TUNIS BUSINESS SCHOOL
UNIVERSITY OF TUNIS

OBJECT-ORIENTED PROGRAMMING

ANALYZING AND PREDICTING CO2 AND NH4 EMISSIONS

PROFESSOR: PROF. AMENI AZOUZ

TEAM:
AHMED KHARRAT
DHAKER MESSAOUD
EYA TRABELSI
WAEL ZAIER
YESMINE CHALGHAM

28/01/2025

WHY THIS TOPIC?

Agriculture accounts for roughly **10-12% of global greenhouse gas (GHG) emissions**, with carbon dioxide (CO2) and methane (CH4) being the primary gases.

The potential to mitigate emissions through improved practices could offset up to **20%** of global agricultural emissions.

WHY THIS PROJECT?

INTRODUCTION

Objective:
Develop a web app to analyze and predict emissions, offering statistical insights and forecasts through visualizations and machine learning models.

Relevance:
Understanding emissions trends is vital for environmental decision-making.

Goals:

- Manage and analyze emissions data.
- Predict trends using machine learning.
- Visualize data for actionable insights.

SYSTEM REQUIREMENTS

1

FUNCTIONAL REQUIREMENTS:

- Upload and process emissions data.
- Perform statistical analysis (e.g., mean, regression).
- Generate predictive models for future trends.
- Visualize data interactively.

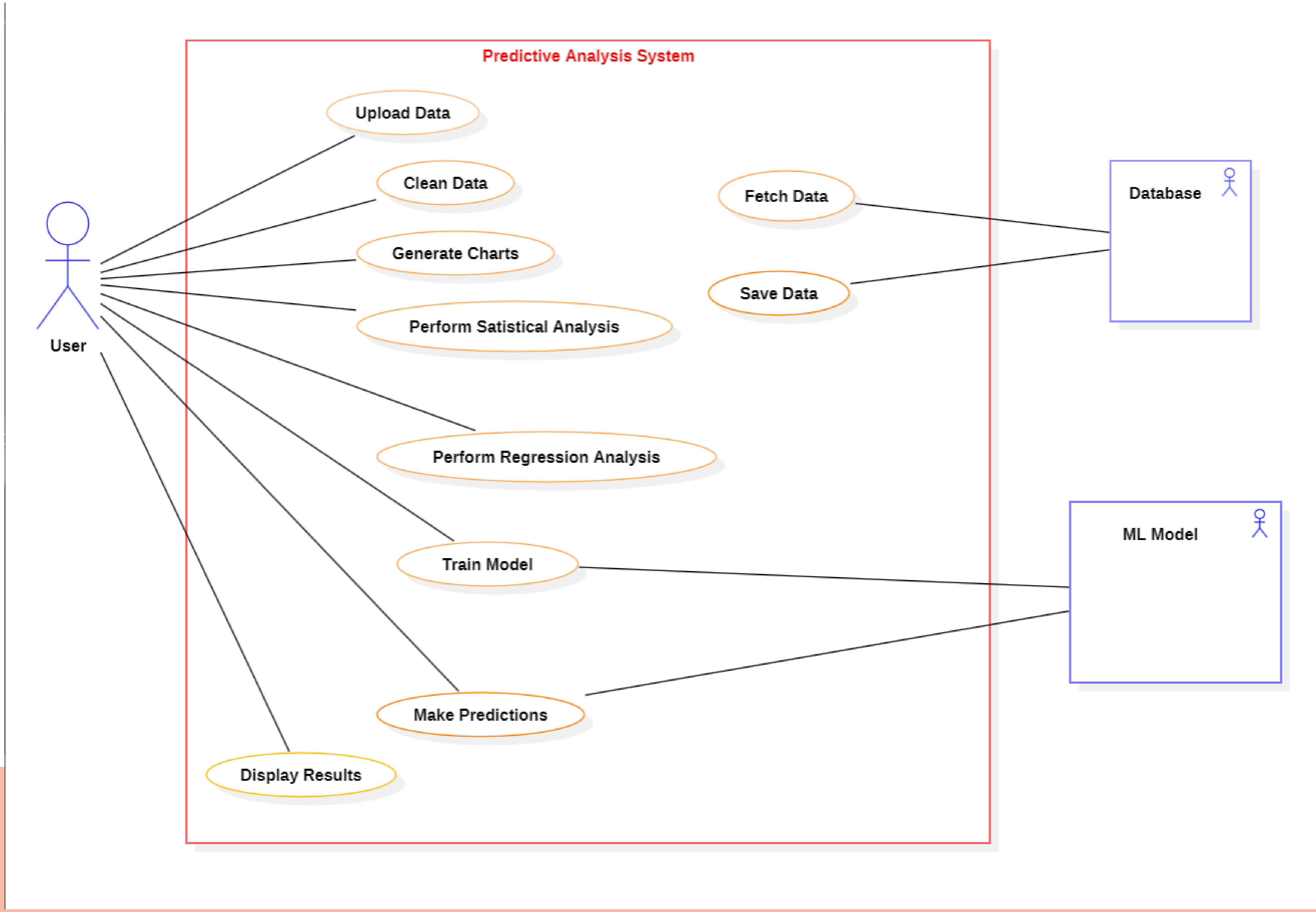
2

NON-FUNCTIONAL REQUIREMENTS:

- High performance with large datasets.
- Scalable and secure database.
- User-friendly UI.

3

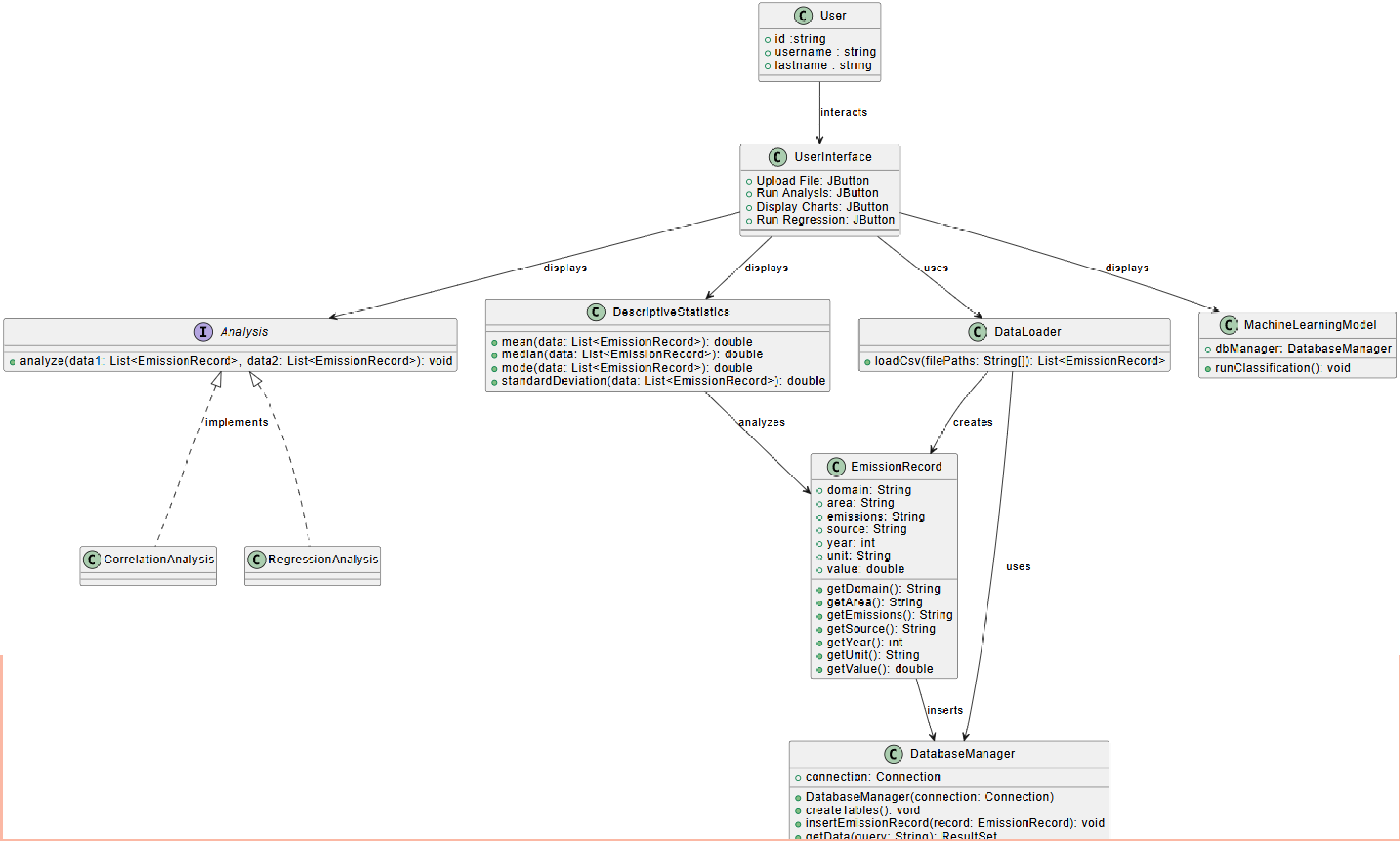
SYSTEM DESIGN - USECASE DIAGRAM



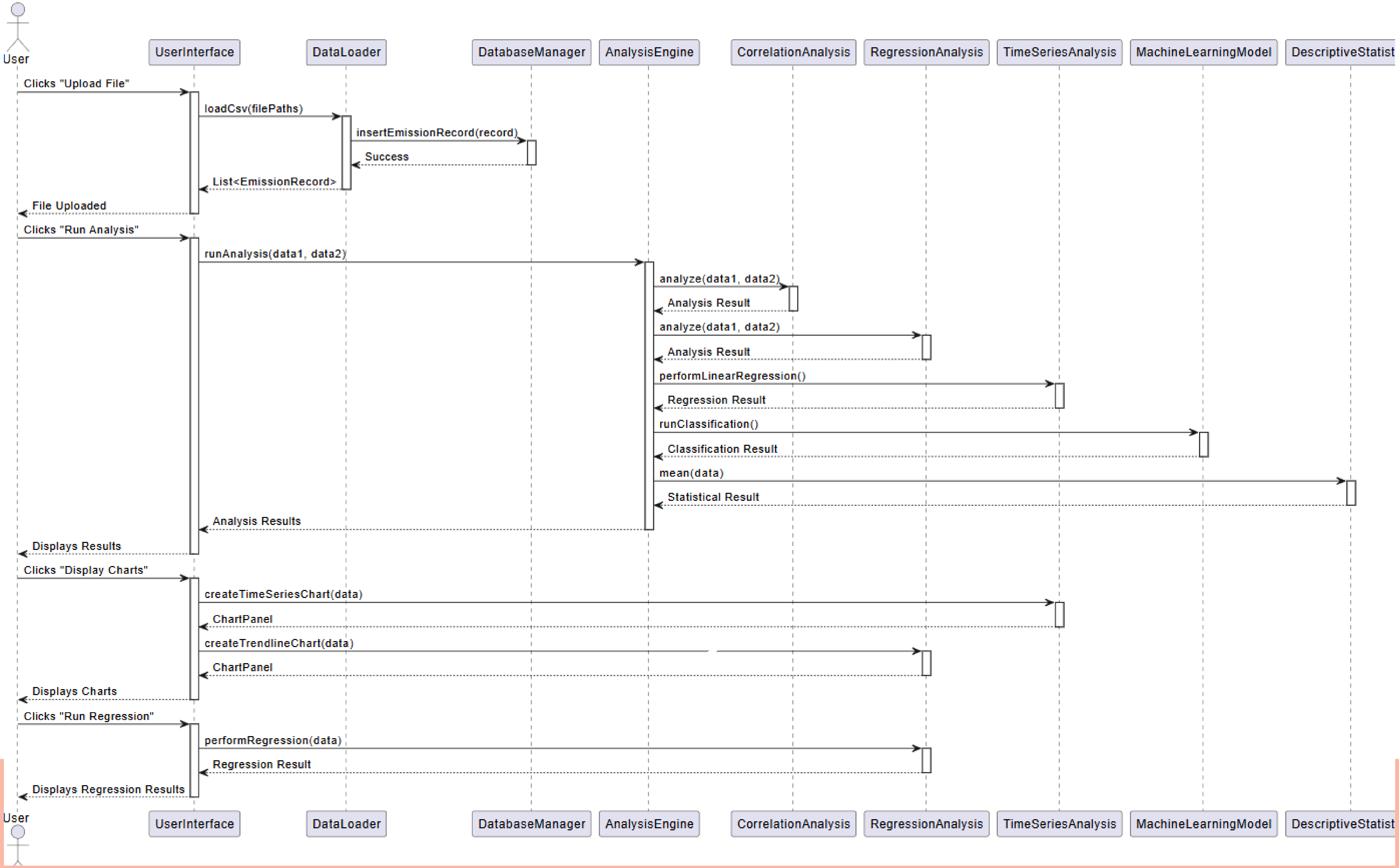
Actors:

- **User:** Uploads, analyzes, and visualizes data.
- **Database:** Stores emissions and analysis data.
- **Machine Learning Model:** Predicts tree nds.

SYSTEM DESIGN - CLASS DIAGRAM



SYSTEM DESIGN - SEQUENCE DIAGRAM



OBJECT-ORIENTED PROGRAMMING (OOP) PRINCIPLES

Abstraction:

- EmissionChart.java is an abstract base class extended by specific charts like CH4Chart.java.

Inheritance:

- Interfaces: Analysis.java defines a blueprint, implemented by CorrelationAnalysis.java.
- CorrelationAnalysis inherits from Analysis to handle specific correlation tasks.

Static Classes:

- Utility calculations (e.g., mean, median) are in DescriptiveStatistics.java.

Encapsulation:

- Private: Fields in DataLoader.java, DatabaseManager.java, and EmissionRecord.java ensure data hiding.
- Protected: Used in CH4Chart.java for limited subclass/package access.

Polymorphism:

- AnalysisEngine.runAnalysis() calls analysis.analyze(data), dynamically invoking the method based on the passed Analysis type (e.g., CorrelationAnalysis).

LIBRARIES USED

Apache Spark:

- used Apache Spark for efficient data processing and analysis. It handles large datasets, performs distributed computing, and supports machine learning tasks like regression or clustering.

Apache Commons:

- A collection of reusable Java libraries that provide common utilities for tasks like file manipulating strings, and collections.

Weka:

- A machine learning library for data mining and analysis, offering tools for classification, clustering, regression, and visualization.

JFreeChart:

- Creates professional-quality charts and graphs, such as bar, pie, and line charts.

IMPLEMENTATION STEPS

1

BACKEND DEVELOPMENT:

- DataLoader: File uploads, cleaning, and validation.
- DatabaseManager: CRUD operations for storage.
- AnalysisEngine: Computes statistics (mean, regression, etc.).
- PredictionModel: Machine learning training and predictions using tools like Tribuo.
- VisualizationEngine: Dynamic graphs via JFreeChart.

2

FRONTEND DEVELOPMENT:

- Build intuitive UI with JPanel and JFrame.
- Seamless integration with backend components.

STEP 1:

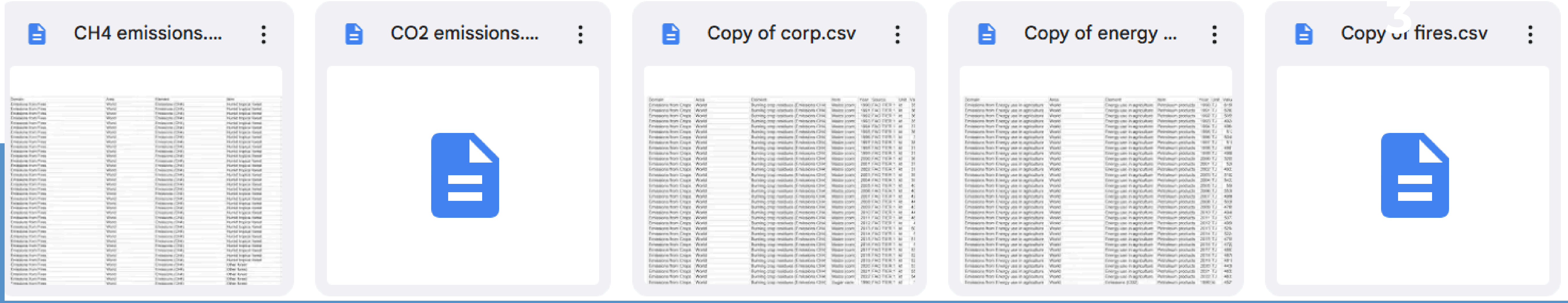
Backend Development

DATA

This project aims to analyze global emissions of CO_2, CH_4 and other related factors using datasets provided by the Food and Agriculture Organization (FAO).

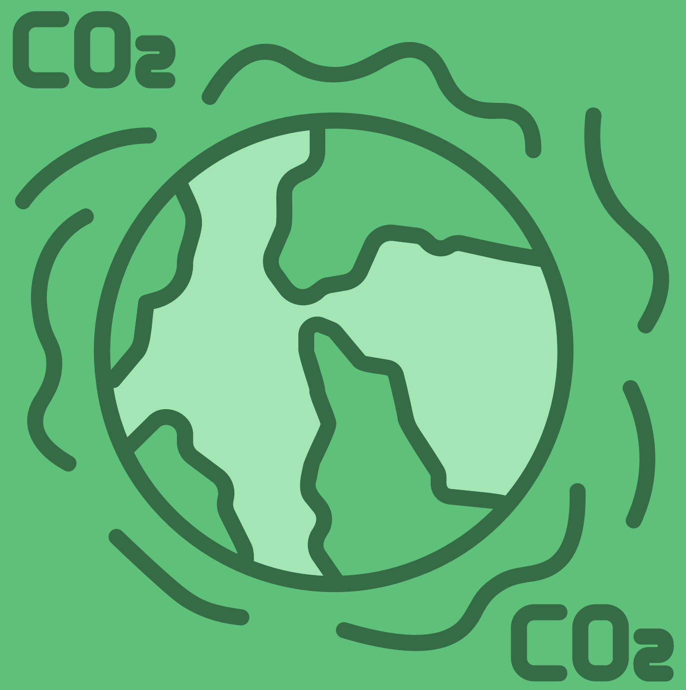
Key Datasets:

- **Forest CO2 Emissions:** Focuses on emissions and absorption trends in forests globally.
- **Livestock Emissions:** Captures methane (CH_4) and from agricultural activities.
- **Fires Emissions:** Tracks emissions from wildfires and their impact on climate.
- **Energy Use:** Highlights energy consumption trends and their contribution to emissions.



MACHINE LEARNING

IMPLEMENTATION



The primary goal was likely to analyze historical emissions data, build predictive models, and classify emissions to better understand emissions patterns and inform strategies for mitigating CO2 and CH4 emissions globally.

This work is directly applicable to environmental policies, international emissions agreements, and strategies for tackling climate change.

```
Spark Version: 3.4.0
root
 |-- Domain: string (nullable = true)
 |-- Area: string (nullable = true)
 |-- Emissions: string (nullable = true)
 |-- Source: string (nullable = true)
 |-- Year: integer (nullable = true)
 |-- Unit: string (nullable = true)
 |-- Value: double (nullable = true)

+--------------------+-----+---------+----------+----+----+------------+
|              Domain| Area|Emissions|    Source|Year|Unit|       Value|
+--------------------+-----+---------+----------+----+----+------------+
|Emissions from Fo...|World|      CO2|Forestland|1990|  kt|-3448167.213|
|Emissions from Fo...|World|      CO2|Forestland|1991|  kt|-3448167.213|
|Emissions from Fo...|World|      CO2|Forestland|1992|  kt|-3448167.213|
|Emissions from Fo...|World|      CO2|Forestland|1993|  kt|-3448167.213|
|Emissions from Fo...|World|      CO2|Forestland|1994|  kt|-3448167.213|
+--------------------+-----+---------+----------+----+----+------------+
only showing top 5 rows

root
 |-- Domain: string (nullable = true)
 |-- Area: string (nullable = true)
 |-- Emissions: string (nullable = true)
 |-- Source: string (nullable = true)
 |-- Year: integer (nullable = true)
 |-- Unit: string (nullable = true)
 |-- Value: double (nullable = true)

+--------------------+-----+---------+--------------------+----+----+---------+
|              Domain| Area|Emissions|              Source|Year|Unit|    Value|
+--------------------+-----+---------+--------------------+----+----+---------+
|Emissions from Fires|World|      CH4|Humid tropical fo...|1990|  kt|3044.7007|
|Emissions from Fires|World|      CH4|Humid tropical fo...|1991|  kt|3044.7007|
|Emissions from Fires|World|      CH4|Humid tropical fo...|1992|  kt|3044.7007|
|Emissions from Fires|World|      CH4|Humid tropical fo...|1993|  kt|3044.7007|
|Emissions from Fires|World|      CH4|Humid tropical fo...|1994|  kt|3044.7007|
+--------------------+-----+---------+--------------------+----+----+---------+
```

FIGURE 1

ANALYZING EMISSIONS DATA USING PYSPARK: SCHEMA AND SAMPLE OUTPUT

- Spark DataFrame schema and a preview of emissions data.
- Key columns include Domain, Area, Emissions, Source, Year, Unit, and Value.
- Sample data for CO_2 emissions from forestlands between 1990-1994.

```
put - Run (App)
|Emissions from Fires|World|      CH4|Humid tropical fo...|1994|  kt|3044.7007|
+--------------------+-----+---------+--------------------+----+----+---------+
only showing top 5 rows

CO2 Data: 3060 records loaded from database.
CH4 Data: 7724 records loaded from database.
CO2 Data - Mean: 104269.20569082572
CO2 Data - Median: 25581.562155
CO2 Data - Mode: -3448167.213
CO2 Data - Standard Deviation: 645471.0438589456
CH4 Data - Mean: 1083.4205807480596
CH4 Data - Median: 32.3632
CH4 Data - Mode: 0.0
CH4 Data - Standard Deviation: 4605.228122040289
Correlation between CO2 and CH4 emissions (World only): -0.029230715414571326
Regression Equation: y = -21.302391795257513x + 44946.10281659706
Testing CO2 Regression Predictions (Including R-squared):
Year: 1990
Actual CO2 Emissions: -3448167.213
Predicted CO2 Emissions: 2554.343144034603
Difference: 3450721.5561440345

Year: 1991
Actual CO2 Emissions: -3448167.213
Predicted CO2 Emissions: 2533.0407522393507
Difference: 3450700.2537522395

Year: 1992
Actual CO2 Emissions: -3448167.213
Predicted CO2 Emissions: 2511.738360444091
Difference: 3450678.951360444

Year: 1993
Actual CO2 Emissions: -3448167.213
Predicted CO2 Emissions: 2490.4359686488315
Difference: 3450657.6489686486

Year: 1994
Actual CO2 Emissions: -3448167.213
Predicted CO2 Emissions: 2469.133576853572
```

FIGURE 2

EMISSIONS DATA ANALYSIS AND REGRESSION MODEL PREDICTIONS: CO2 AND CH4 INSIGHTS

- Statistical analysis of CO2 and CH4 data, including mean, median, mode, and standard deviation.
- Regression analysis for CO2, showing the regression equation and testing predictions with actual vs. predicted emissions.

out - Run (App)

```
Year: 1994
Actual CO2 Emissions: -3448167.213
Predicted CO2 Emissions: 2469.133576853572
Difference: 3450636.3465768537

CO2 Regression R-squared: 0.05646832976470206
Regression Equation: y = -0.2834796197097371x + 598.09633356539
Testing CH4 Regression Predictions (Including R-squared):
Year: 1990
Actual CH4 Emissions: 3044.7007
Predicted CH4 Emissions: 33.97189034301607
Difference: 3010.7288096569837

Year: 1991
Actual CH4 Emissions: 3044.7007
Predicted CH4 Emissions: 33.68841072330633
Difference: 3011.0122892766935

Year: 1992
Actual CH4 Emissions: 3044.7007
Predicted CH4 Emissions: 33.404931103596596
Difference: 3011.2957688964034

Year: 1993
Actual CH4 Emissions: 3044.7007
Predicted CH4 Emissions: 33.12145148388686
Difference: 3011.579248516113

Year: 1994
Actual CH4 Emissions: 3044.7007
Predicted CH4 Emissions: 32.83797186417712
Difference: 3011.8627281358226

CH4 Regression R-squared: -1.357369082767125
Performing Time Series Analysis for CO2...
Linear Regression Model for CO2:
Slope: 1815.0894336631334
Intercept: -3537100.340477104
R-Square: 7.219921400188332E-4
Performing Time Series Analysis for CH4
```

FIGURE 3

REGRESSION AND TIME-SERIES ANALYSIS OF CO_2 AND CH_4 EMISSIONS DATA

- Continuation of regression analysis for CH_4 emissions.
- R-squared values for CO_2 and CH_4 regression models.
- Predicted vs. actual CH_4 emissions for 1990-1994, with differences highlighted.

CODE HIGHLIGHTS

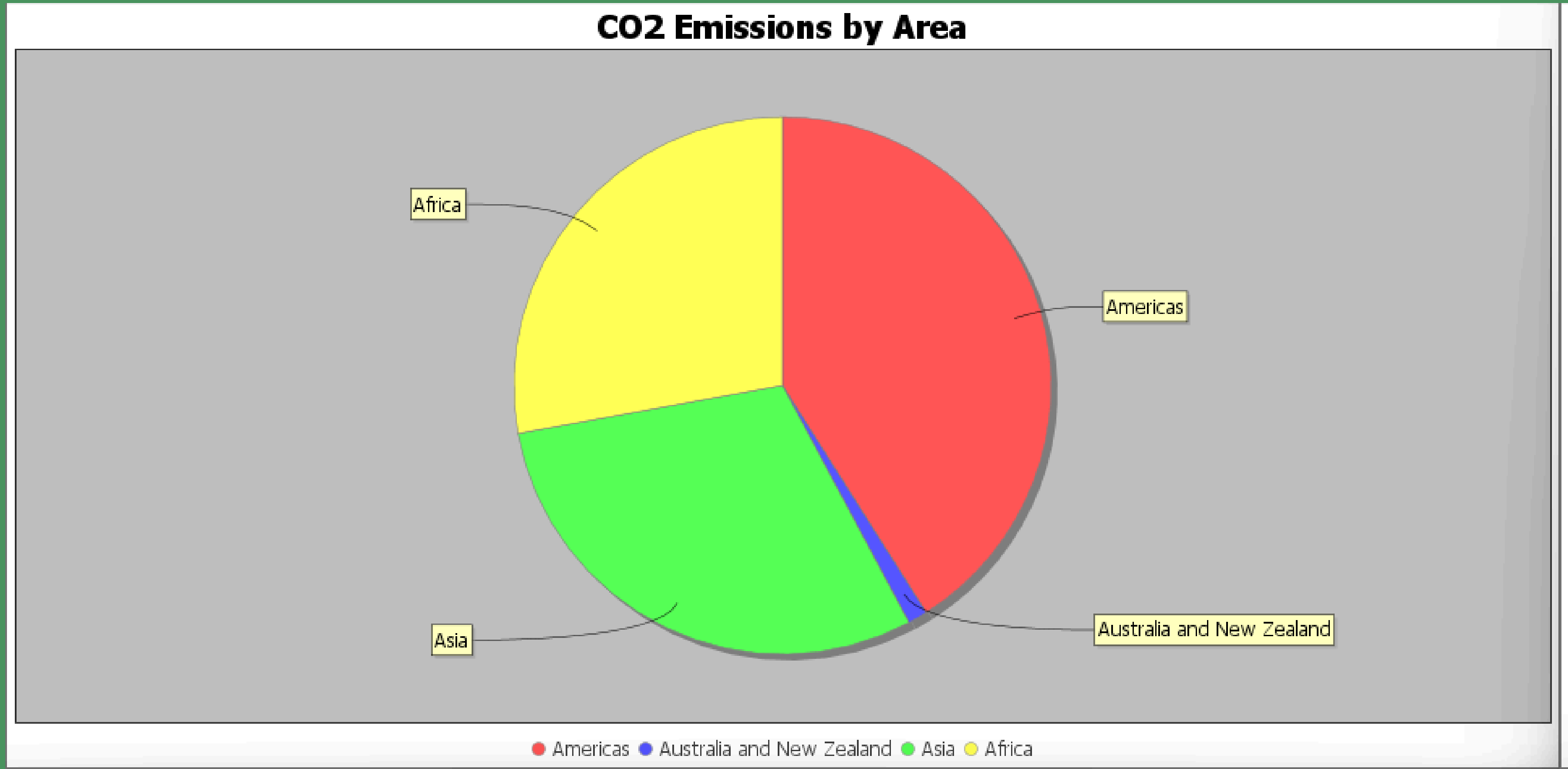
put - Run (App)

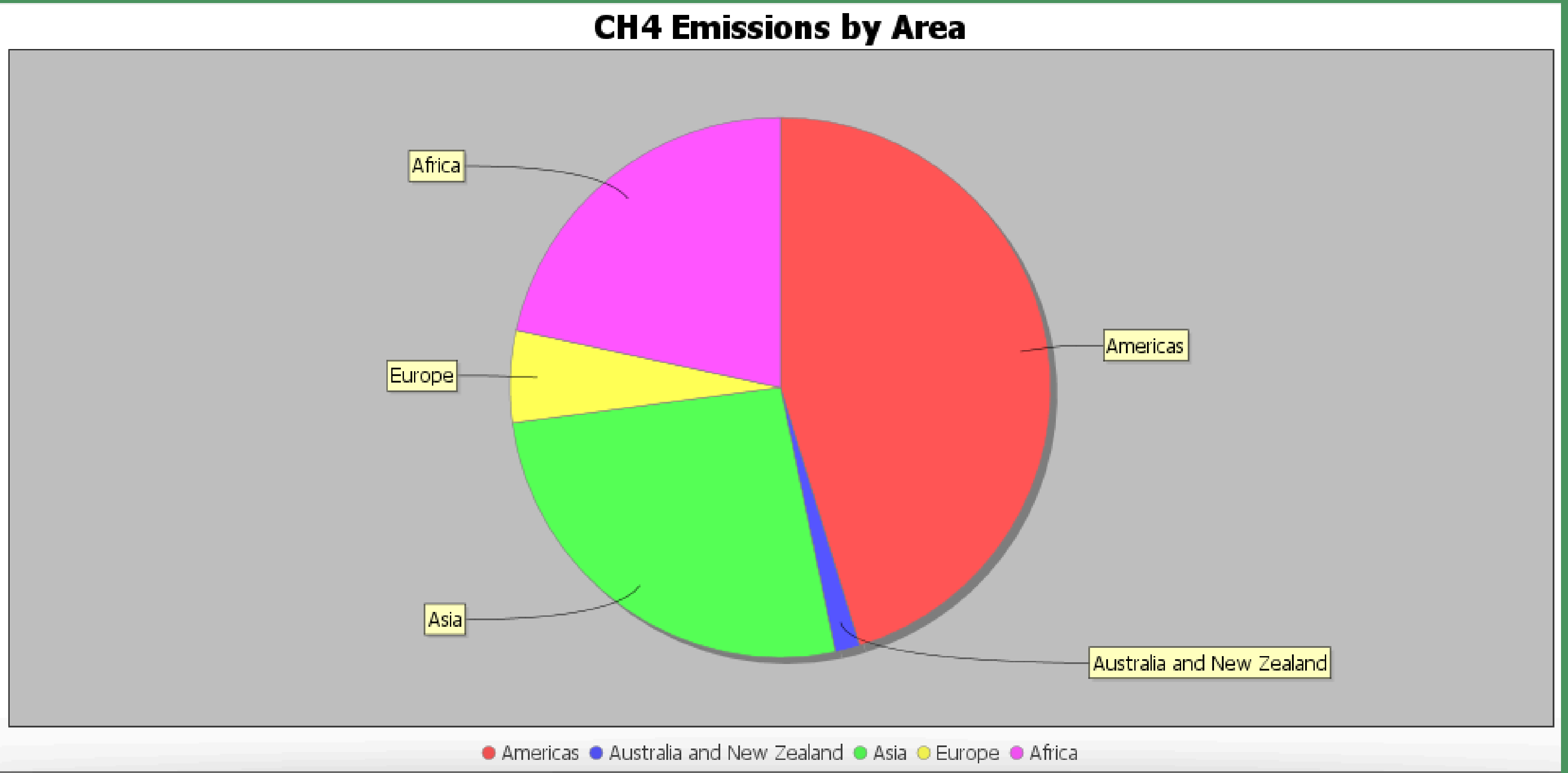
```
Slope: 1815.0894336631334
Intercept: -3537100.340477104
R-Square: 7.219921400188332E-4
Performing Time Series Analysis for CH4...
Linear Regression Model for CH4:
Slope: 6.219470326870274
Intercept: -11393.328075287656
R-Square: 1.6710350244199623E-4
Decision Tree for co2 emissions:
J48 pruned tree
------------------

value <= 736025.8791
|   value <= -690813.5631: Low (3264.0)
|   value > -690813.5631: Moderate (66288.0)
value > 736025.8791: High (3888.0)

Number of Leaves  :     3

Size of the tree :      5

Decision Tree for ch4 emissions:
J48 pruned tree
------------------

value <= 5663.539732: Moderate (177144.0)
value > 5663.539732: High (8232.0)

Number of Leaves  :     2

Size of the tree :      3

[Incubating] Problems report is available at: file:///C:/Users/ahmed/Docum

Deprecated Gradle features were used in this build, making it incompatible

You can use '--warning-mode all' to show the individual deprecation warnin
```

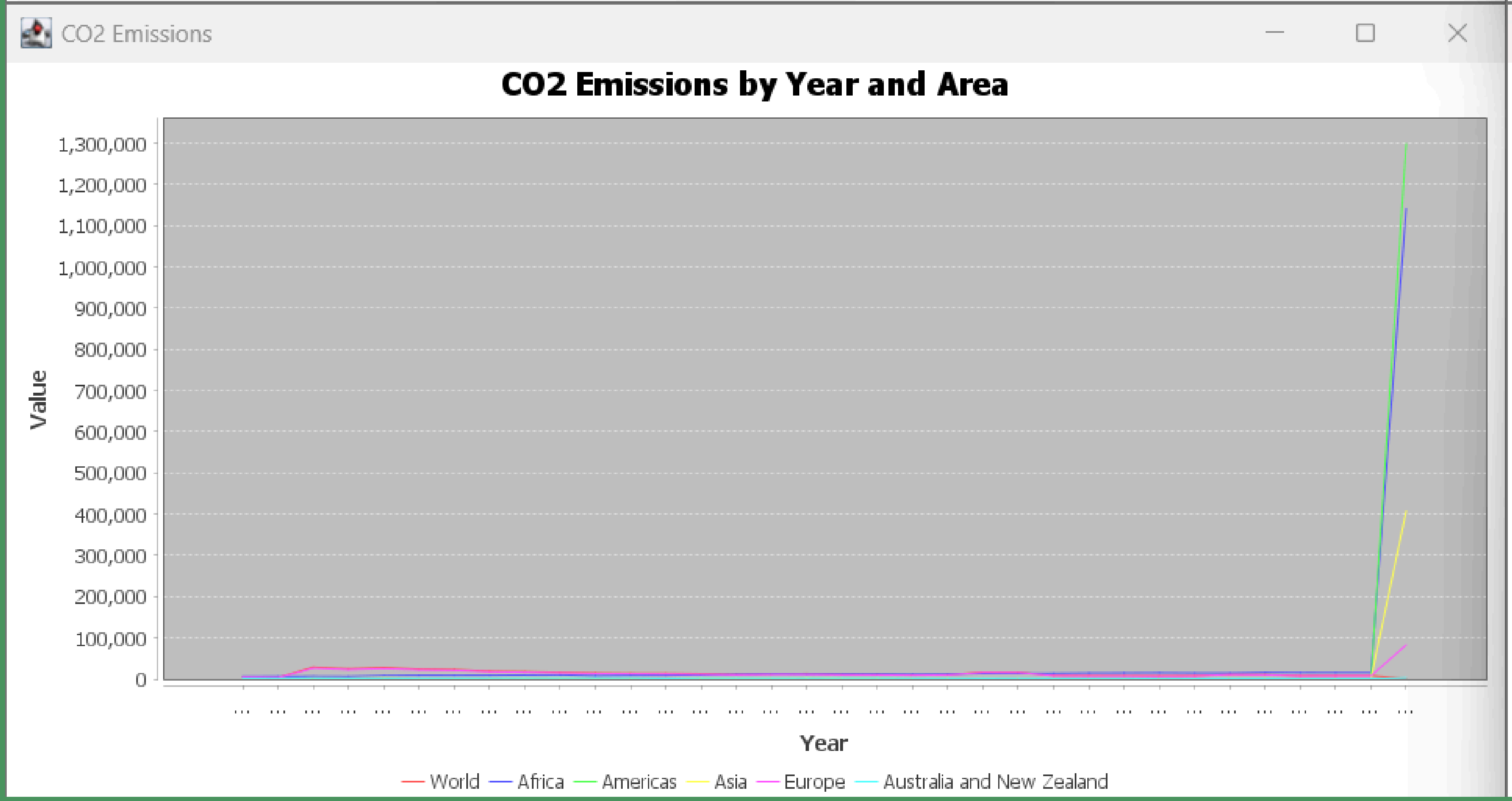
FIGURE 4

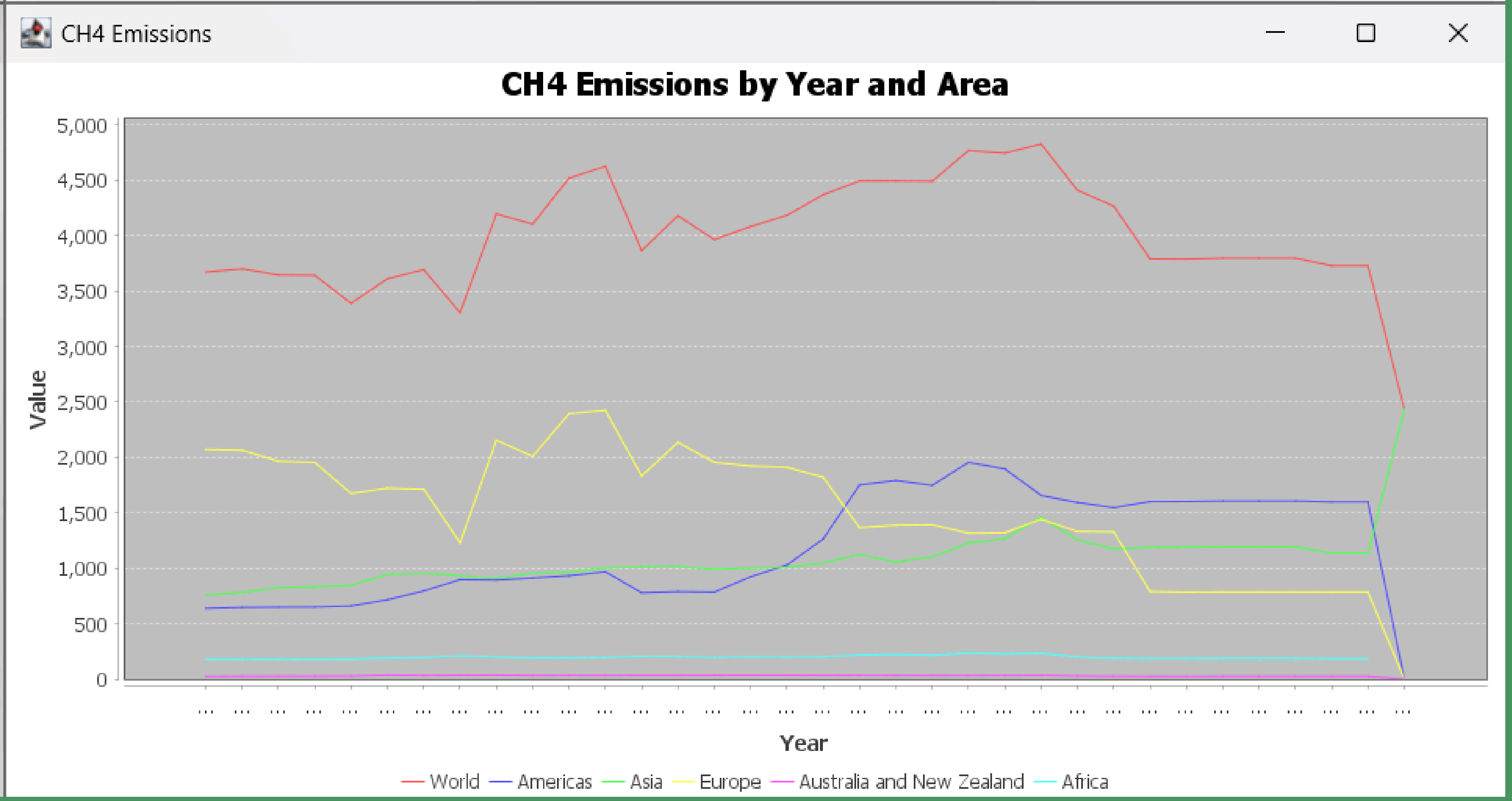
REGRESSION AND DECISION TREE MODELING FOR CO2 AND CH4 EMISSIONS ANALYSIS

- Continuation of regression analysis for CH4 emissions.
- R-squared values for CO2 and CH4 regression models.
- Predicted vs. actual CH4 emissions for 1990-1994, with differences highlighted.



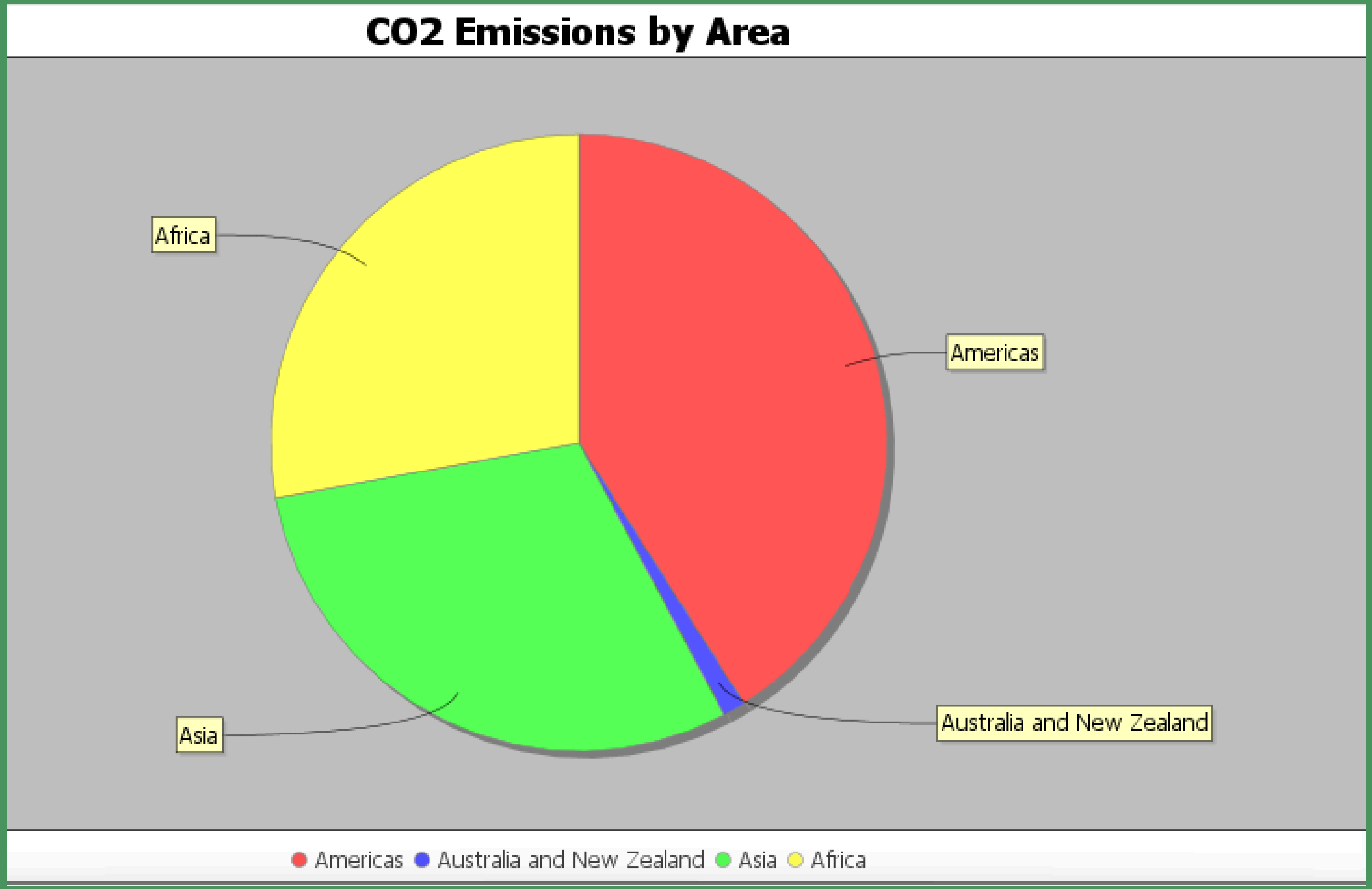






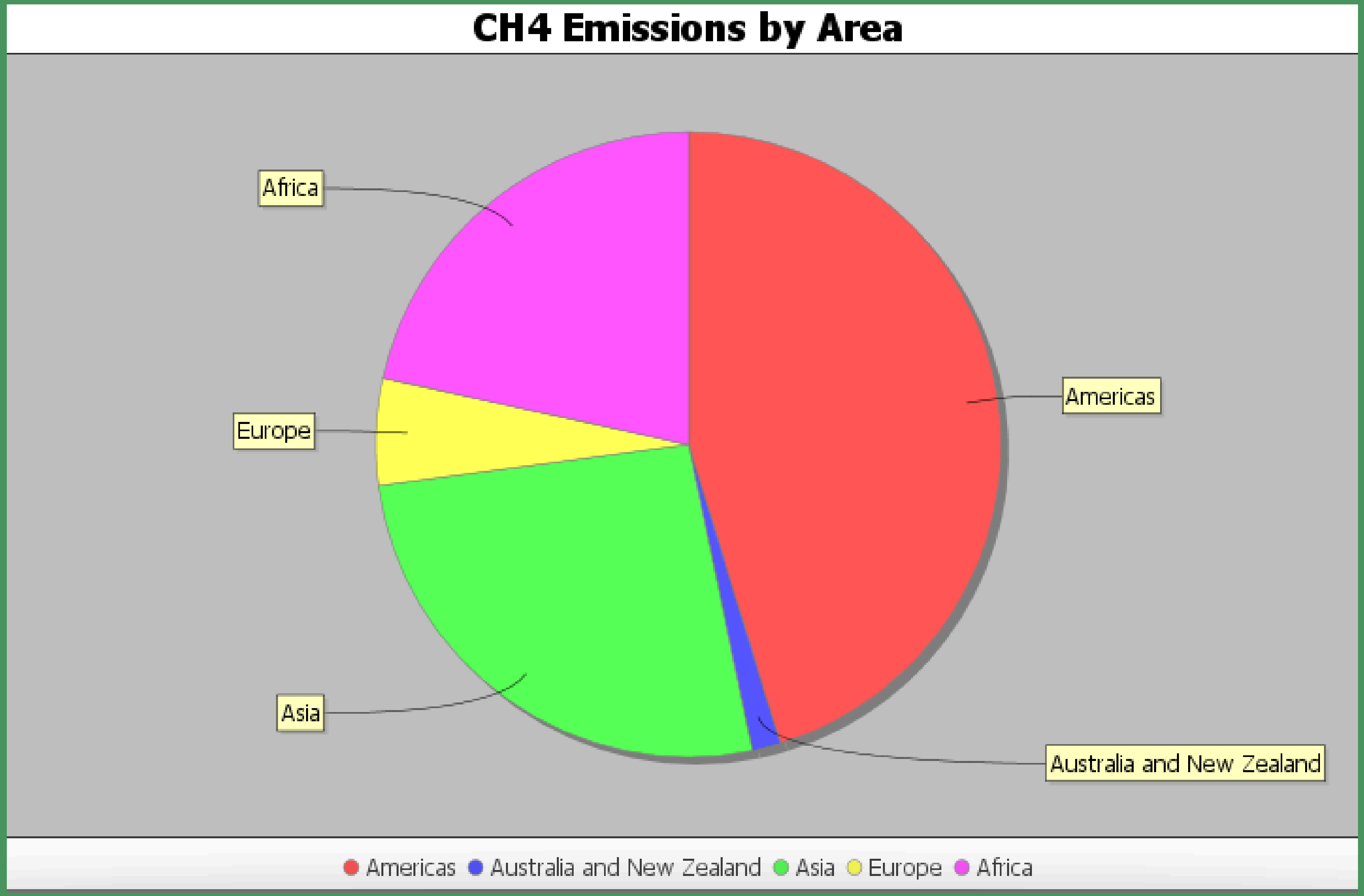
CODE HIGHLIGHTS

FIGURE 5



Displays the proportional contribution of different regions (Americas, Asia, Africa, etc.) to total CO_2 emissions.

FIGURE 6

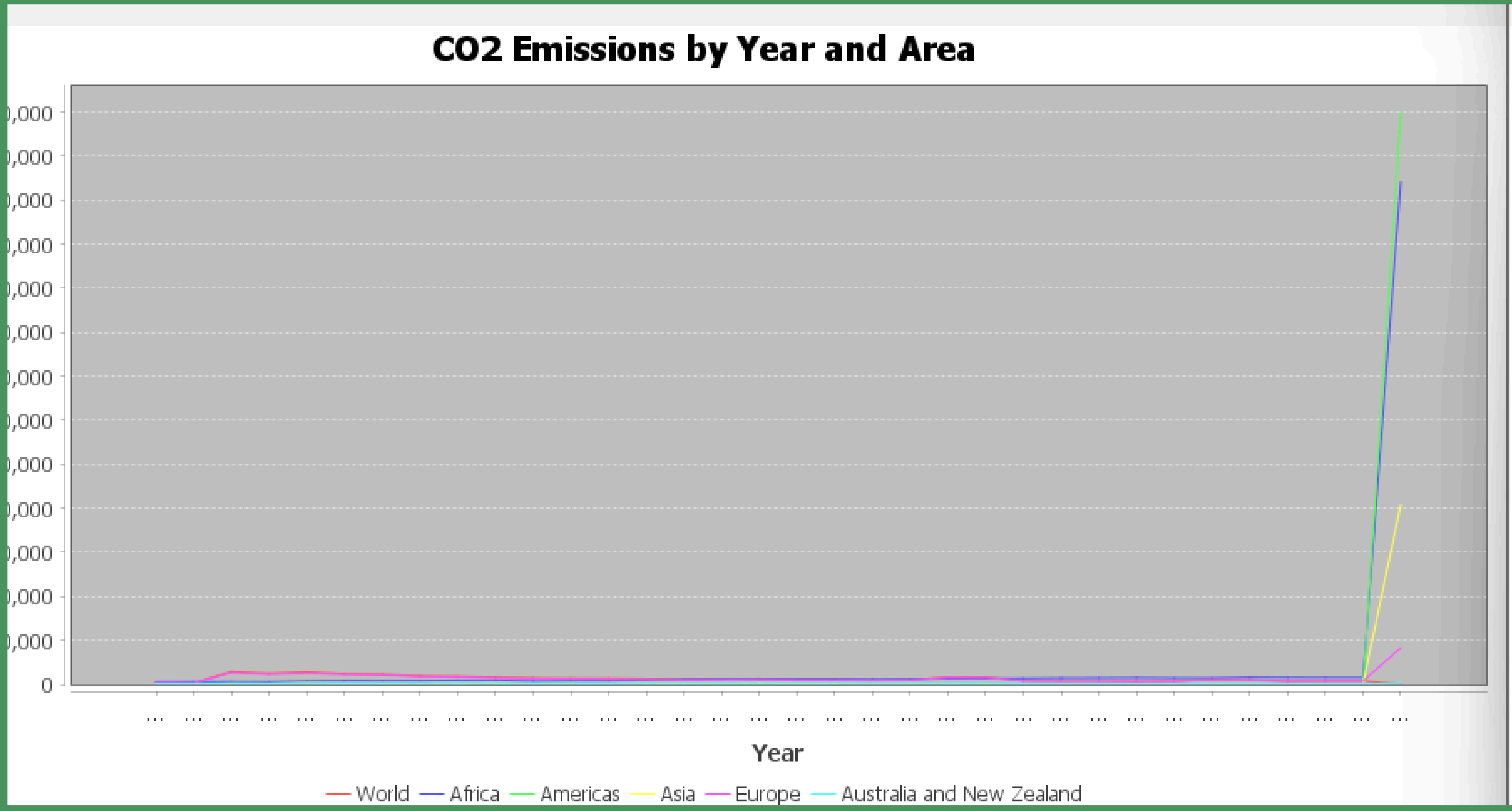


Shows the distribution of CH_4 emissions across regions like Americas, Europe, Asia, Africa, and Australia/New Zealand.

Asia is the largest contributor to CO_2 emissions, followed by the Americas and Africa, indicating that industrial and population density in Asia may significantly impact global CO_2 emissions.

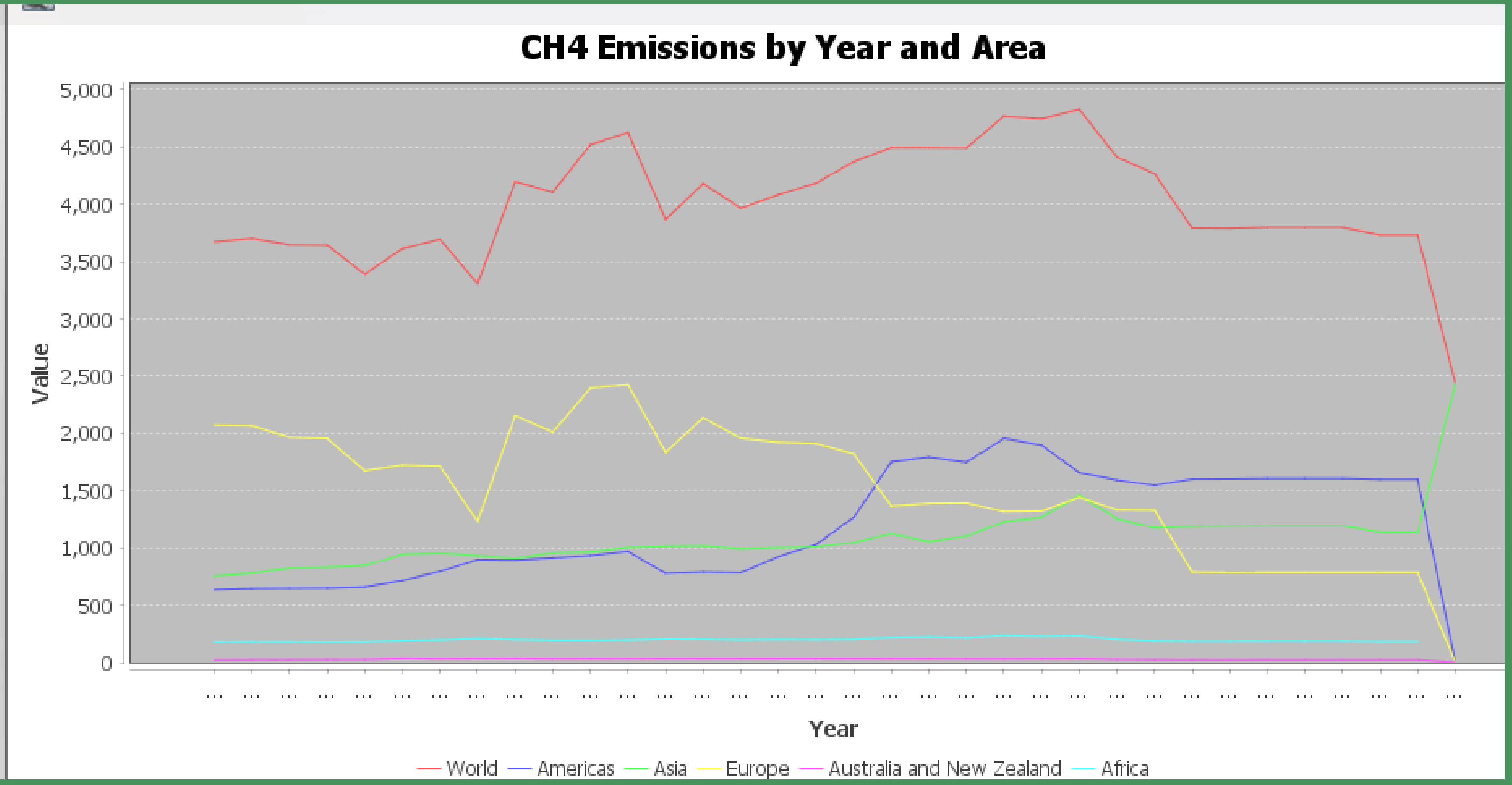
The Americas lead in CH_4 emissions, followed by Asia and Europe. This suggests that agricultural activities (e.g., livestock) and energy production in the Americas play a significant role in methane emissions.

FIGURE 7



Trends in CO2 emissions over time, broken down by different regions and globally.

FIGURE 8



Historical trends of CH4 emissions by region, highlighting variations and patterns across time.

- CO2 emissions show a sharp rise over time, particularly in specific regions like Asia.
- Other regions, such as Africa and Australia/New Zealand, have relatively minimal contributions, though their trends are stable or increasing slightly.

- CH4 emissions remain relatively stable over time compared to CO2 emissions.
- The Americas exhibit the highest CH4 emissions consistently, with a slight decline in recent years.
- Other regions show minor fluctuations, suggesting a less dynamic change in CH4 emissions over time.

COMPARATIVE OBSERVATIONS:

CO_2 emissions are rising more sharply over time, especially in industrialized or densely populated regions.

CH_4 emissions are more stable but concentrated in regions with significant agricultural or natural gas activities.

Efforts to reduce emissions will likely need region-specific strategies, focusing on industrial activities in Asia for CO_2 and agricultural practices in the Americas for CH_4.

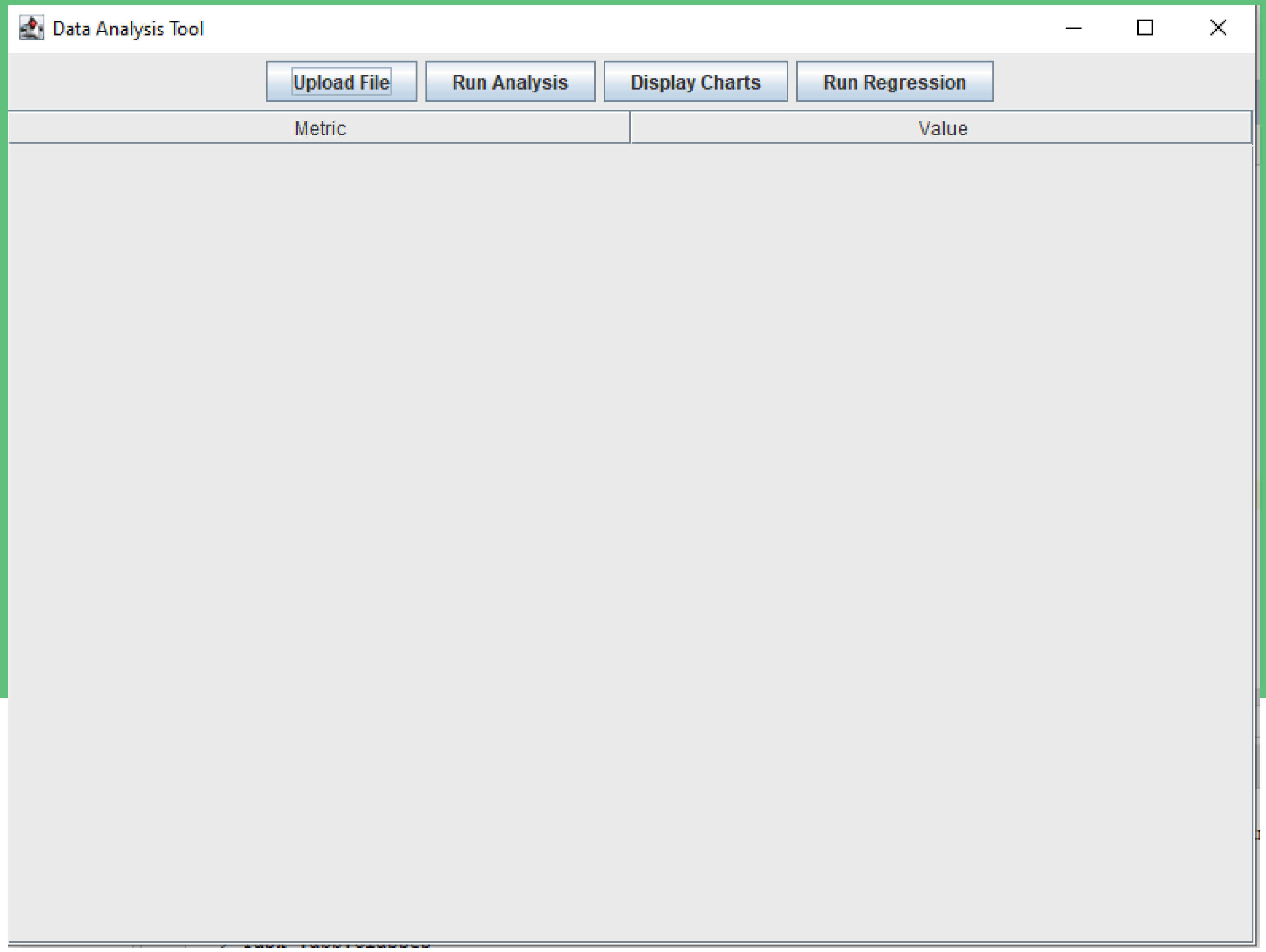
RECOMMENDATIONS

- Asia requires targeted interventions to manage CO_2 emissions, possibly through cleaner energy and improved industrial practices.
- CH_4 reduction efforts should focus on agricultural methane capture and reducing emissions from livestock, particularly in the Americas.
- Global collaboration is essential to mitigate these emissions, as regions contribute differently to CO_2 and CH_4 profiles.

STEP 2:

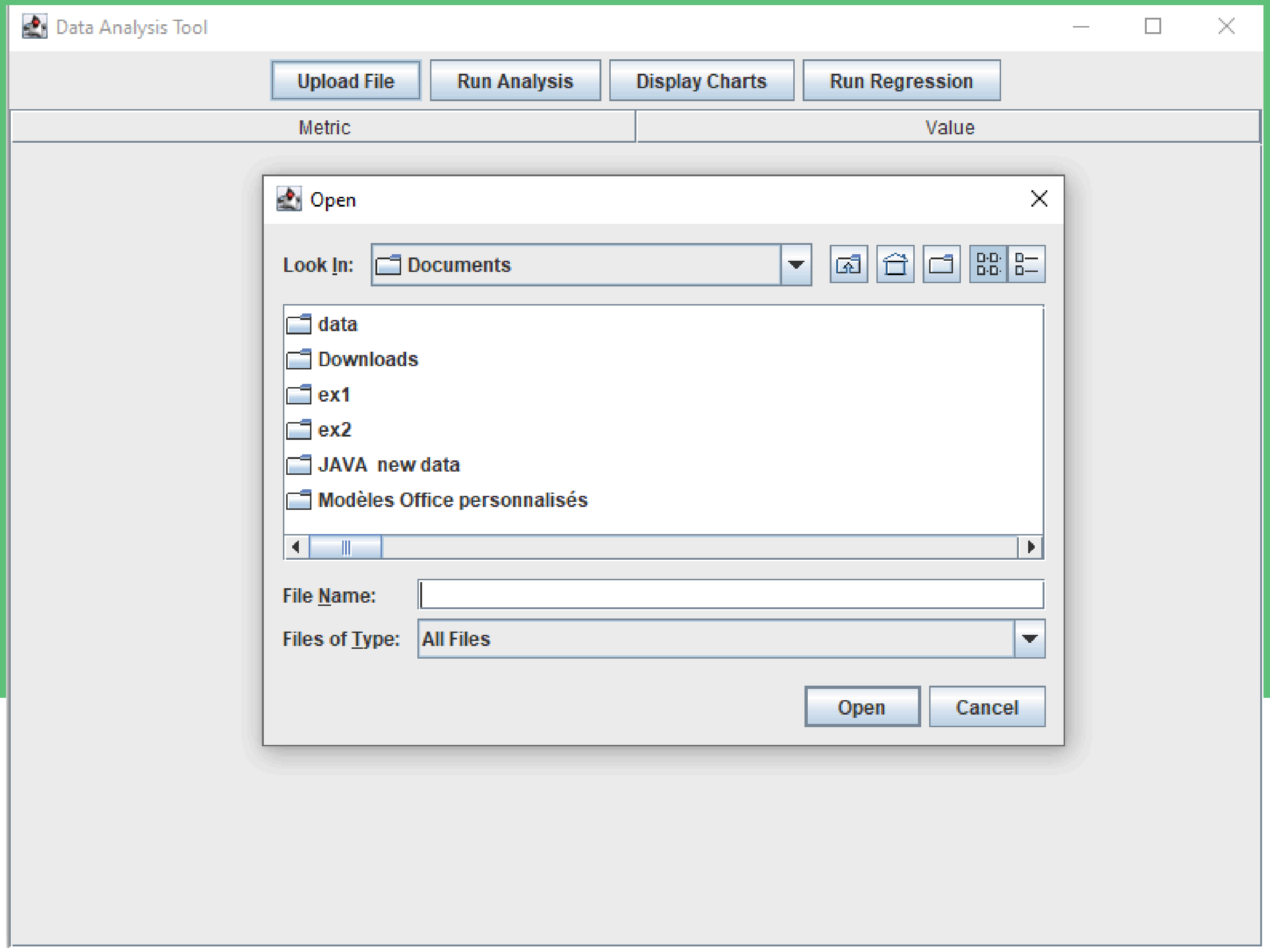
Frontend Development

USER INTERFACE



Step 1: Launch the Tool

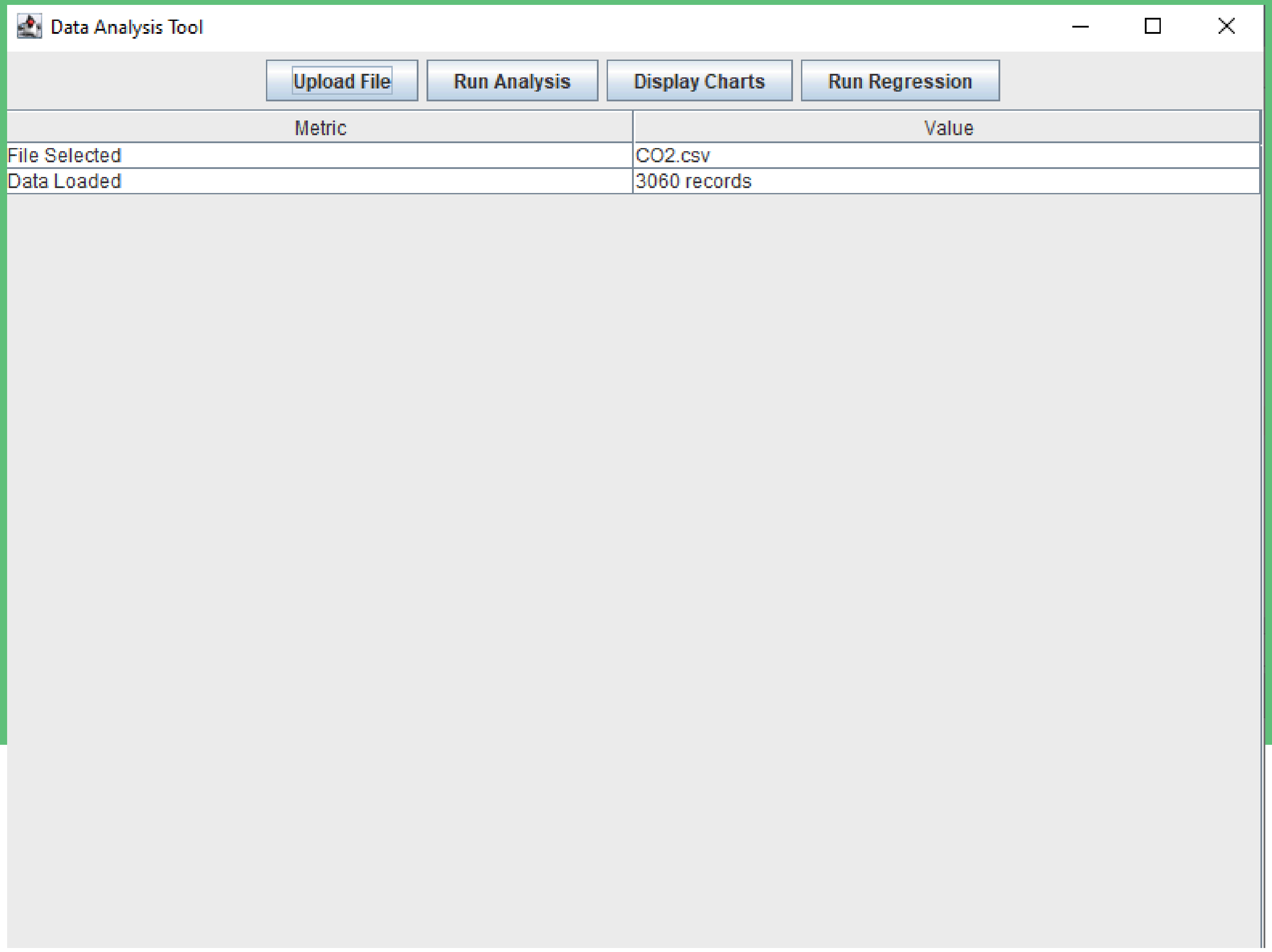
Open the Data Analysis Tool, which provides options to upload files, analyze data, display charts, and run regression models.

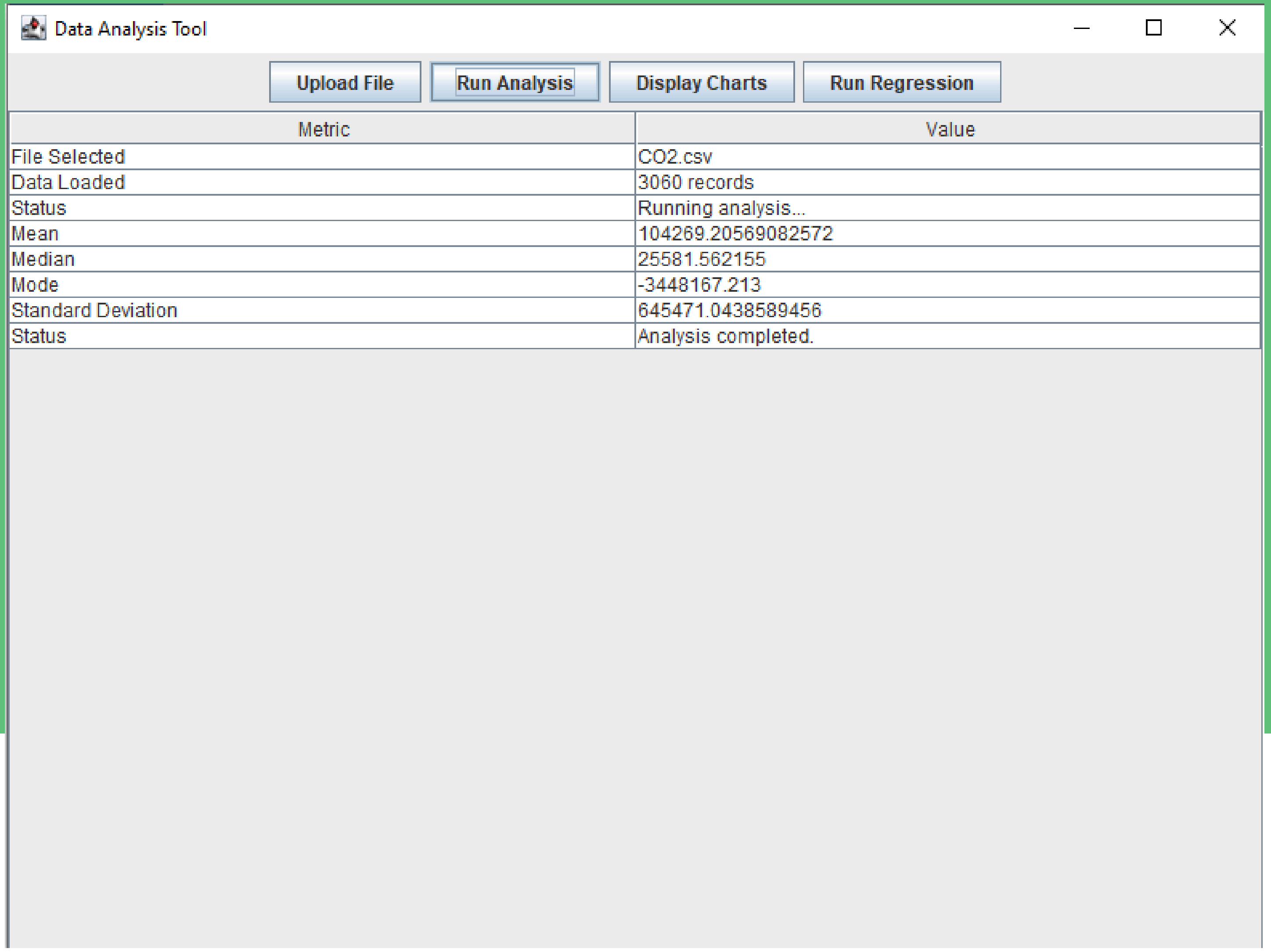


Step 2: Upload File

Click the "Upload File" button to open a file dialog.

Select the desired dataset (e.g., CO2.csv) from your directory.





Step 3: Confirm File Loading

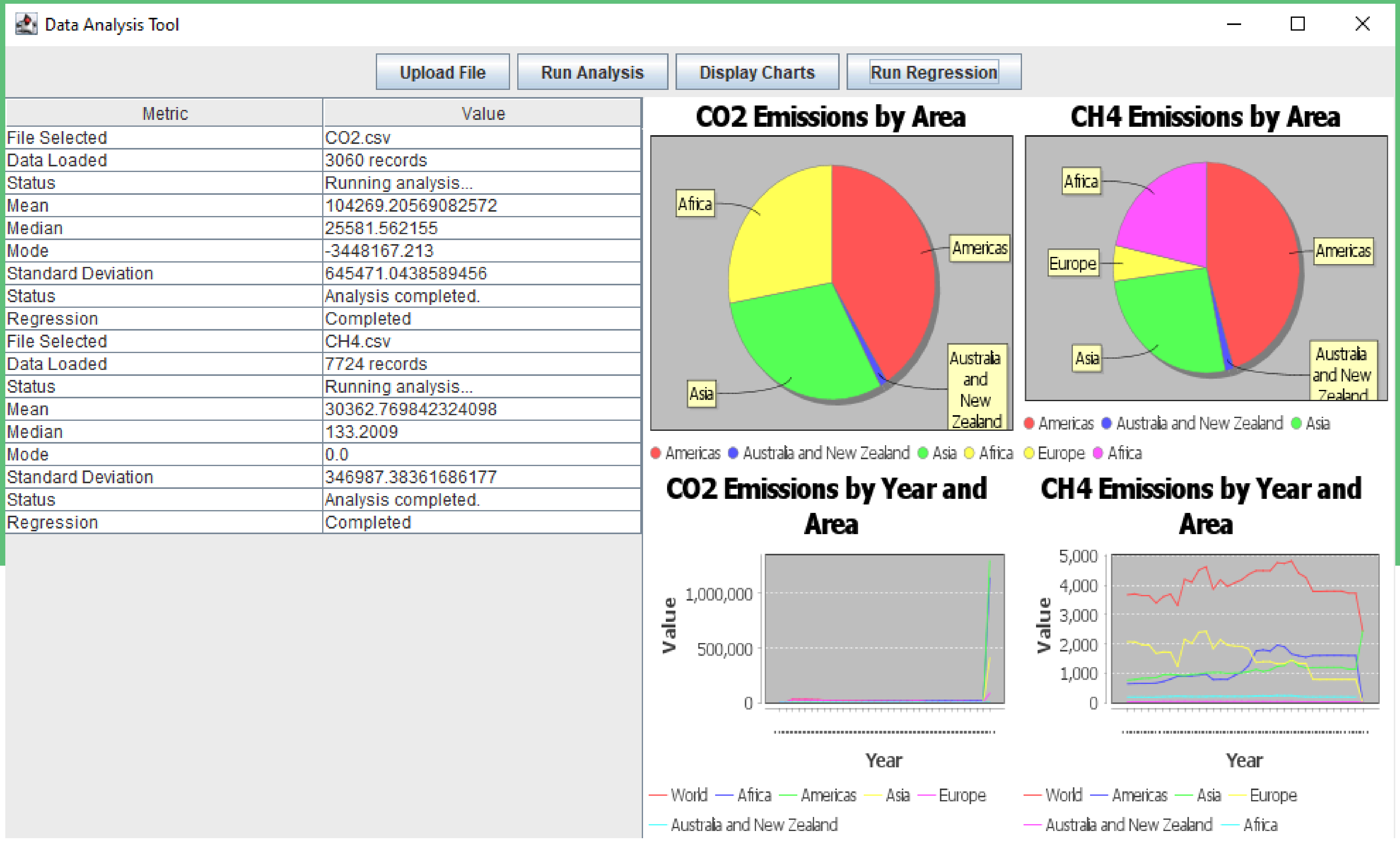
The tool displays the selected file name and confirms the number of records loaded.

Step 4: Run Analysis

Click "Run Analysis" to calculate statistical metrics such as:

Mean, Median, Mode, and Standard Deviation.

The results are displayed in the interface.



Step 4: Run Analysis

Click "Run Analysis" to calculate statistical metrics such as:
Mean, Median, Mode, and Standard Deviation.
The results are displayed in the interface.

DEMO

LET'S SEE HOW IT WORKS

THANK YOU!

WELCOMING YOUR QUESTIONS