REGIS UNIVERSITY

DATA SCIENCE PRACTICUM II

GREENHOUSE GAS EMISSIONS

Future Footprints: Predicting Trends in Greenhouse Gas Emissions

Presented By Peter La

Table of Contents

Table of Contents

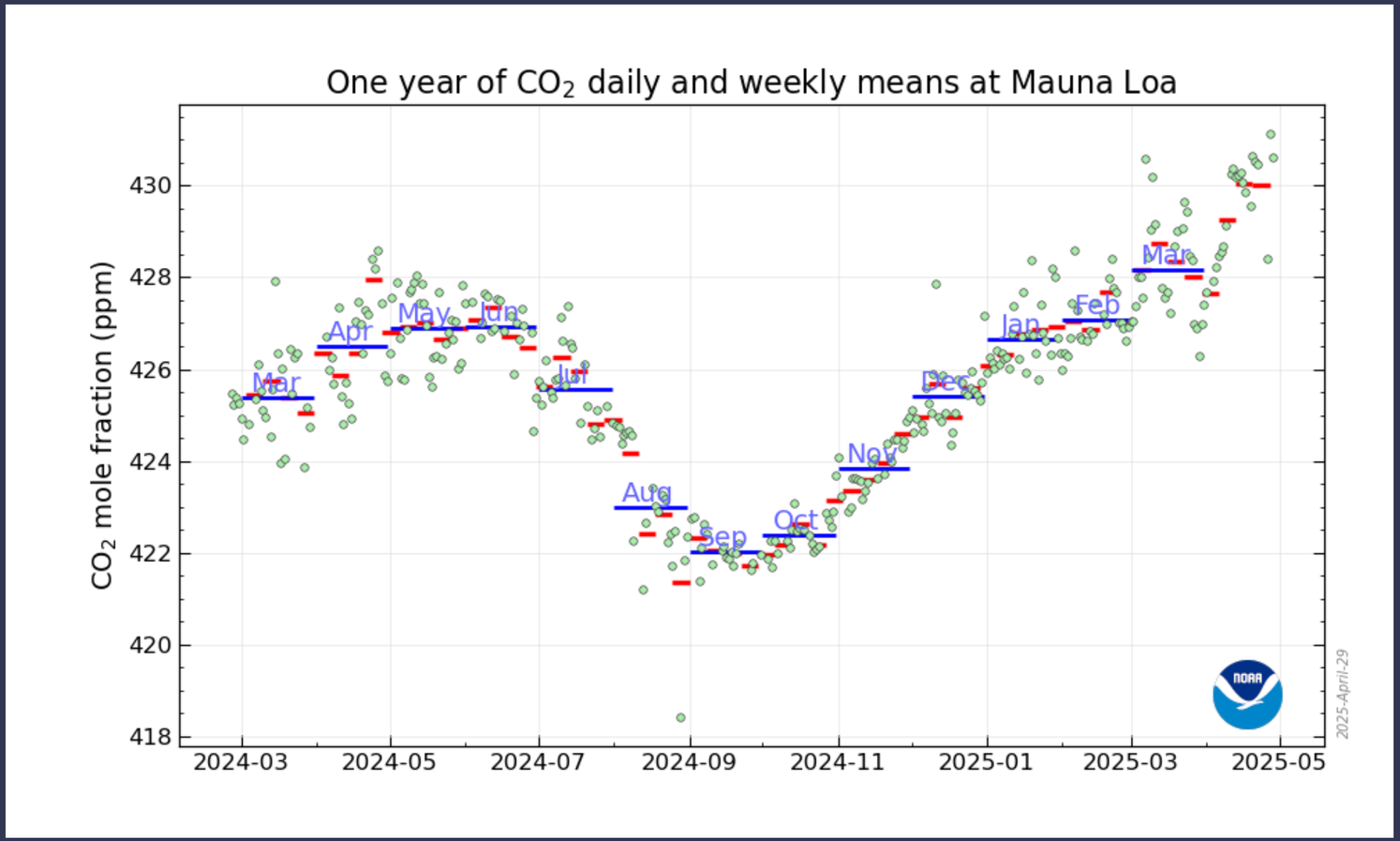
Background

Greenhouse gases (GHGs) play a critical role in Earth's climate system. However, human activities have significantly increased their concentrations in the atmosphere.

The NOAA Global Monitoring Laboratory reported that atmospheric CO_2 levels reached approximately 430 parts per million (ppm) in early 2025, a significant rise from pre-industrial levels of around 280 ppm. Similarly, the levels of other GHGs have increased sharply.

Addressing the increase in greenhouse gases is thus critically important for environmental sustainability, economic stability, and human health.

Trends in Atmospheric Carbon Dioxide (CO2)



Atmospheric increase of CO2 in weekly averages of CO2 observed at Mauna Loa, Hawaii.

Preliminary weekly (red line), monthly (blue line), and daily (green points) averages at Mauna Loa for the last year.

Project Overview

Using an unsupervised learning approach, we can explore and analyze the dataset on greenhouse gas emissions to reveal hidden patterns or insights without predefined labels.

- **Clustering Analysis**
  Group countries or sectors with similar emission characteristics.
- **Dimensionality Reduction**
  Reduce complexity and identify the variables that contribute most strongly to emissions.

Project Timeline

Week	Key Activities	Deliverables/Goals
1	Project Setup & Planning	Project goals, dataset acquisition
2	Exploratory Data Analysis	Initial insights, distributions, correlations
3	Clustering Analysis	Identify groups of countries with similar patterns
4	Dimensionality Reduction (PCA)	Reduced dimensionality, PCA visualizations
5	Anomaly Detection	Identify countries or sectors with unusual emission profiles
6	Models Comparison	Evaluating multiple predictive models
7	Visualization	Visualize emission trends across different countries
8	Reporting & Presentation	Document and present summary of findings

Dataset Overview

Basic Information

- Rows: 44,079
- Columns: 22

Key Columns

- Sector metrics: Year, incorporated_country, Primary activity, Primary sector
- Financial metrics: Market_Cap_USD, Revenue_USD, netIncome_USD, etc.
- Emission metrics: Scope_3_emissions_amount, Scope_3_emissions_type, country_ghg_avg
- Country statistics: country_population_avg, country_gdp_avg

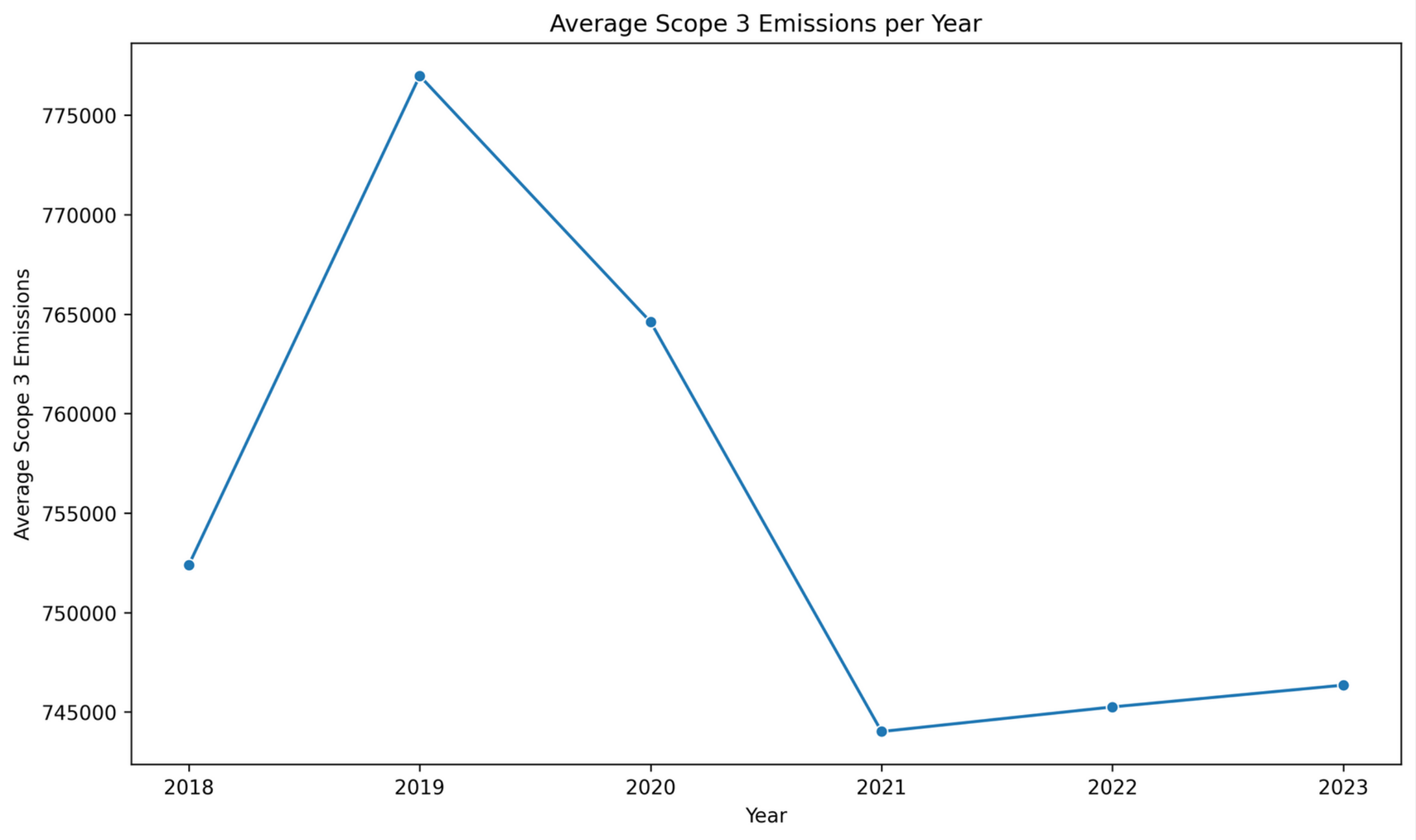
Exploratory Data Analysis

EDA is conducted to explore the relationships among greenhouse gas emissions, country statistics, and financial metrics.

We can gain a better understanding of the impact of environmental policies and practices on the economies of different countries.

- Emission Trends
- Correlation Matrix
- Emission Distribution
- Anomaly Detection
- Emissions by Sector

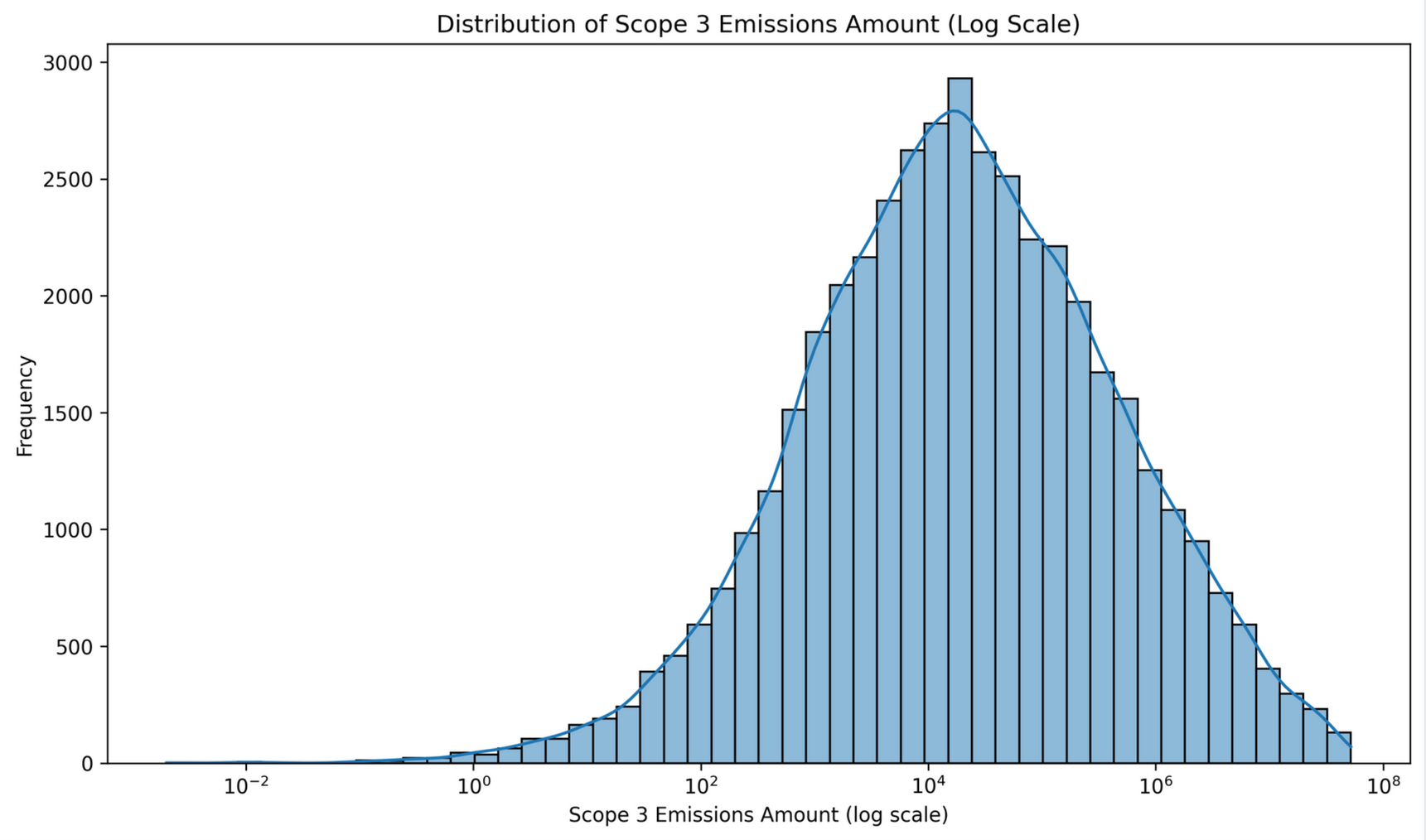
Emission Trends Over Time



This chart illustrates the average annual emissions year over year.

The sharp decrease in emissions from 2020 to 2021 was primarily due to the global lockdowns resulting from the COVID-19 pandemic.

Distribution of Scope 3 Emissions

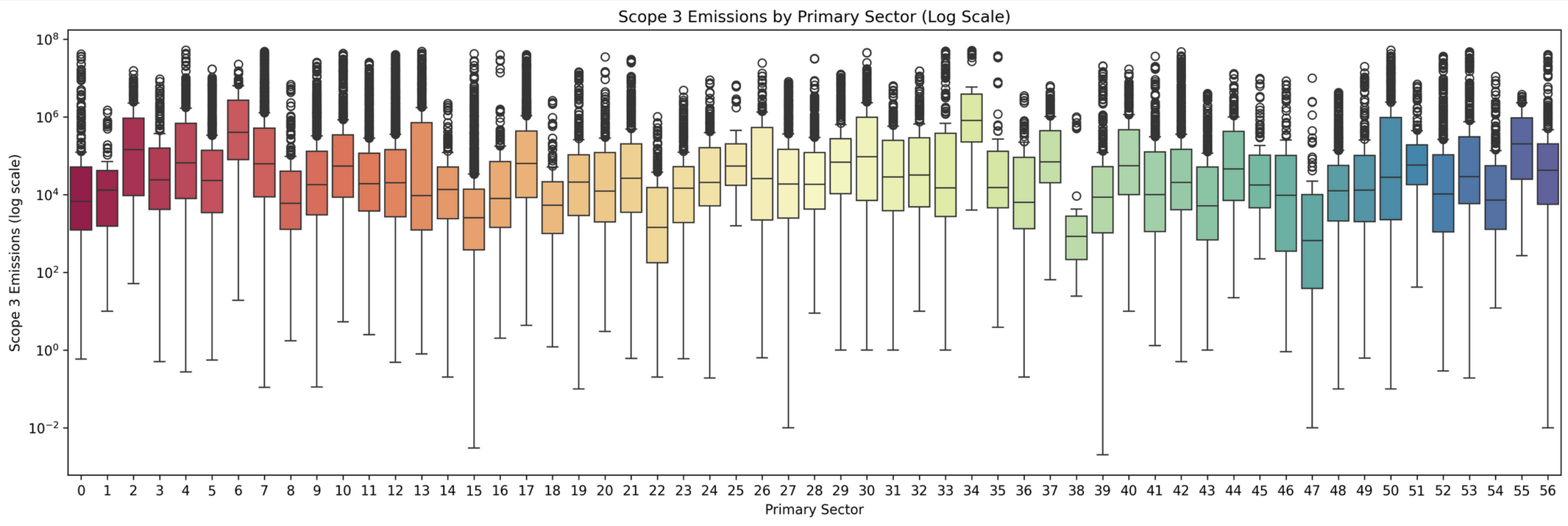


The distribution of Scope 3 emissions is highly skewed to the right.

A log scale transformation helps visualize the spread of data effectively.

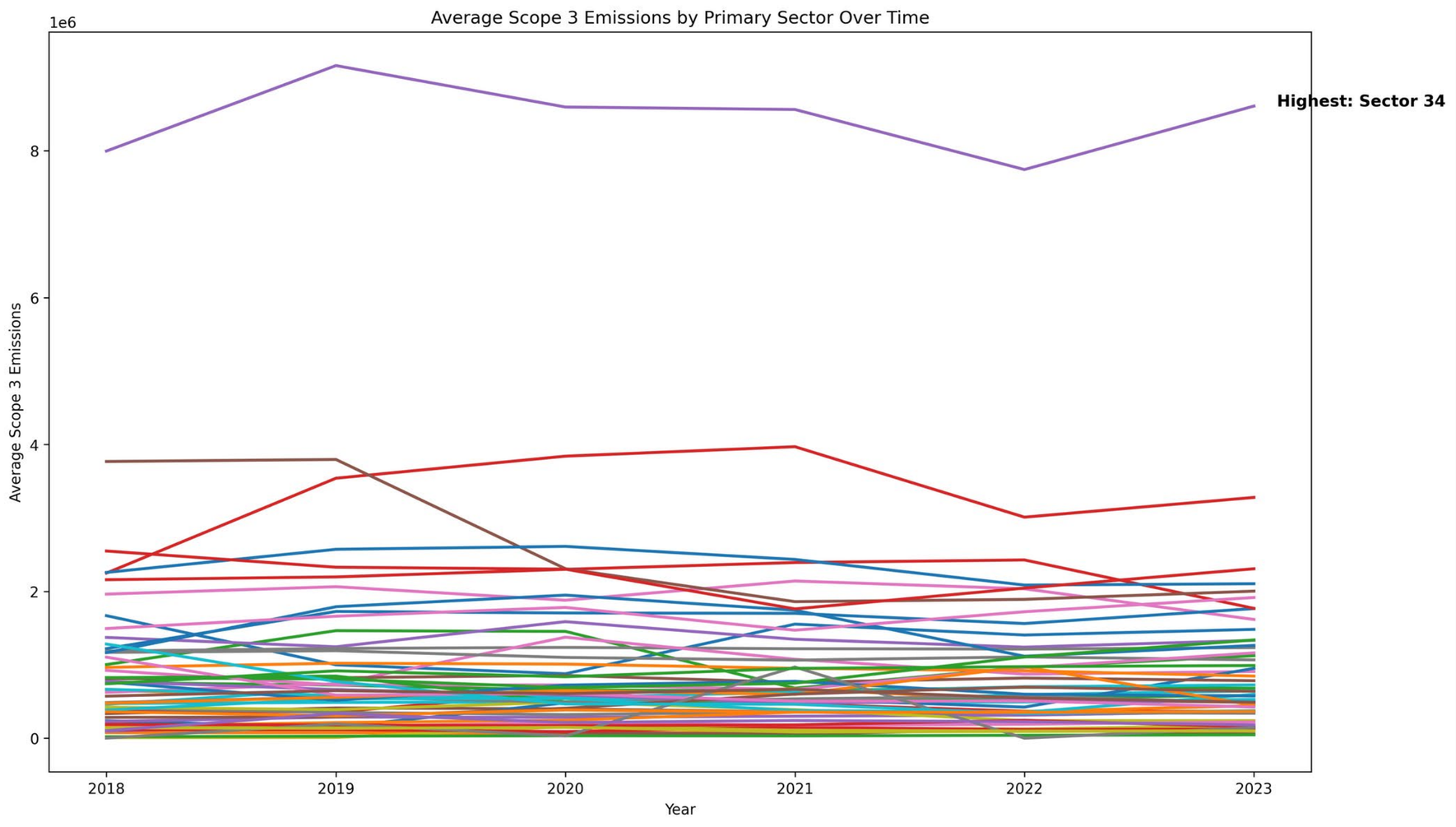
High-emission outliers disproportionately contribute to total emissions.

Scope 3 Emissions by Primary Sector



Some sectors display significantly higher median emissions and a broader range, indicating that emissions are strongly sector-dependent.

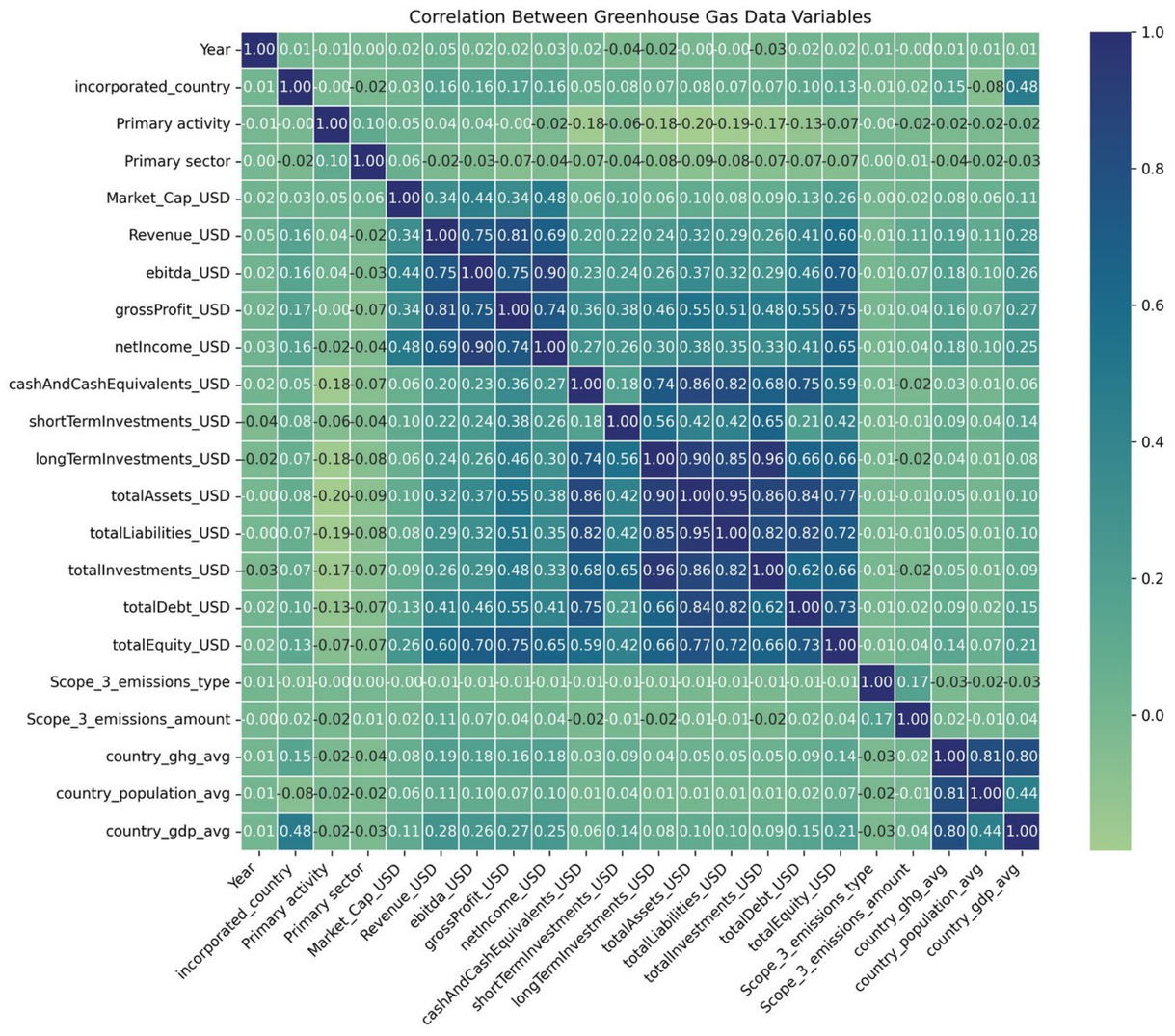
Can We Pinpoint Specific Contributors?



The plot identifies a sector with the highest average emission rate over time. Sector 34 is oil & gas processing (refineries).

Sector-specific strategies may be needed to curb emissions effectively.

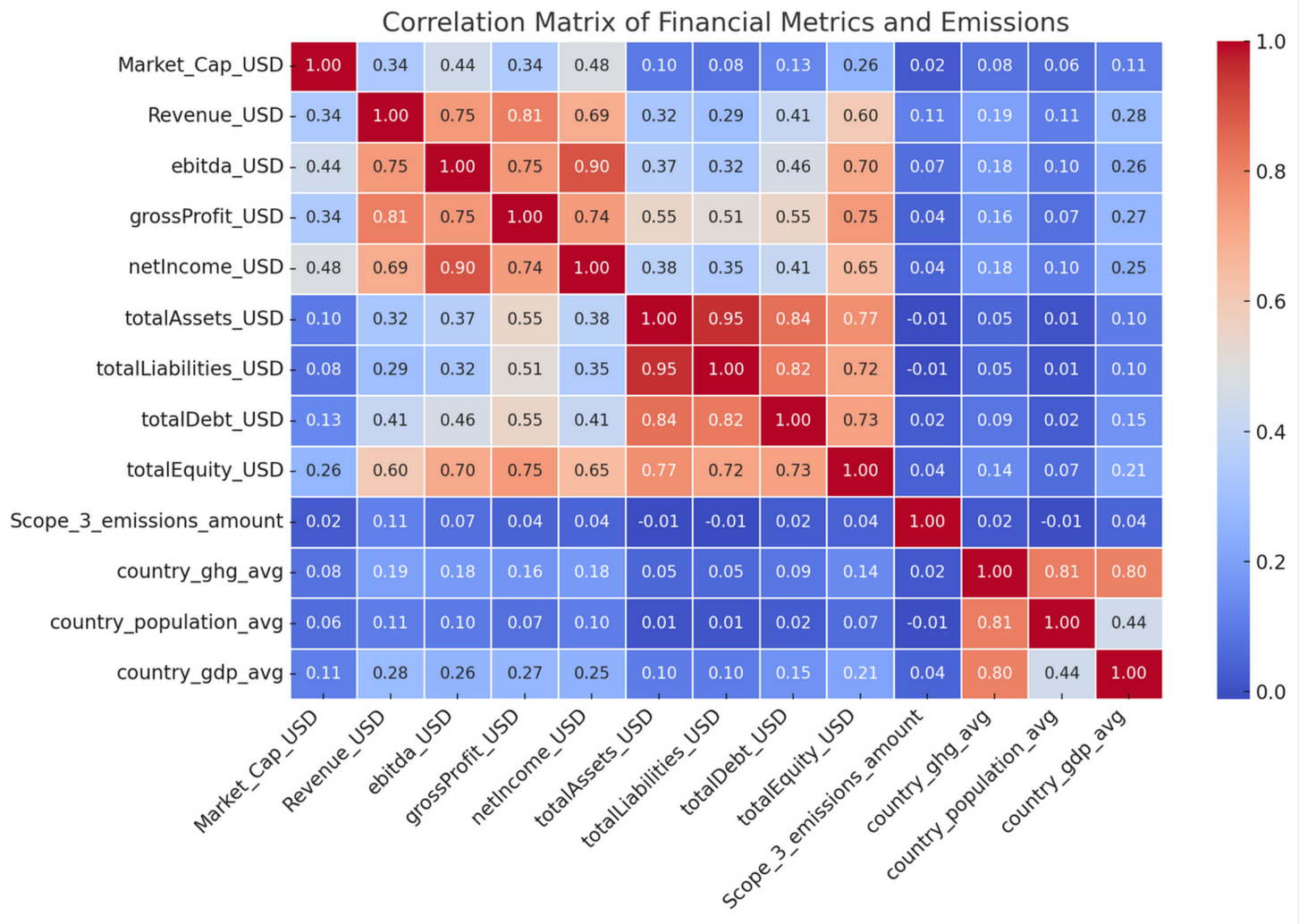
Correlation Matrix (All variables)



The heatmap clearly identifies smaller sections of strong correlations between financial metrics and emissions.

Variables with no inherent ordering or meaningful distance for clustering will be dropped from the analysis.

Correlation Matrix (Financial Metrics & Emiss



Strong positive correlations among financial metrics.

Emissions have notably weak correlations with most financial metrics.

Country-level metrics strongly correlate with emissions metrics.

Clustering Analysis

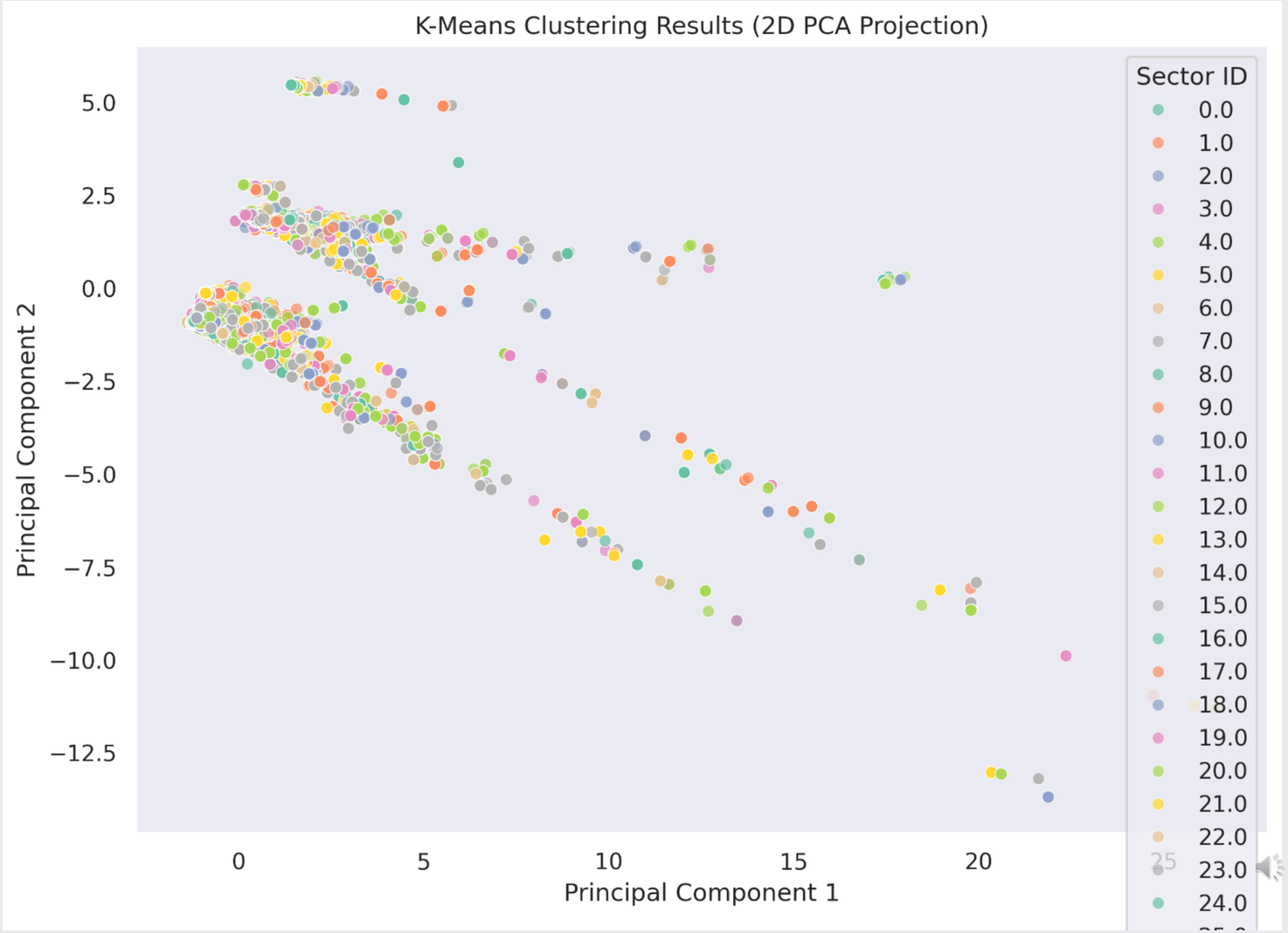
Clustering helps reveal patterns among entities (e.g., countries, sectors) with similar emission and economic profiles. The selected features below were chosen because they are factors that logically contribute to emission patterns.

- Scope_3_emissions_amount
- country_ghg_avg
- country_population_avg
- Revenue_USD
- Market_Cap_USD
- netIncome_USD

K-Means Clustering Results

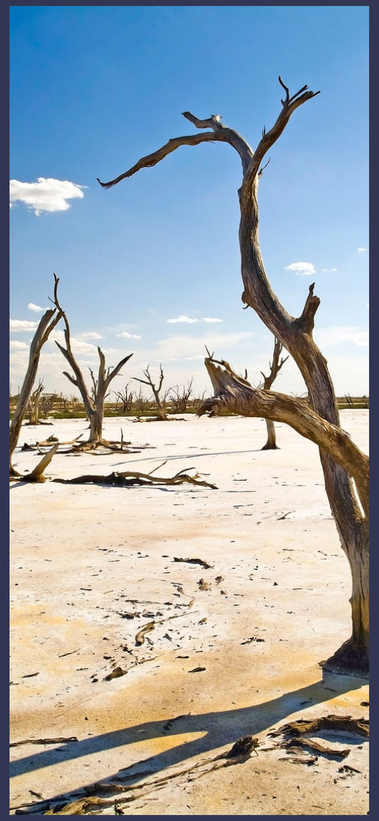
The silhouette score and the PCA plot of the model both show that the clustering of 11 (k=11) yields the best results.

The scatter plot visualizes the results of a K-Means clustering analysis, specifically showing how data points are clustered based on their Sector ID.



Interpreting the Clusters

- **Labeling Sector ID**
  The plot's colors denote different Sector IDs, effectively illustrating how entities from various sectors are distributed across clusters.
- **Overlapping Sectors**
  Some sectors appear across multiple clusters, showing that emissions behavior varies even within a single industry.
- **Outliers**
  These might be countries or companies with extreme Scope 3 emissions, exceptionally high or low market cap.
- **Conclusion**
  The unsupervised learning approach effectively illustrates multidimensional patterns beyond simple categorical grouping.



Future Steps

It would be valuable to identify which sectors dominate each cluster by cross-tabulating Cluster versus Sector ID, gaining insights into which industries contribute the most to each emission behavior group.

This can support targeted policy-making and sustainability strategies tailored to the emission profiles of different clusters.

THANK YOU

This discussion aims to provide valuable insights into the importance ofsustainable strategies to address greenhouse gas emissions.

415-963-1915

https://github.com/lokidamenace/GHG-Emissions

peter.h.la@outlook.com