Energy Consumption and Production Analysis

SDG 12: Responsible consumption and production

Concept of the Project:

The goal of this project is to analyse global energy consumption and production patterns and develop a machine learning model to predict future energy consumption. By using historical data on various forms of energy consumption and production, the project aims to uncover insights that can inform energy policies and strategic planning. This will support sustainable energy use by providing a data-driven foundation for decision-making and policy development.

Problem Statement:

Energy consumption and production are critical components of global economic development and environmental sustainability. With the increasing emphasis on renewable energy sources and the need to reduce carbon footprints, understanding trends and predicting future energy consumption are more important than ever. This project utilizes datasets containing information on primary energy consumption, electricity generation, fossil fuel consumption, renewable energy shares, and more to develop predictive models and derive actionable insights.

Objectives of the project:

- **Data Preparation and Cleaning:** Collect and preprocess data from multiple sources to ensure consistency and accuracy.
- **Exploratory Data Analysis (EDA):** Conduct thorough EDA to understand historical trends, identify patterns, and uncover insights in energy consumption and production.
- **Feature Engineering:** Select and engineer features that are most relevant for predicting energy consumption.
- **Model Development:** Develop and train machine learning models to predict future energy consumption based on historical data.
- **Model Evaluation:** Evaluate the performance of the models using appropriate metrics to ensure accuracy and reliability.
- **Insights and Recommendations:** Provide insights derived from the data analysis and model predictions to inform policy and decision-making.

Data Sources:

The project will use air quality datasets from the following sources:

1. **OurWorldinData:** Energy Production and Consumption-link
2. **OurWorldinData:** Ensure access to affordable, reliable, sustainable and modern energy for all-link
3. **Kaggle:** Global Electricity Demand and Generation Dataset-link
4. **Kaggle:** Renewable Power Generation and weather Conditions-link

Features:

The project will utilize the following datasets:

1. **Change in Energy Consumption**: Annual percentage change in primary energy consumption.
2. **Consumption:** Total fossil fuels consumption, nuclear electric power consumption, total renewable energy consumption, primary energy net exports, and total primary energy consumption.
3. **Electricity Generation**: Annual electricity generation in TWh.
4. **Global Energy Substitution**: Substituted energy by various sources (e.g., renewables, biofuels, solar, wind, hydropower, nuclear, gas, oil, coal, traditional biomass).
5. **Primary Energy Consumption**: Total primary energy consumption in TWh.
6. **Production:** Total fossil fuels production, nuclear electric power production, total renewable energy production, primary energy imports, and total primary energy production.
7. **Renewable Energy Share:** Share of renewable energy in the total final energy consumption.

Tool for Analysis (Use any tool, even excel):

The following tools and technologies will be used for data analysis:

1. Python: For data cleaning, analysis, and visualization, using libraries such as Pandas, NumPy, Matplotlib, and Seaborn.
2. Jupyter Notebooks: For documenting the analysis process and visualizations.
3. Scikit-learn: For developing predictive models and machine learning algorithms.

Hypothesis:

Hypothesis Testing for Energy Consumption and Production Analysis

Hypothesis testing is a critical component of any data-driven project. It allows us to make inferences about populations based on sample data and provides a statistical framework for decision-making. In the context of this project, we can formulate and test several hypotheses related to energy consumption and production.

Hypothesis 1: Change in Primary Energy Consumption

- Null Hypothesis (H_0): There is no significant annual change in primary energy consumption over the years.
- Alternative Hypothesis (H_1): There is a significant annual change in primary energy consumption over the years.

Hypothesis 2: Relationship Between Fossil Fuel Consumption and Renewable Energy Share

- Null Hypothesis (H_0): There is no significant relationship between fossil fuel consumption and the share of renewable energy in total final energy consumption.
- Alternative Hypothesis (H_1): There is a significant relationship between fossil fuel consumption and the share of renewable energy in total final energy consumption.

Hypothesis 3: Impact of Renewable Energy Substitution on Total Energy Consumption

- Null Hypothesis (H_0): The substitution of renewable energy sources does not significantly impact total primary energy consumption.
- Alternative Hypothesis (H_1): The substitution of renewable energy sources significantly impacts total primary energy consumption.

Methodology:

- **Data Preparation:**
- Load datasets and parse dates correctly.
- Drop irrelevant columns and handle missing values appropriately.
- Standardize and scale features for modeling.
- **Exploratory Data Analysis:**
- Generate summary statistics and visualizations to understand data distributions and trends.
- Identify significant patterns and correlations in the data.
- **Feature Engineering:**
- Select relevant features for predicting total primary energy consumption.
- Engineer new features if necessary, such as lagged variables or interaction terms.
- **Model Development:**
- Split the data into training and testing sets.
- Train a linear regression model as a baseline.
- Explore more advanced models (e.g., decision trees, random forests, gradient boosting) to improve predictions.
- **Model Evaluation:**
- Use metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared to evaluate model performance.
- Compare performance across different models and select the best-performing one.
- **Insights and Recommendations:**
- Analyse model predictions and generate insights on future energy consumption trends.
- Provide recommendations based on findings to support energy policy and strategic planning.

Probable Outcomes:

- **Comprehensive Data Analysis:** Detailed understanding of historical energy consumption and production patterns.
- **Predictive Model:** A machine learning model capable of predicting future energy consumption with high accuracy.
- **Actionable Insights:** Insights and recommendations to inform energy policy, focusing on sustainable and efficient energy use.

Conclusion:

This project aims to harness the power of data and machine learning to provide valuable insights into global energy consumption and production. By developing robust predictive models and uncovering key trends, the project will contribute to informed decision-making and strategic planning in the energy sector, ultimately supporting the transition to sustainable energy systems