

CPSC 368 Group 10 Assignment 2:

Idea 1: Global Climate & Sustainability Analysis

1. Explanation of the Proposed Idea

This project aims to conduct a comprehensive analysis of global climate change and sustainability efforts. By examining CO2 production, sustainability metrics, and climate-related development indicators across various countries, the project will uncover patterns, correlations, and potential strategies for effective climate action and sustainable development.

2. Relation to Climate and Sustainability

This analysis directly addresses the critical issues of climate change and sustainable development. It provides insights into CO2 emissions, a major factor in global warming, evaluates countries' progress on sustainability, and assesses how climate-related factors affect development. This holistic approach is crucial for understanding and addressing the multifaceted challenges of global sustainability.

3. Datasets

a. Dataset 1: [Countrywise CO2 Production](#)

ii. Reliability: Sourced from Wikipedia, so it is highly trustworthy.

b. Dataset 2: [World Sustainability Dataset](#)

ii. Reliability: A 'global sustainability' dataset was created for the TrueCue Women+Data Hackathon. The dataset tracks the performance of 173 countries against a range of sustainability metrics over 19 years.

c. Dataset 3: [World Development Indicators: Climate](#)

ii. Reliability: The data was sourced from the World Bank, so it's highly trustworthy

4. Data Analysis

a. Objective and Methods

- The analysis aims to explore global patterns in CO2 emissions, sustainability performance, and the impact of climate change on development indicators.
- The primary methods are descriptive statistics and simple regression analysis. This approach is best suited for the datasets because it provides a clear and straightforward understanding of the data, which is essential for identifying key trends and relationships.
 - Descriptive statistics will summarize and reveal trends within the datasets.
 - Regression analysis will help to understand and quantify the relationships between CO2 emissions, sustainability metrics, and development indicators.

- Alternative Approaches:
 - Time Series Analysis: This method is not chosen because the datasets include data points from various years and countries with different temporal resolutions, making it challenging to create a consistent time series for analysis.
 - Inferential Statistics: While useful for hypothesis testing, inferential statistics are not chosen here because the data is more aligned with descriptive and regression analysis objectives. Our data may not meet the assumptions required for valid inferential statistical tests, such as normality and random sampling.
- b. Purpose of the Data Analysis:
 - This analysis is crucial for understanding how different nations contribute to global CO2 emissions and how these emissions, along with other sustainability factors, impact development.
 - Such insights are vital for shaping effective environmental policies and sustainable development strategies at both national and international levels.
- c. Statistical Rigour:
 - Descriptive statistics will provide a foundational understanding of the datasets.
 - Regression analysis will uncover relationships between variables, controlling for other factors.

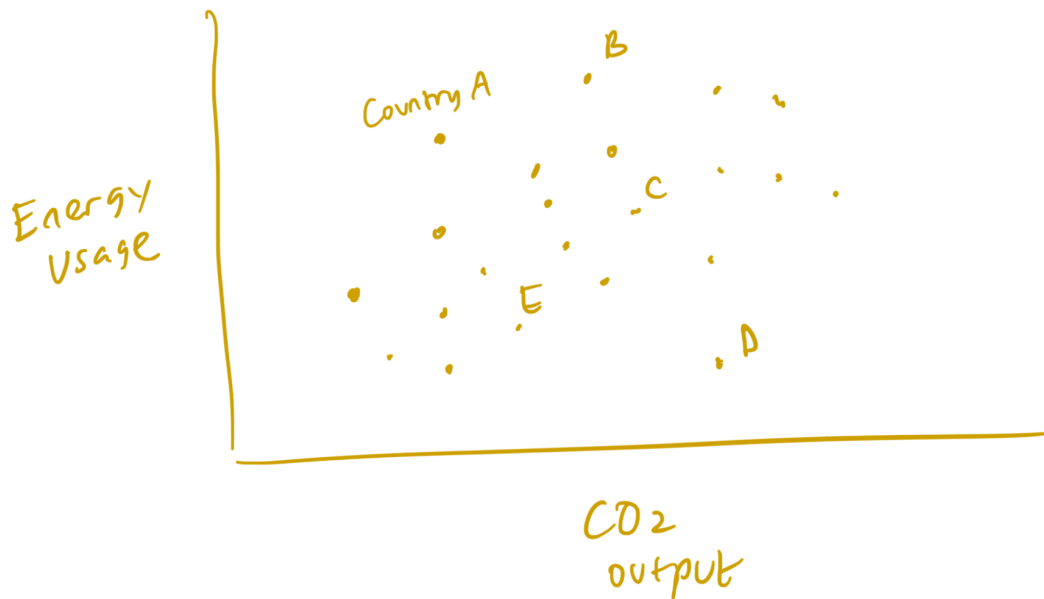
By using well-established statistical methods and carefully interpreting results within the datasets' contexts, the analysis ensures that the conclusions drawn are both statistically valid and relevant to real-world scenarios. The simplicity of these methods also aids in the clear communication of findings to a non-technical audience, which is essential for policy formulation and public understanding.

5. Visualization

Simple, clear visualizations using Matplotlib will be used:

- Line Graphs: To show trends over time in CO2 emissions and sustainability metrics.
- Bar Charts: For comparing CO2 emissions and sustainability scores among countries.
- Scatter Plots: To explore correlations between sustainability metrics and development indicators.

These visualizations will be designed to communicate complex data in an accessible and interpretable manner, suitable for a wide range of audiences.



Idea 2: Impact of Renewable Energy on Carbon Emissions and Economic Growth

1. Explanation of the Proposed Idea

This project aims to conduct a comprehensive analysis of the impact of renewable energy adoption on carbon emissions and economic growth across various countries. By examining renewable energy consumption data, CO2 emissions, and economic performance indicators, the project will identify trends, correlations, and potential strategies for promoting renewable energy use to achieve sustainable development and reduce global carbon footprints.

2. Relation to Climate and Sustainability

This analysis directly addresses the pivotal issues of climate change and sustainable development by focusing on renewable energy as a key factor in mitigating carbon emissions and fostering economic sustainability. It explores the balance between environmental sustainability and economic growth, providing insights into how countries can transition towards greener energy sources while maintaining or enhancing their economic development.

3. Datasets

a. Dataset 1: Renewable Energy (1960-2023)

- Reliability: Given the comprehensive coverage of renewable energy sources and the extensive timespan, this dataset is expected to be highly reliable for analyzing renewable energy trends.

- Link: <https://www.kaggle.com/datasets/imtkaggleteam/renewable-energy-1960-2023>

b. Dataset 2: Carbon Dioxide Emissions of the World (1990-2018)

- Reliability: Sourced from reputable climate data organizations, this dataset's credibility is solid for investigating CO2 emissions trends and their correlations with renewable energy adoption.

- Link: <https://www.kaggle.com/datasets/ankanhore545/carbon-dioxide-emissions-of-the-world>

c. Dataset 3: Global Economy Indicators

- Reliability: Compiled by the United Nations Statistics Division and other international agencies, ensuring high trustworthiness for analysing economic trends related to renewable energy and sustainability.

- Link: <https://www.kaggle.com/datasets/prasad22/global-economy-indicators>

4. Data Analysis

a. Objective and Methods

- This initial step will serve to summarize the datasets, providing a foundational overview of renewable energy usage, CO2 emission levels, and key economic performance metrics across various nations. By highlighting central tendencies, dispersion, and distribution shapes, this analysis will set the stage for a deeper exploration of the data.

- Based on these insights gained from descriptive statistics, regression analysis will be employed as a powerful tool to quantify the relationships between renewable energy consumption, carbon emissions, and economic outcomes. This approach will enable the identification of statistically significant predictors of environmental and economic health, offering a nuanced understanding of how shifts in energy consumption patterns may influence carbon footprints and economic vitality.

- Alternative Approaches

- Machine Learning Models: Not chosen due to the initial focus on identifying direct correlations and trends rather than predictive modelling.

- Time Series Analysis: While valuable, it's not selected as the primary approach due to the project's broad scope, aiming to compare across countries rather than forecasting future trends.

b. Purpose of the Data Analysis

The analysis aims to reveal how renewable energy adoption impacts carbon emissions and economic growth, informing sustainable energy policies and economic strategies. This insight is vital for supporting global sustainability efforts and guiding decisions that balance environmental and economic considerations.

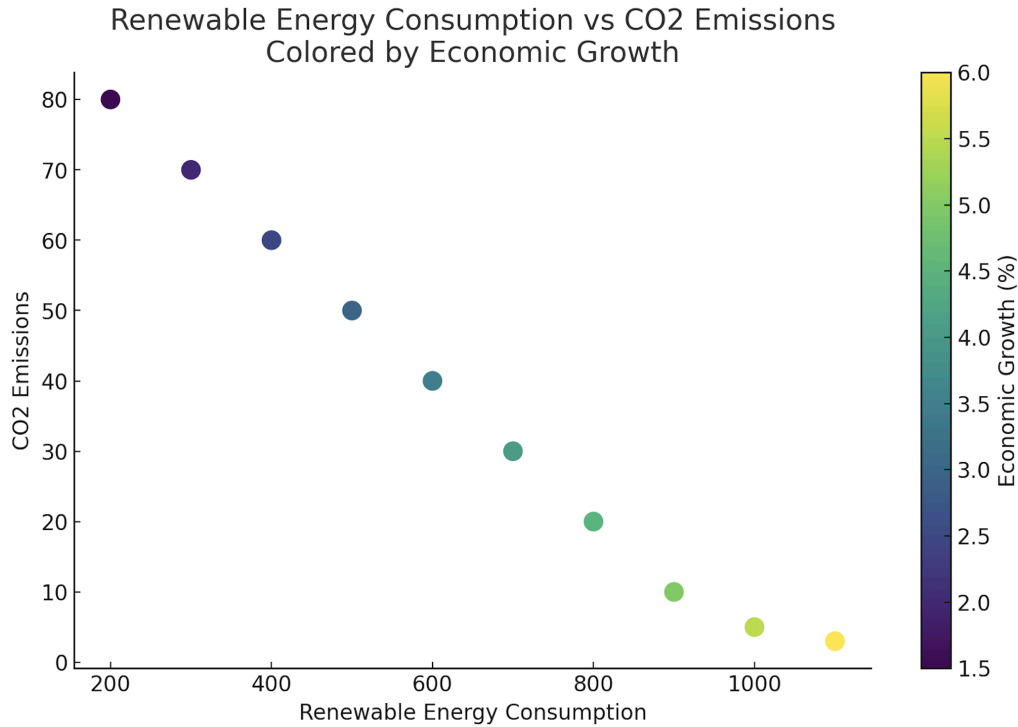
c. Statistical Rigour

The study employs regression analysis to ensure statistical rigour, analyzing the relationships between renewable energy use, CO2 emissions, and economic indicators. This approach allows for a reliable examination of how changes in energy consumption influence environmental and economic outcomes, providing a solid foundation for the conclusions drawn.

5. Visualization

We can use visualizations created using Matplotlib or Altair:

- Scatter Plots or line plots: For exploring correlations between renewable energy consumption and CO2 emissions, with economic growth factors potentially included as a third dimension such as colour encodings. We can implement drop-down menus to study the relations between the three by country.



Idea 3: Exploration of Relationship Between Food Production and Consumption with Greenhouse Gas Emissions

1. Explanation of the Proposed Idea

This idea relates to analysis of the environmental sustainability of food production and consumption in various countries. Specifically, we want to explore the relationship between food production, consumption, and greenhouse gas emissions, focusing on the agriculture and food production sectors.

2. Relation to Climate and Sustainability

It is well known that the production of animal products such as meat is less sustainable, as it can take up more land to provide food. By investigating the environmental impact of meat-related activities, we hope to quantify the impact of meat consumption and production and compare it to more sustainable practices. This project would help to understand the differences between different animal and crop production to encourage sustainable farming practices.

3. Datasets

- a. Greenhouse Gas Emissions Dataset:

Link: https://edgar.jrc.ec.europa.eu/report_2023#data_download

- The dataset is sourced from the Emission Database for Global Atmospheric Research (EDGAR), a government regulated website which is government funded research. Because this data is included in an official report by the European Commission, we can be reasonably sure that this data is trustworthy and reliable.

b. World Meat Consumption and World Crop Production

Link: <https://data.oecd.org/agroutput/crop-production.htm#indicator-chart>

- This dataset is compiled from official agricultural reports. Published by the Organization for Economic Co-operation and Development, this dataset is reliable because it is a respected international organization that collaborates with many countries to suggest policies for better lives. It provides detailed information on the number of animals slaughtered for meat consumption, and information about crop production.

c. Meat Consumption Dataset:

Link:

<https://www.kaggle.com/datasets/pragadesh06/meat-production-datasets1961-2018?select=global-meat-production.csv>

- These datasets are sourced from the Food and Agriculture Organization of the United States. Compiled from FAOSTAT, their large aggregate data, the dataset of interest from the kaggle site is the animals-slaughtered-for-meat dataset that summarizes the amount of animals slaughtered for meat by country since 1961. This data is reasonably trustworthy and reliable because it is provided by a government organization.

4. Data Analysis

a. Objective and Methods

- We aim to conduct a regression analysis to understand the relationship between GHG emissions and variables such as the number of animals slaughtered for meat and per capita meat consumption. This approach allows us to quantify the environmental impact of meat-related factors.
- Regression analysis provides a robust statistical framework for identifying the various factors contributing to GHG emissions. It allows for controlling confounding variables, providing a deeper understanding of the relationships between the datasets.
- Alternative approaches:
 - Time Series Analysis: While effective for tracking trends over time, it might be too complex given the large amount of time dependent data. There may also be confounding variables associated with the time difference such as technology differences and global events.
 - Analysis of Variance Test: Using ANOVA testing methods to see if there is a significant effect of meat consumption and production on GHG emissions by country. Although interesting, this test would not be the best because of the limited ways we could visualize and present our findings.

b. Purpose of the data analysis:

- Understanding how meat production and consumption contribute to GHG emissions is crucial for sustainable agricultural practices and environmental conservation. This

analysis can inform policies aimed at reducing the environmental impact of the food production/meat industry, aligning with the broader theme of sustainability.

c. Statistical Rigour:

- Regression analysis ensures statistical rigour by examining the relationship between variables while considering potential confounding factors.

5. Visualization

For this idea, an appropriate visualization would be two scatter plots with the estimated regression lines for production vs. GHG emissions and consumption vs GHG emissions. This would be less complex and more interpretable than a single regression line in a 3D space. Alternatively, a heatmap of the globe could be used to express the findings of our analysis.

