PREDICTING CO2 EMISSIONS FROM THE AGRI-FOOD SECTOR: ANALYZING FAO AND IPCC DATA FOR SUSTAINABLE STRATEGIES

PROFILE PRESENTATION

Team: oa-5

Meet Our Team!

LUNGILE BALOYI

AMOS MAPONYA

KHADIJAH KHAN

OBET SEGWETE MABOWA

MIEKE SPAANS

Introduction

The agri-food sector is responsible for 20-25% of total CO_2 emissions.

Despite our best efforts, there is a gap in effective strategies to combat CO_2 emissions.

Urgent action is required to understand and implement effective strategies to significantly cut CO_2 emissions and mitigate the severe effects of climate change.

Problem Statement

We are a team of data scientists tasked to train machine learning models to predict future CO2 emissions in the agricultural sector.

We analyzed the dataset supplied by the FAO and the IGPCC.

Our aim with this presentation is:

- to communicate our process
- evaluate our results
- foster trust between us and the stakeholders of this project.

Our process

Exploratory data analysis

The first step in fully understanding our data.
We looked at the relationships between variables and $CO2$ emissions.
This process also involves visualizing our data through the use of graphs.

Model building: regression

We trained three different models:

- LASSO
- Linear regression
- Decision tree

The aim is to diversify and increase the chances for a successful model.

Model evaluation

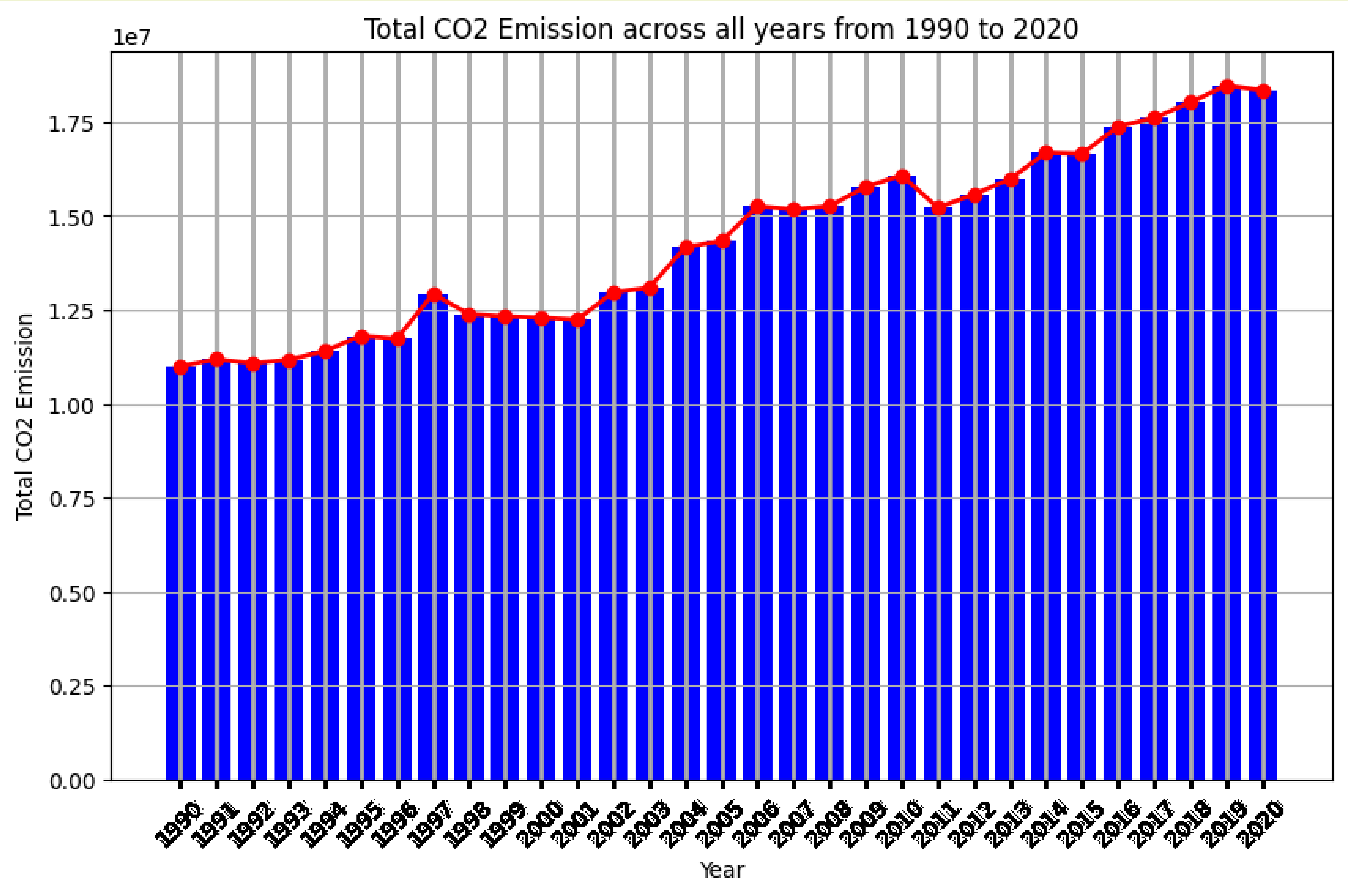
A model is only useful if it is:

- consistent
- accurate

Here we assess the three models we trained.

Exploratory Data Analysis:

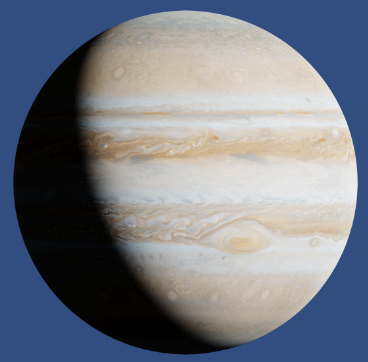
Analyzing the global change in CO2 emissions


Total CO2 Emission across all years from 1990 to 2020
1e7
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Total CO2 Emission
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year

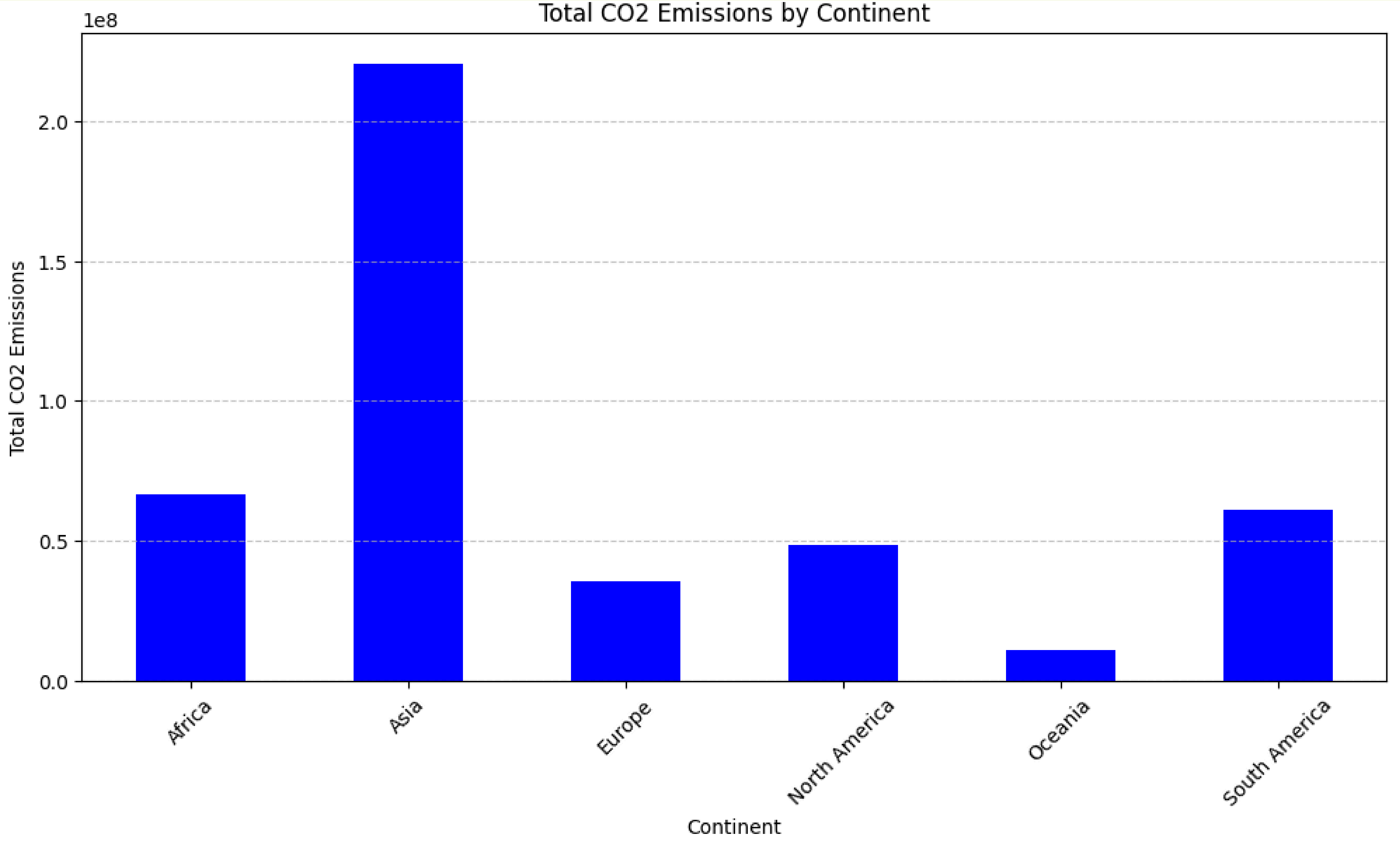
The planets: Diversity in the solar system

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

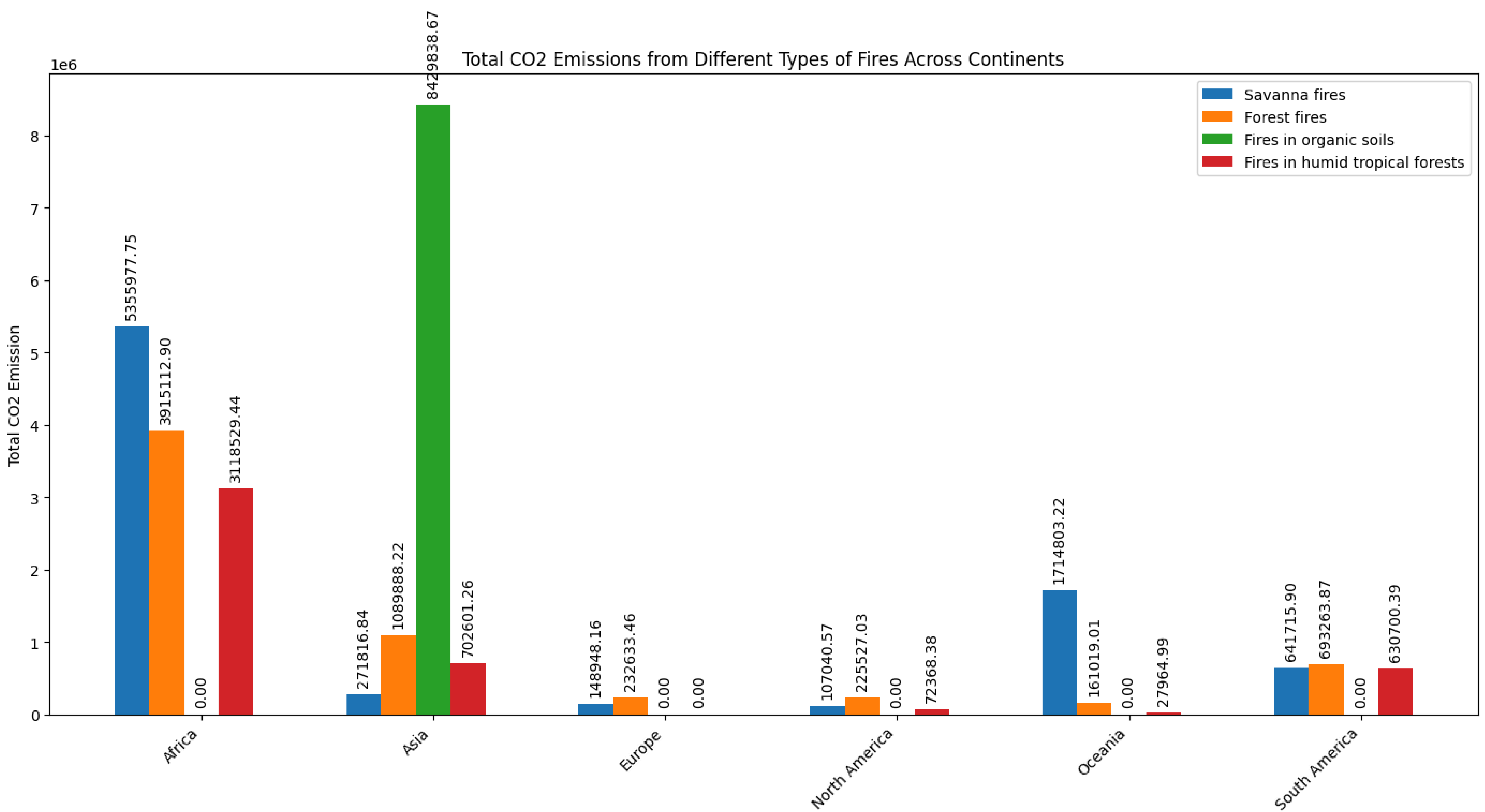
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.



Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

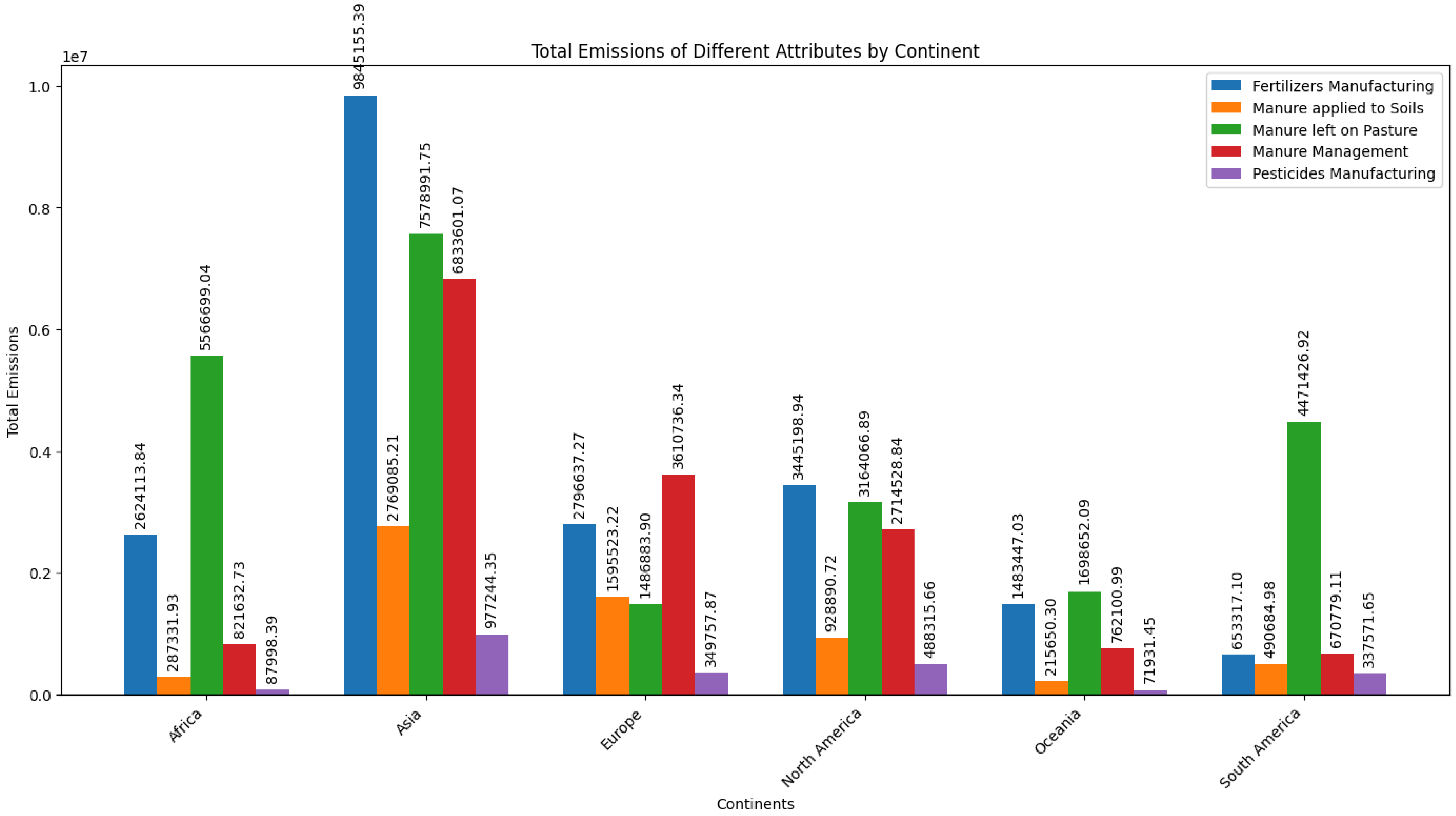

Total CO2 Emissions by Continent
1e8
Total CO2 Emissions
0.0
0.5
1.0
1.5
2.0
Africa
Asia
Europe
North America
Oceania
South America
Continent

Analyzing CO_2 emissions of fires in each continent



Total CO2 Emissions from Different Types of Fires Across Continents
1e6
Total CO2 Emission
0
1
2
3
4
5
6
7
8
Savanna fires
Forest fires
Fires in organic soils
Fires in humid tropical forests
5355977.75
3915112.90
0.00
3118529.44
271816.84
1089888.22
8429838.67
702601.26
148948.16
232633.46
0.00
0.00
107040.57
225527.03
0.00
72368.38
1714803.22
161019.01
0.00
27964.99
641715.90
693263.87
0.00
630700.39
Africa
Asia
Europe
North America
Oceania
South America

Analyzing CO_2 emissions of manure in each continent



Total Emissions of Different Attributes by Continent
1e7
1.0
0.8
0.6
0.4
0.2
0.0
Total Emissions
Continents
Africa
Asia
Europe
North America
Oceania
South America
Fertilizers Manufacturing
Manure applied to Soils
Manure left on Pasture
Manure Management
Pesticides Manufacturing
2624113.84
287331.93
5566699.04
821632.73
87998.39
9845155.39
2769085.21
7578991.75
6833601.07
977244.35
2796637.27
1595523.22
1486883.90
3610736.34
349757.87
3445198.94
928890.72
3164066.89
2714528.84
488315.66
1483447.03
215650.30
1698652.09
762100.99
71931.45
653317.10
490684.98
4471426.92
670779.11
337571.65

Variables used to train our models

We trained our model through variable selection.

Final variables:

- **Farming:** Crop Residues, Rice Cultivation, On-farm energy use
- **Manure and fertilizers:** Fertilizers Manufacturing, Manure applied to Soils, Manure left on Pasture, Manure Management
- **Food specific:** Food Household Consumption, Food Packaging, Food Processing, Agrifood Systems Waste Disposal, IPPU (industrial energy processes)
- **Population data:** Urban population, Total Population - Male, Total Population - Female

Defining our model metrics

MSE:

- Average of squared residuals
- A residual is the distance between the correct value and the model's incorrect prediction
- Lower value = more accurate prediction

R-squared:

- Provides information about how well our model's regression line fits the actual data
- Presented as a decimal value
- Higher value = more accurate fit

Process to choose our models

	model	best_score	best_params
0	linear_regression	0.901914	{'normalize': False}
1	lasso	0.895835	{'alpha': 1, 'selection': 'cyclic'}
2	decision_tree	0.974016	{'criterion': 'mse', 'splitter': 'random'}

Model 1: LASSO regression

Overview

- Definition:

With LASSO, some of our input variables have been scaled down to zero, so that only the variables with the greatest influence remain.

- Initial MSE: 5.632 billion
- Final MSE: 4,995 billion
- Final R-squared: 0.9242
- From this we gather that:

Overall, the LASSO model demonstrates strong performance, with high R^2 values indicating good explanatory power.

Model 2: Linear regression

Overview

- Definition:

This model is a simple linear regression model. It is the most basic model to train. None of the input variables are shrunk down.

- Initial MSE: 5.303 billion
- Final MSE: 4,732 billion
- R-squared: 0.9282
- From this we gather that:

Overall, our linear regression model performed slightly better than our LASSO model. We noted a decrease in MSE and slight increase in R2.

Model 3: Decision tree

Overview

- Definition:

Decision tree learning employs a divide and conquer strategy to identify the optimal way to split our data within a tree.

- Initial: 1.825 billion
- Final MSE: 1,015 billion
- R-squared: 0.9846
- From this we gather that:

Overall, the Decision Tree model demonstrates exceptional performance, with the highest R2 values and lowest MSE of all three models.

Final model metrics

Model	Test MSE	Test R-squared	Final MSE	Final R-squared
Linear Regression	5.30737e+09	0.896812	4.73225e+09	0.928172
LASSO	5.63209e+09	0.890499	4.9953e+09	0.924179
Decision Tree	1.82518e+09	0.964514	1.01596e+09	0.984579

Model Conclusion

By evaluating these different factors:

Performance

Generalization

Complexity

Interpretability

Overall: The decision tree is the best choice.

FUTURE GOALS

Future Outlook

Future of agriculture:

- Agriculture accounts for 20-25% of global emissions.
- Agri-food sector is vulnerable to climate change impacts.

This study aims to address these gaps and provide a holistic analysis and actionable strategies for stakeholders.

Future studies from our model's variables:

- Responsible manure management
- Improved food processing
- Sustainable food and waste disposal

Conclusion

Our team of data scientists successfully:

- cleaned and analyzed our data
- trained machine learning models
- evaluated each model's performance

This project aimed to communicate our process, evaluate results, and foster trust with stakeholders.

Here are three key points:

- Collaboration and knowledge-sharing are crucial for effective implementation
- CO_2 emissions are increasing each year, with Asia contributing the most
- We were able to train an accurate regression model

References And Insights

Please find below, important links:

- Project Repository: Click here
- Facilitator Github Usernames: Click here
- Managing Environments: Click here
- Creating environments from requirements.txt using "conda create": Click here
- Jupyter notebook markdown cheatsheet: Click here
- Trello: Click here
- Video on how to set-up your Trello board: click here

Business Motivation

Agriculture is a testament to the power of patience and perseverance in business. Every harvest is a result of months of hard work and dedication.

Thank you for your time.
Any questions?