[Kopie van Untitled1.ipynb - Colab (google.com)]

Question 1: What is the biggest predictor of a large CO2 output per capita of a country?

Answer is: co2_per_gdp

The dataset is sourced from "Our World in Data" and is available as a CSV file at **https://raw.githubusercontent.com/owid/co2-data/master/owid-co2-data.csv**

I used `pd.read_csv(url)` to load the dataset into a Pandas DataFrame.

Using `print(df.columns)`, I printed the column names to verify which columns are available in the dataset.

I selected relevant columns that are typically predictors of CO2 emissions.

These columns were chosen because they represent factors related to economic activity, energy consumption, and population, which are commonly linked to CO2 emissions.

After inspecting the actual column names using `print(df.columns)`, I found the actual column names and adjusted the column selection accordingly.

Specifically, I renamed columns to match the expected names for analysis.

Dropped rows with missing values using `df.dropna()`.

Focused on data from the most recent year available in the dataset.

Used `sns.pairplot()` to create pair plots, which help visualize relationships between variables.

Defined the predictors (`X`) and the target variable (`y`)

Split the data into training and testing sets using `train_test_split()`.

Fitted a linear regression model using `LinearRegression()` from scikit-learn.

Evaluated the model using mean squared error and R-squared metrics.

Analyzed the coefficients of the linear regression model to determine which predictor has the largest impact on CO2 emissions.

Identified the biggest predictor by finding the variable with the largest absolute coefficient.

Question 2: Which countries are making the biggest strides in decreasing CO2 output?

Answer is: OECD

The data is sourced from "Our World in Data," and is available in CSV file at **https://raw.githubusercontent.com/owid/co2-data/master/owid-co2-data.csv**

I used the `pd.read_csv(url)` function from the Pandas library to load this data into a Pandas DataFrame.

To focus on the analysis, I selected columns relevant to CO2 emissions: 'country', 'year', and 'co2'.

I removed rows with missing values using `df.dropna(inplace=True)` to ensure that the analysis would be based on complete data.

I defined a period from 2010 to 2020 to analyze changes in CO2 emissions over the last decade.

The data was filtered to include only records within this period.

I pivoted the DataFrame to have years as columns, with countries as rows and CO2 emissions as the values. This format helps in calculating the change in CO2 emissions over time.

The change in CO2 emissions was calculated by subtracting the CO2 emissions in 2010 from those in 2020 for each country.

The countries were sorted based on the calculated change to identify those with the largest decreases.

I selected the top 10 countries with the largest decreases in CO2 emissions.

A horizontal bar plot was created to visually represent these countries and their corresponding changes in CO2 emissions.

Question 3: Which non-fossil fuel energy technology will have the best price in the future?

Answer is: Solar PV and Onshore wind

Projected Costs of Generating Electricity 2020 – Analysis - IEA

I searched for credible sources that provide comprehensive analyses of the future costs of various renewable energy technologies. The primary sources used were:

National Renewable Energy Laboratory (NREL): The NREL publishes the Annual Technology Baseline (ATB), which includes detailed projections of the costs and performance of renewable energy technologies.

International Energy Agency (IEA): The IEA publishes reports on the projected costs of generating electricity, which include data on various energy technologies and their expected costs.

I conducted a web search for datasets and reports from these organizations that specifically address future cost projections for non-fossil fuel energy technologies. The search included terms like "future cost projections of renewable energy technologies dataset" and "NREL ATB 2020" to find the most relevant data.

I reviewed the following key reports:

NREL's Annual Technology Baseline (ATB) 2020: This report provides projections for the levelized cost of energy (LCOE) for various renewable technologies, including solar PV, onshore wind, offshore wind, and other emerging technologies. It highlights trends and expected cost reductions due to technological advancements and increased deployment (NREL Website) (NREL Data).

IEA's Projected Costs of Generating Electricity 2020: This report includes a comprehensive analysis of the LCOE for different energy technologies, comparing renewables with fossil fuel-based generation and nuclear power. The IEA report emphasizes the competitiveness of solar PV and onshore wind in various regions, especially under carbon pricing scenarios (IEA).

The analysis from these reports showed that:

- **Solar Photovoltaic (PV): Solar PV is projected to continue experiencing significant cost reductions, making it one of the least expensive options for electricity generation by 2050.**
- **Onshore Wind: Onshore wind is also expected to remain highly cost-competitive, with ongoing technological improvements and economies of scale driving down costs.**
- **Regional Variations: The competitiveness of these technologies can vary by region due to factors like resource availability, local policies, and market conditions.**

Conclusion

Based on the detailed projections from NREL and IEA, solar PV and onshore wind are the non-fossil fuel energy technologies expected to have the best prices in the future. These findings are supported by extensive data and analysis provided by these reputable sources.