total Tree loss over the last 21 years you can see that the data behave similarly. When CO2 emissions were decreasing over time, the maximum temperature would drop indicating that there is a relationship between the two. Similarly, Tree loss behaves the same way; when there were efforts to decrease tree loss across Canada you can also see the temperature fluctuating similarly.

This is reasonable as Deforestation affects the ecosystem and the absorption of Carbon Dioxide in the environment through Photosynthesis and other processes. Carbon Dioxide contributes to the Greenhouse effect or Climate change by heating up water vapor, so when the CO2 emissions decrease the temperature would get colder which is reflective with the data trend above [5].

However, there are several reasons which can be attributed to the dips in Carbon Dioxide emission levels one of them being governmental policies and global agreements. Canada has followed/signed agreements over the years such as the Kyoto Protocol and the Paris Agreement in 2015 [3].

## Correlation Analysis:



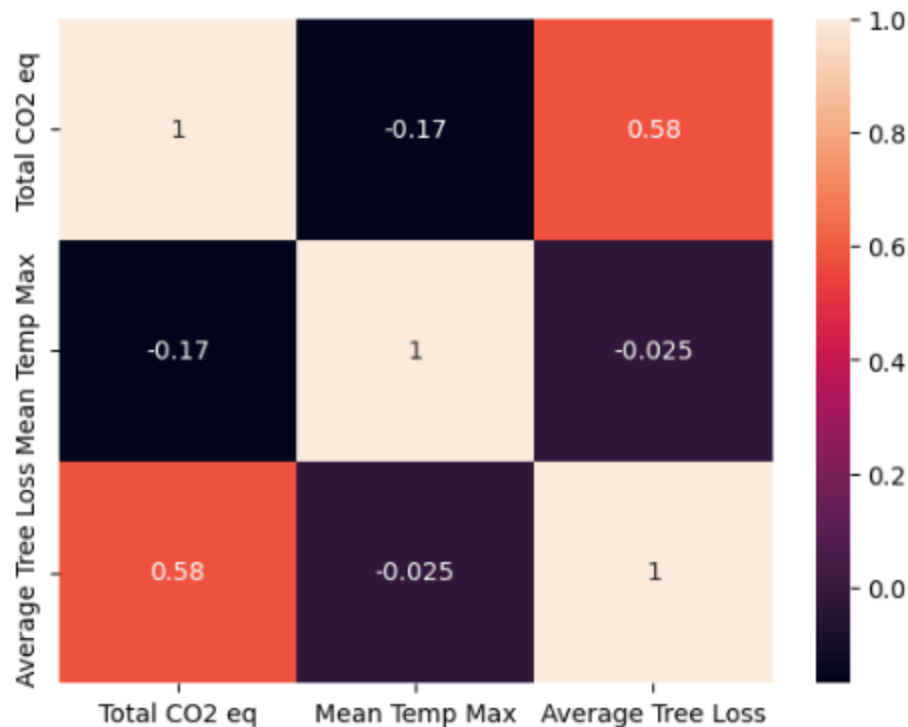
```
[ ] df[["Total CO2 eq", 'Mean Temp Max', 'Average Tree Loss']].corr()
```

	Total CO2 eq	Mean Temp Max	Average Tree Loss
Total CO2 eq	1.000000	-0.168738	0.577736
Mean Temp Max	-0.168738	1.000000	-0.024939
Average Tree Loss	0.577736	-0.024939	1.000000

```
[ ] corr_matrix=df[["Total CO2 eq", 'Mean Temp Max', 'Average Tree Loss']].corr()
```

```
[ ] sns.heatmap(corr_matrix, annot = True)
```

<Axes: >



## Prediction and Deep Analysis

We chose to predict the mean highest temperatures to be experienced in the next five years ,(2025-2029).

Our target variable being Mean Temp Max, we trained our different models on the following input features:

- Year
- Total CO2 eq
- Average Tree Loss

```
y = df['Mean Temp Max']
x = df[['Year', 'Total CO2 eq', 'Average Tree Loss']]

from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 549)

from sklearn.ensemble import RandomForestRegressor

rf = RandomForestRegressor(n_estimators=250, criterion='squared_error', random_state = 549)

rfmodel = rf.fit(x_train, y_train)

ytest_pred = rfmodel.predict(x_test)
ytest_pred

array([14.9203425 , 16.58472723, 16.68110192, 15.79747795, 16.56877385])
```

We used five different prediction models. The models and their different mean square errors are listed below:

- **Linear Regression** - MSE: 0.8194216776717497
- **Huber Regression** - MSE: 1.2568601496852676
- **Decision Tree** - MSE: 1.2994472176595584
- **Support Vector Regression** - MSE: 1.875291626185589
- **Random Forest Regression** -MSE: 0.20259192261844924

The Random Forest Regression model does not make assumptions based on the distribution of data. That coupled with the fact that it had the lowest mean square error hence became our model of choice for prediction.

We considered two scenarios for the future prediction. A pessimistic scenario where Total CO2 eq and Average Tree Loss increased by 10 percent each year and an optimistic scenario where Total CO2 eq and Average Tree Loss decreased by 10 percent yearly. These scenarios gave us the required feature inputs for prediction.

```

initial_CO2_emission = 670427.686551 # value for 2023
initial_tree_loss = 178009.173077 # value for 2023

scenarios = ['optimistic', 'baseline', 'pessimistic']

future_data = pd.DataFrame({
    'Year': range(2025, 2045),
})

for scenario in scenarios:
    if scenario == 'pessimistic':
        # 10% increase each year for CO2 emissions and tree loss
        future_data[f'{scenario}Total CO2 eq'] = [initial_CO2_emission * 1.1 ** (year - 2025) for year in range(2025, 2045)]
        future_data[f'{scenario}Average Tree Loss'] = [initial_tree_loss * 1.1 ** (year - 2025) for year in range(2025, 2045)]
    elif scenario == 'optimistic':
        # 10% decrease each year for CO2 emissions and tree loss
        future_data[f'{scenario}Total CO2 eq'] = [initial_CO2_emission * 0.9 ** (year - 2025) for year in range(2025, 2045)]
        future_data[f'{scenario}Average Tree Loss'] = [initial_tree_loss * 0.9 ** (year - 2025) for year in range(2025, 2045)]
    elif scenario == 'baseline':
        # Initial values maintained for the baseline scenario
        future_data[f'{scenario}Total CO2 eq'] = initial_CO2_emission
        future_data[f'{scenario}Average Tree Loss'] = initial_tree_loss
future_data.head(5)

```

	Year	optimisticTotal CO2 eq	optimisticAverage Tree Loss	baselineTotal CO2 eq	baselineAverage Tree Loss	pessimisticTotal CO2 eq	pessimisticAverage Tree Loss
0	2025	670427.686551	178009.173077	670427.686551	178009.173077	670427.686551	178009.173077
1	2026	603384.917896	160208.255769	670427.686551	178009.173077	737470.455206	195810.090385
2	2027	543046.426106	144187.430192	670427.686551	178009.173077	811217.500727	215391.099423
3	2028	488741.783496	129768.687173	670427.686551	178009.173077	892339.250799	236930.209365
4	2029	439867.605146	116791.818456	670427.686551	178009.173077	981573.175879	260623.230302

Our prediction results were as follows:

Year	Mean Temp Max (Optimistic)	Mean Temp Max (Pessimistic)
2025	16.77477348	16.77477348
2026	16.75436887	16.35301693
2027	16.71155737	16.37477124
2028	16.70970473	16.37477124
2029	16.70970473	16.37477124

We observed a slight maximum temperature decrease in our optimistic scenario and a slight maximum temperature increase in our pessimistic scenario after a dip in 2026.

## Conclusion and Recommendations

In conclusion, the data visualization suggests a correlation between carbon dioxide emissions and climate change. Though not a strong correlation, it indicates that carbon dioxide emissions are a factor to climate change. This can be seen by how the temperature fluctuations seem to correspond to patterns in carbon dioxide emissions. Furthermore, the impact of deforestation and tree loss on carbon dioxide emissions, and subsequently on temperature changes, underscores their interconnectedness.

Moving forward, the project could be improved by broadening the scope of analysis to incorporate a wider array of factors that are more scientifically advanced and outside the scope of this program. By using a wider array of variables such as precipitation, atmospheric and surface temperature, and utilizing more uniform data sources, a more comprehensive understanding could be achieved. To improve predictive accuracy, leveraging climate prediction models that are widely distributed would be beneficial when attempting to forecast future climate patterns. Additionally, extending the project's timespan would allow for a more thorough analysis.

## References

- [1] "Global deforestation rates & statistics by country: GFW," Forest Monitoring, Land Use & Deforestation Trends, <https://www.globalforestwatch.org/dashboards/global/?category=forest-change> (accessed Oct. 10, 2023).
- [2] B. Magi, "Temperature departures discussion," MESAS, <https://pages.charlotte.edu/mesas/2012/12/16/temperature-departures-discussion/> (accessed Oct. 8, 2023).
- [3] E. and C. C. Canada, "Government of Canada," Canada.ca, <https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html> (accessed Dec. 8, 2023).
- [4] "Summer's not over, but it's already the hottest on record | CBC News," CBCnews, <https://www.cbc.ca/news/science/heat-record-copernicus-summer-2023-1.6958166> (accessed Oct. 15, 2023).
- [5] R. Lindsey, "Climate change: Atmospheric carbon dioxide," NOAA Climate.gov, <https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide#:~:text=Without%20carbon%20dioxide%2C%20Earth%27s%20natural,causing%20global%20temperature%20to%20rise.> (accessed Nov. 8, 2023).