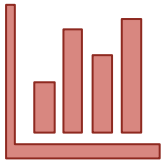
World Energy Consumption Analysis

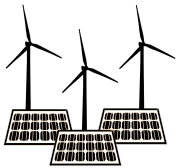
DS8003 - Management of Big Data and Tools

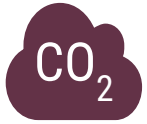
Final Project

World Energy Consumption: Is world striving for carbon neutrality by 2050-70, while over 50% of its energy is still sourced from coal and oil in today's date?

- Significant shifts in global energy dynamics over the past 50 years, driven by technology advancements, emerging energy sources, and growing climate awareness, highlight the need for understanding and analyzing these changes
- In this project, we are trying to analyze consumption trends for primary energy sources on global as well as country level over the past few decades, to understand the shifts
- We are exploring different factors such as -
    - which parts of the world consumes most energy and what trends do we see in their dependency on the various sources of energy (e.g. coal, oil, hydro, wind etc.)?
    - How is a country's energy consumption and growth related?
    - How is world moving towards renewable energy usage for global climate change goals?
    - Which countries have the highest percentage of renewable energy in their energy mix for electricity generation?
- What insights developing countries can take from developed nations and smart cities to enhance their energy mix or vice versa







Work Distribution: balanced individual learning alongside project completion

Project was segmented in four major tasks after proposal - Data Discovery & EDA, Analysis on data and insight generation, Visualization & Documentation. While learning concepts like mapreduce individually and later applying our learnings into the project, we divided responsibilities for other tasks among us to meet timeline.

01 **Amarpreet Kaur** - Data Discovery, EDA, Hive SQL queries, Insight Analysis - 1, 4, Report

02 **Ruchi Parmar -** Proposal, Data Statistics & Wrangling, EDA, Insight Analysis - 2, 3, 5, Presentation

03 **Kartikey Chauhan -** Proposal, Data Transformation, Docker, Airflow, Visualization

04 **Hamna Ashraf -** Data Discovery

Data description

Data Schema

Level of data - country-year & region-year level
It consists of data starting from year 1900 till year 2022

Key Features

- Country's GDP, population
- Consumption data for renewable & non-renewable energy sources like Coal, Oil, Hydro, Wind, Solar etc.

Data shape

Data Contains 17432 rows and 122 features

Data Source

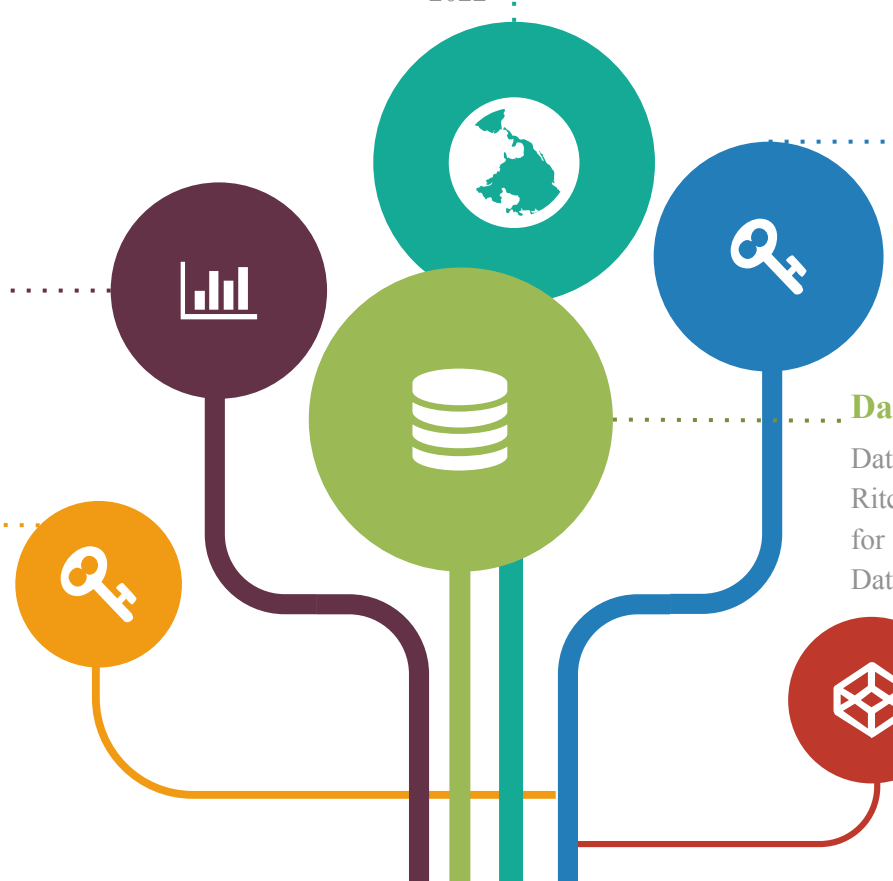
Dataset is originally collected by Hannah Ritchie, Max Roser and Edouard Mathieu for their research work - Our world in Data

Key Features

- Production data for Coal & Gas
- Change percentage in energy consumption year by year
- Electricity production and share of different energy sources for the same

Null information

- Original data has ~68% nulls
- Most of the countries have data available from 1990s



Process & Tools Description

Original Data CSV file

HDFS - File Storage

Stored the original file in Hadoop HDFS because of its large scale, and storing it in HDFS will facilitate running parallel operations.

Hadoop MapReduce - Basic Data Discovery

Basic EDA in data discovery stage such as counting nulls, finding min and max, that can be executed in parallel, performed using MapReduce - to learn & understand how mapreduce work at ground level.

Spark - EDA & Data Transformation

We transformed our data according to analysis requirements using Spark RDD transformations and a variety of operations provided by Spark.

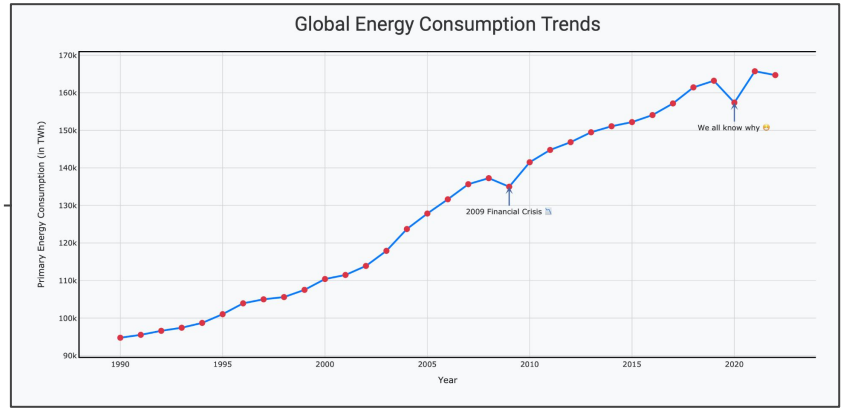
Hive - Data Query for Insight Analysis

Utilizing Hive data queries, and its partitioning functionalities to store data on country/year level and analyzed to find insights.

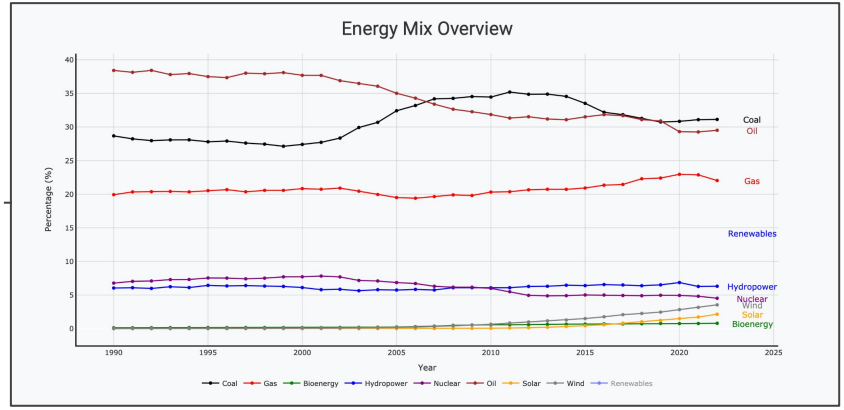
Plotly - Visualization tool

We converted our insights into Plotly visualizations to enhance user experience and understanding.

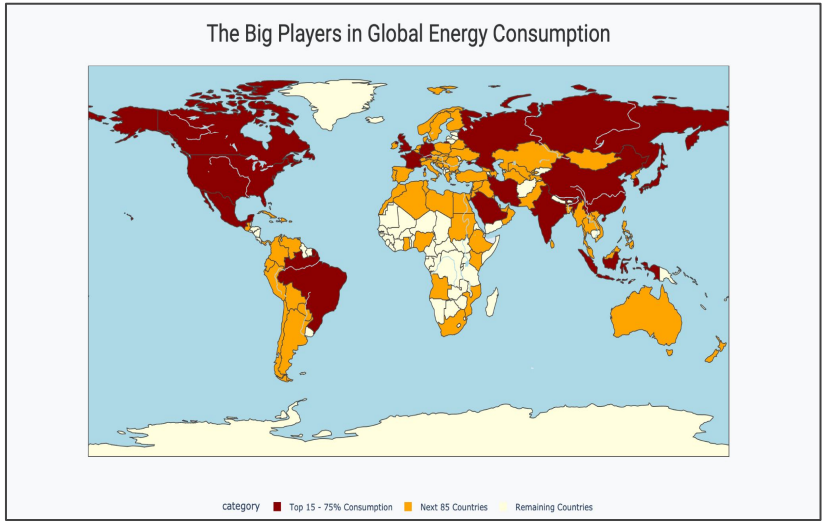
Analysis & Insights - part I



The world's energy consumption has surged by over 70 percent in the past three decades. A snapshot of global energy utilization in the year 2021 reveals a total consumption of 164710.98 terawatt-hours (tWH) of primary energy. We see that global energy consumption has increased nearly every year for more than 3 decades, with exceptions in 2009 and 2020.



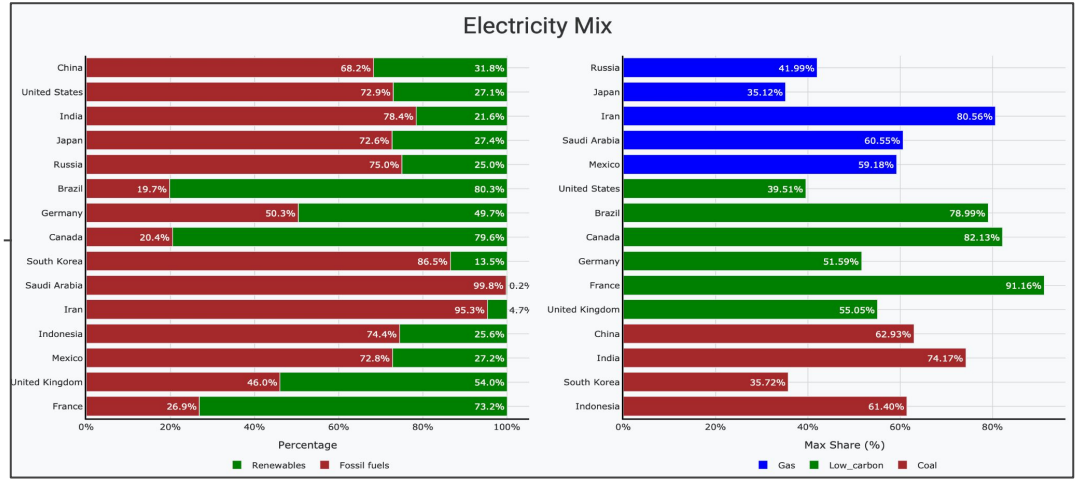
Coal consumption increased significantly between 1990 and 2010, then stabilized. There has been a marked increase in the use of Solar and Wind energy sources. Nuclear energy usage has remained constant, possibly due to safety concerns, despite its potential as a fossil fuel alternative.

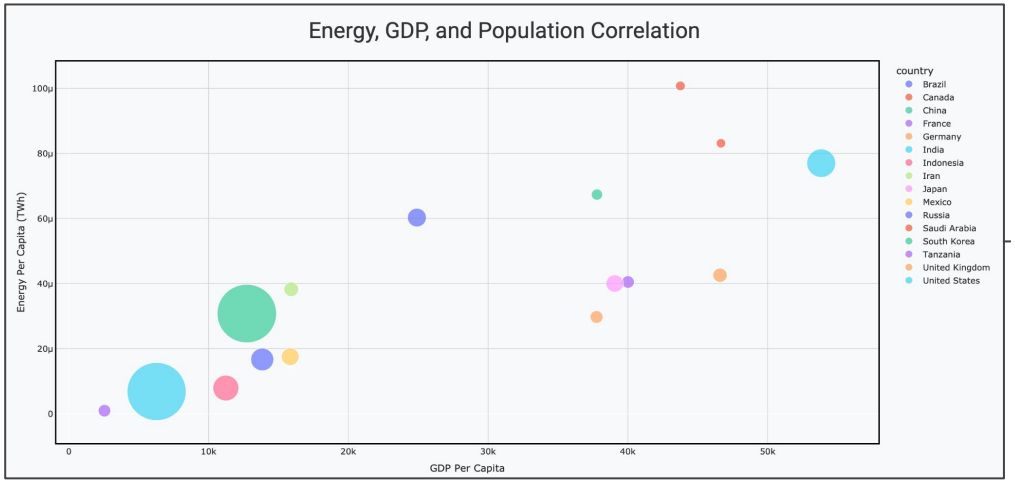


Top 15 countries, accounting for 75% of global energy consumption, highlight the importance of major economies leading the clean energy transition. Over 130 countries with consumption below 100 tWH face the dual challenge of energy poverty and sustainable development. These nations have a unique opportunity to learn from the larger economies' experiences, adopting best practices and pivot to cleaner, more sustainable energy solutions.

Analysis & Insights - part II

Countries like Brazil, Canada, and France, among the top electricity-generating countries, distinguish themselves by prioritizing renewable energy. In contrast, major contributors like China, United States, India, Russia, and Japan exhibit a higher reliance on fossil fuels in their electricity generation. This disparity underscores the global challenge of transitioning towards cleaner energy sources and emphasizes the need for concerted efforts to promote sustainability.The 'Max Share' % reveals key insights into each country's primary electricity generation sources in 2021. For instance, China and India rely heavily on coal , while United Kingdom emphasizes low-carbon sources. Brazil and Canada showcase a commitment to low-carbon energy, with shares of 79% and 82%, respectively. France stands out with a remarkable 91% share from low-carbon sources, primarily nuclear and renewables.





Global energy consumption patterns highlight a complex relationship with GDP and population size. Countries with smaller populations, such as the USA and Canada, show higher per capita energy use compared to populous nations like China and India. This underscores the role of energy accessibility and availability in shaping consumption trends, indicating that factors beyond population size are key determinants of national energy usage.

Lessons Learned

Transformation & Analysis with Big data tools

learned the strengths of different tools and how to leverage them for optimal performance and efficiency in data transformation processes like map-reduce, RDD transformations, partitioning etc.

Visualization of analysis and insights

Learned how analysis and insights can be transformed into visualization for better user experience and story telling

Workflow Automation & Collaboration

- Understanding how to use Apache Airflow for workflow automation, task scheduling, and overall project orchestration.

- Learned to containerize every aspect using Docker, creating a centralized environment that fosters parallel collaboration, enhancing efficiency for each team member

Life Cycle of Data Analysis

Gained a comprehensive understanding of the entire data lifecycle, its journey from the original source through data injection, exploratory data analysis (EDA), thorough analysis, visualization, and beyond, including future scope considerations