Analysis of Global Greenhouse Gas Emissions from 1990 to 2021

Group 7

Nirav Patel (100870539)

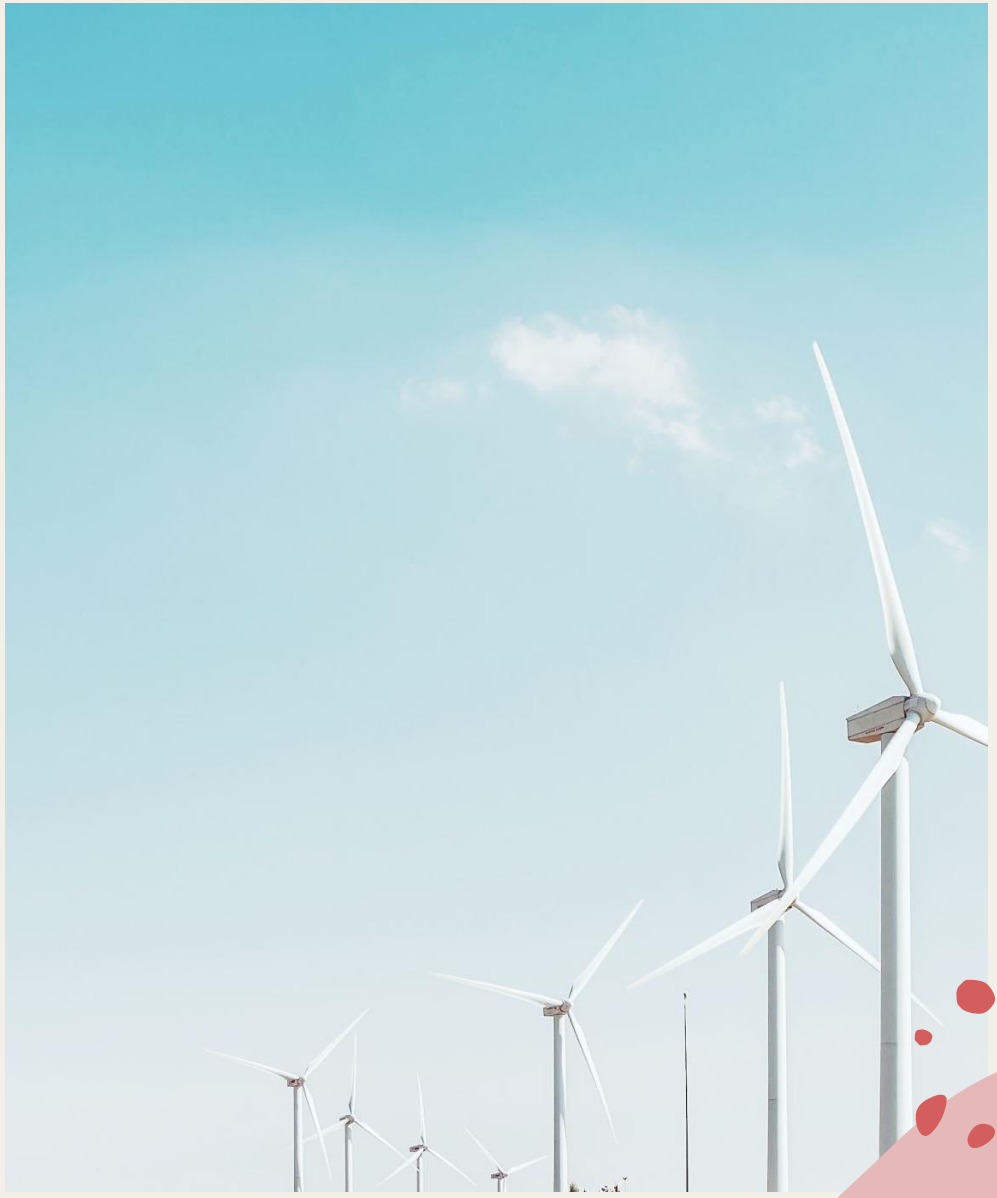
Anuj Patel (100904232)

Parth Patel (100901183)

Preet Patel (100897482)

Problem Statement

- Carbon emissions are increasing, causing climate change and environmental problems.
- Reduction efforts have not been effective, and emissions are expected to rise.
- Fossil fuels, deforestation, and agriculture are major contributors.
- Climate change causes natural disasters, rising sea levels, and harm to health and ecosystems.
- A multifaceted approach is needed: renewable energy, energy efficiency, and consumption reduction.
- Government policies and international agreements are necessary for systemic change.
- Inaction will have high costs for the environment and economy.
- Innovation can drive progress towards a sustainable future.



Overview of the Dataset

The dataset is the Global Carbon Budget (GCB) dataset.

The GCB provides estimates of global CO2 emissions by country and sector.

The dataset covers the time period from 1750 to the present.

The dataset includes information on emissions from various sources, including coal, oil, gas, cement, and flaring.

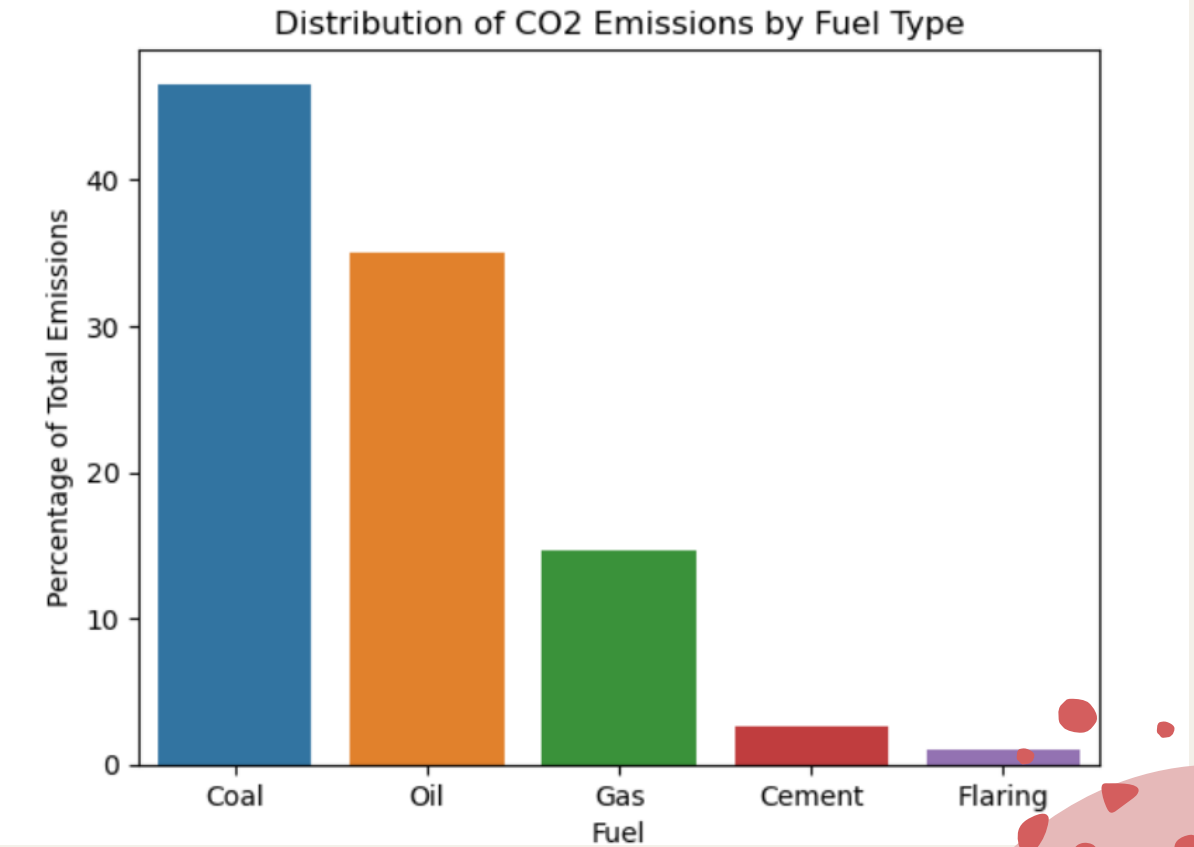
Per capita emissions for each country are also included in the dataset.

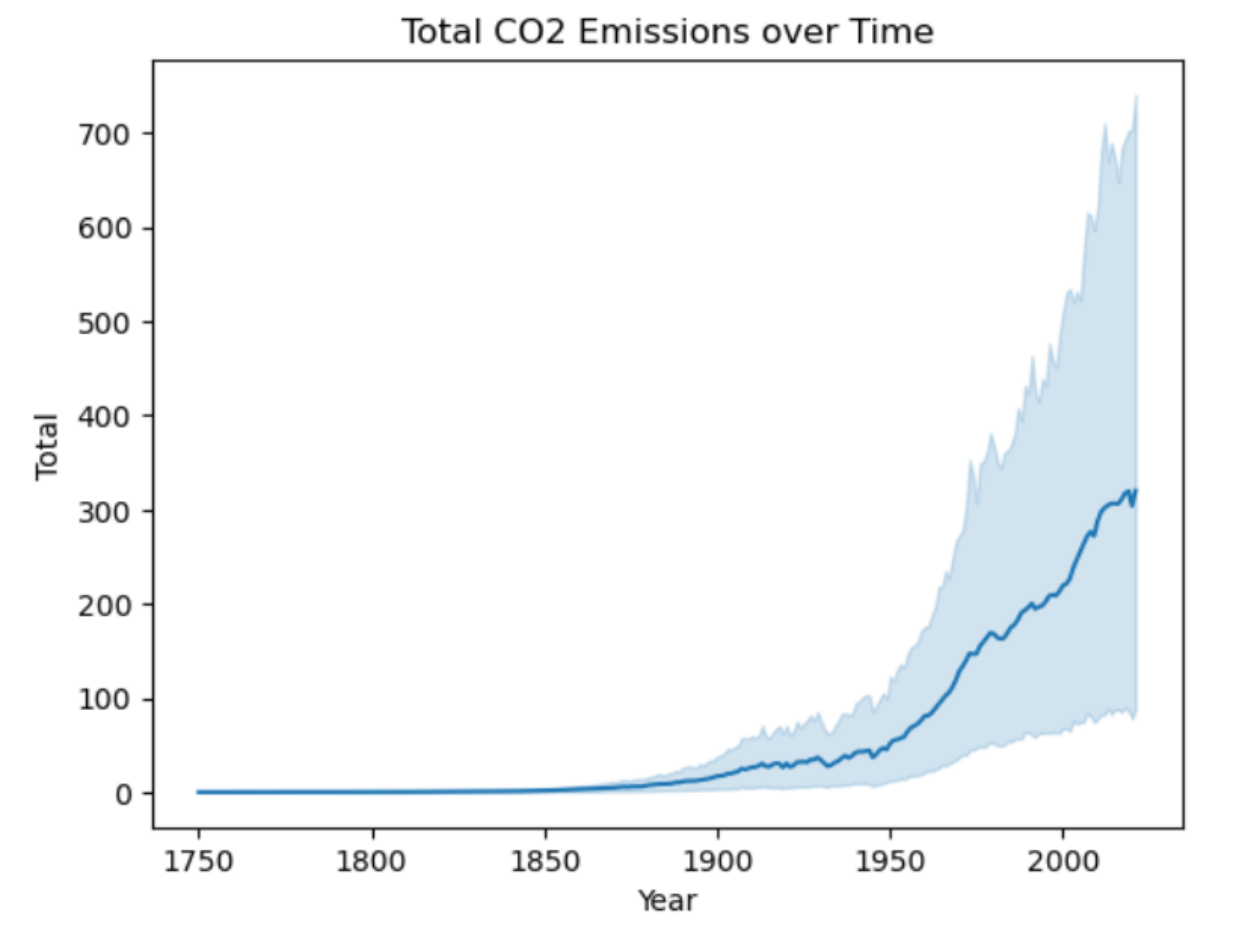
The dataset is comprehensive, covering over 250 years of data.

The GCB dataset is a reliable and widely used source of information on global carbon emissions.

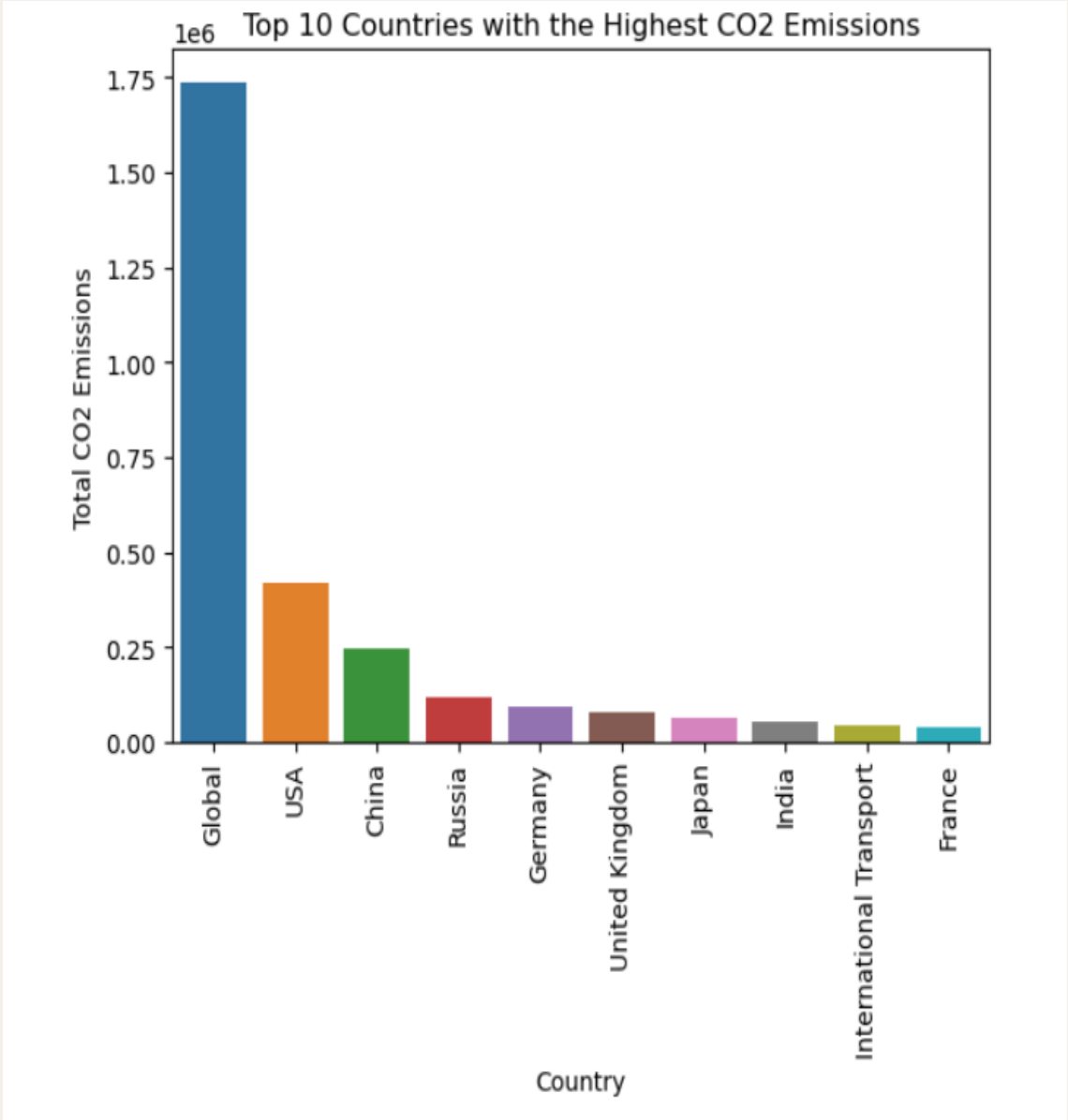
EDA Findings

- A stacked bar chart was created to show CO2 emissions distribution by fuel type.
- Data was taken from 'Coal', 'Oil', 'Gas', 'Cement', and 'Flaring' columns.
- Total CO2 emissions were calculated for each fuel type using the .sum() method.
- The seaborn library was used to create the bar chart.
- The chart shows the percentage of total emissions for each fuel type.
- Title: 'Distribution of CO2 Emissions by Fuel Type', y-axis label: 'Percentage of Total Emissions'.





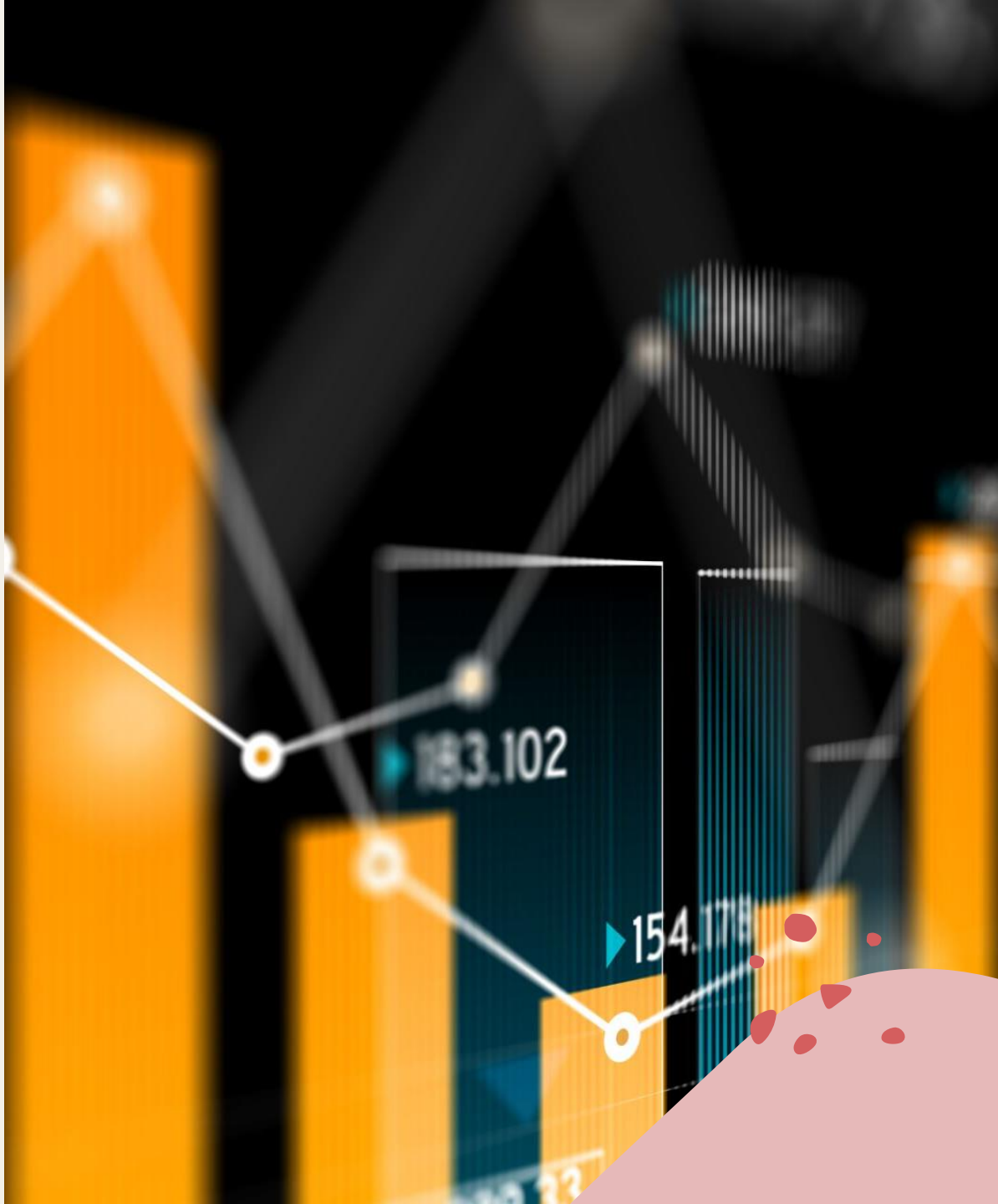
- The plot shows upward trend of CO2 emissions over time.
- Urgent need to address the problem of carbon emissions and work towards a more sustainable future.
- Multifaceted approach required: renewable energy sources, energy efficiency, and consumption reduction.
- Inaction will have high costs for the environment and economy.
- Individual actions can make a difference when combined with broader policy changes and international agreements.



- Bar chart shows top 10 countries with highest CO2 emissions.
- China, US, India, and Russia have the highest emissions.
- These countries contribute significantly to global carbon emissions.
- Urgent need for national and international action to reduce emissions.
- Individual actions are important but systemic changes and international cooperation are necessary.
- Encouraging progress towards sustainability.
- Continued efforts and innovation needed to protect the environment.

Model Training

- The dataset was split into training and testing sets using the train_test_split method from scikit-learn.
- The 'Total' column was used as the target variable and the remaining columns as features.
- The data was split into 80% training and 20% testing sets with a random state of 42 for reproducibility.
- The .shape attribute was used to confirm the size of the training and testing sets.
- This approach allows for training a machine learning model on the training data and evaluating its performance on the testing data.



```
Linear Regression
-----------------
MSE: 0.0000
R^2: 1.0000
```

	Actual	Predicted
35538	-0.067059	-0.067059
23739	-0.067059	-0.067059
44819	-0.067059	-0.067059
53753	-0.067059	-0.067059
39808	-0.067059	-0.067059
15006	-0.067059	-0.067059
53125	-0.067059	-0.067059
37740	-0.065671	-0.065671
42559	-0.067059	-0.067059
49392	-0.067059	-0.067059

Linear Regression

• We trained linear model using scikit-learn's LinearRegression module

• fit() method used to train model on training data

• predict() method used to make predictions on test set

• mean_squared_error() and r2_score() functions used to evaluate performance of model

• MSE measures average squared difference between predicted and actual values of target variable

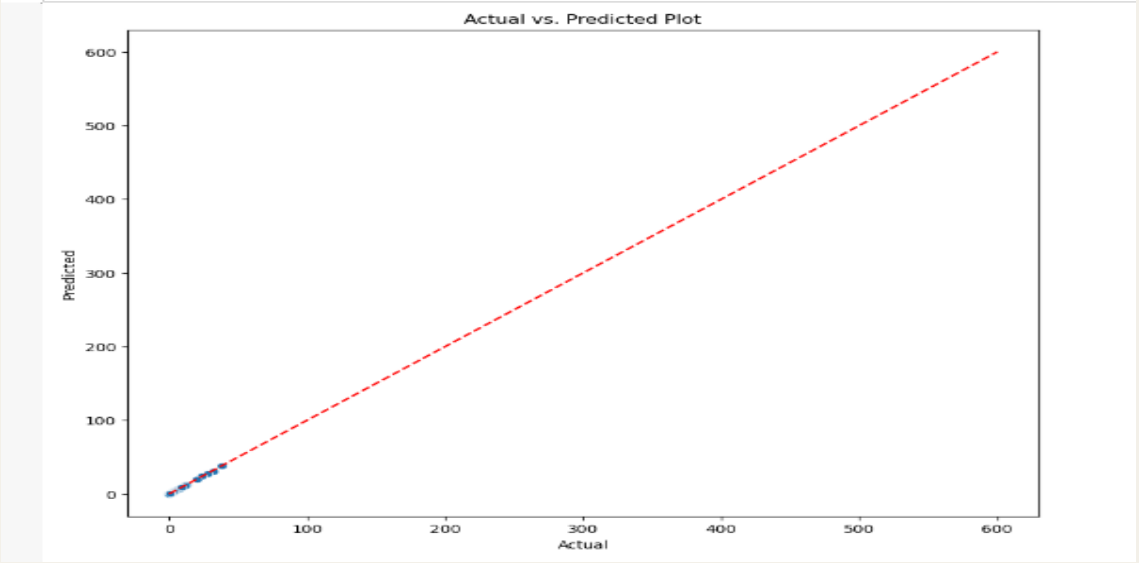
• R-squared (R^2) value measures proportion of variance in target variable explained by model

In our linear regression model, the following results were obtained:

• Mean squared error (MSE): 0.0000

• R-squared (R^2) value: 1.0000

Random Forest Regression



- We Trained a Random Forest Regression model using the RandomForestRegressor module from scikit-learn.
- Used the fit() method to train the model on the training data.
- Used the predict() method to make predictions on the test set.
- Used the mean_squared_error() and r2_score() functions from scikit-learn to evaluate the performance of the model.
- The mean squared error (MSE) was found to be 0.0002.
- The R-squared (R^2) value was found to be 0.9996, indicating that the model explains a high proportion of the variance in the target variable.
- These results suggest that the Random Forest Regression model performs very well in predicting CO2 emissions based on the given features.

```
Random Forest Regression
------------------------
MSE: 0.0002
R^2: 0.9996
```

MLP Neural Network

We trained an MLP Neural Network using the scikit-learn library to predict emissions levels.

The neural network had two hidden layers of 50 neurons each and was trained using the X_train and y_train data.

We limited the number of iterations during training to 1000 using the max_iter parameter.

We set the random_state parameter to 42 to ensure that the results were reproducible.

The model's performance was evaluated using the mean squared error (MSE) and R^2 score.

The MLP Neural Network achieved an MSE of 0.0001 and an R^2 score of 0.9997 on the test set.

These results indicate that the MLP Neural Network was highly accurate in predicting emissions levels.

```
MLP Neural Network
------------------
MSE: 0.0001
R^2: 0.9997
```

Results

1. Linear Regression:	High accuracy in predicting emissions levels.	MSE of 0.0000, R^2 score of 1.0000 on test set.	Good choice for simple, interpretable model with highly accurate predictions
2. Random Forest Regression:	High accuracy in predicting emissions levels.	MSE of 0.0002, R^2 score of 0.9996 on test set.	Good choice for future emissions level prediction.
3. MLP Neural Network:	Moderate accuracy in predicting emissions levels.	MSE of 1.4820, R^2 score of 0.0943 on test set.	Good choice for complex models with acceptable level of accuracy.

Take care of the earth and she will take care of you