This project focuses on predicting CO2 emissions using ML.

**1. Data Exploration-Preprocessing**: The Python code fetches data using API and saves as CSV file. If request fails, it returns a failed status. The initial step involves identifying and addressing missing values, consistent data types, checking duplicates, and validating that numerical features contain no negative values.

**2. Feature Engineering:** Features are categorized as numerical or categorical. Scaling and encoding are applied, with a pipeline used to streamline preprocessing and reduce data leakage risks.

**3.** **Data Visualization & Outlier Detection:** Visual tools like boxplots help identify outliers in CO2 emissions, with Z-score and IQR methods used to detect and handle these outliers.

**4**.**Feature Selection & ANOVA**: ANOVA test assesses the significance of categorical variables on CO2 emissions.

**5**. **Model Building & Evaluation**: ML models like Linear Regression, Random Forest, and XGBoost are trained using pipeline, with the data split into training and test sets. Model performance is evaluated using metrics like MAE, MSE, RMSE, R-squared.

**6.** **Outlier Treatment & Feature Transformation**: Outliers are managed through log transformation, scaling, and capping at the 99th percentile, integrated into the pipeline for consistency.

**7.** **Hyperparameter Tuning & Cross-Validation**: GridSearchCV is used for tuning model hyperparameters, and cross-validation ensures robust performance estimates.

**8.** **Final Evaluation**: The final model is evaluated on the test set, with key metrics such as RMSE, MAE, and R-squared used to report its performance. A residuals analysis is conducted to assess the normality and distribution of errors, to ensure the model's accuracy and reliability. Learning curves are plotted to identify any overfitting or underfitting issues in the model,like Random Forest and XGBoost.

**Summary**: This project offers a detailed, pipeline-driven approach to predicting CO2 emissions, enhancing reproducibility and reliability.

**Contributions**: It was a complete group effort and everyone contributed as per their understanding.