Impact Analysis of Socio-economic Factors on Renewable Energy Consumption & Solar Radiation Potential in Emerging Economies

Project Github: https://github.iu.edu/judonato/Big_Data_Final_Project

Introduction

This project investigates the intricate relationship between socio-economic development and renewable energy utilization in emerging economies. Utilizing the Google Cloud Platform, this study analyzes a wide array of socio-economic metrics from World Bank Open Data and solar radiation data from the National Solar Radiation Database. Focusing on Indonesia, China, India, South Africa, and Brazil, the research aims to elucidate the influence of various socio-economic factors on renewable energy consumption and assess the potential for solar energy in these rapidly developing nations. This exploration is pivotal in understanding how emerging economies can transition to sustainable energy practices in the face of global environmental challenges.

Background

The selection of this topic is driven by the critical need to understand and facilitate the energy transition in emerging economies, a concept central to addressing global climate change challenges. The "new green dilemma" as discussed by Xu et al. highlights the complex scenario faced by emerging economies, balancing the pursuit of sustainable energy solutions and the developmental challenges posed by established economies (Xu et al, 2020). This research is further motivated by the findings of Salim and Rafiq, who analyze the factors influencing renewable energy consumption in major emerging economies, revealing a bidirectional link between renewable energy, income, and pollutant emissions (Salim & Rafiq, 2012). The International Energy Agency emphasizes the significant yet underfunded role of these economies in the global clean energy transition, a gap widened by the Covid-19 pandemic (IEA, 2021). Lastly, the influence of educational attainment on renewable energy use in emerging market economies, as explored by Sart et al., underscores the importance of socio-economic factors in the adoption of sustainable energy practices (Sart et al, 2022). This project seeks to provide insights into these dynamics, offering guidance for policy and investment decisions in the sustainable energy sector of emerging economies.

Methodology

<u>Part 1: Data Extraction and Storage Using GCP</u>

Technological Setup:
- Google Cloud Platform (GCP): Primary platform for data processing and analysis.
- Virtual Machines (VM): Hosted on GCP to execute data extraction and transformation scripts.
- Google Cloud Storage (GCS): Used for storing extracted and processed data.

Steps:

1. Data Extraction from National Solar Radiation Database:
  - Developed a Python script `BigData_ETL.py` to extract data for the year 2019 from the National Solar Radiation Database using their API (https://developer.nrel.gov/api/hsds).
  - The script focuses on Global Horizontal (ghi) irradiance data for countries including Indonesia, China, India, South Africa, and Brazil.
  - The extraction process involves reading HDF5 files, filtering data based on specified criteria, and generating outputs like time indices and country-specific metadata.

2. Infrastructure Setup on GCP:
  - Configured a Virtual Private Cloud (VPC) to establish a secure and isolated network environment for the project's resources.

Name ↑	Subnets	MTU	Mode
vpc-network-1	39	1460	Auto

Fig (1): VPC network creation
  - Set up firewall rules for the VPC to manage and secure network traffic, specifying protocols, ports, and destination filters.

3. VM Configuration and Setup:
  - Launched and set up a VM instance on GCP. This involved choosing the right machine type, operating system, and network settings.

☐	Status	Name ↑	Zone
☐	✔	instance-1	us-central1-a

Fig (2): VM Creation
  - Installed necessary software and libraries (like pip, git, numpy, pandas, matplotlib, datetime, h5pyd) on the VM for data processing and analysis.

4. Data Processing and Storage:
  - Executed the `BigData_ETL.py` script on the VM, which processed and extracted relevant data.

```
judonato@i535-instance:~/project/Big_Data_Final_Project$ python3 BigData_ETL.py
CSVs and PNG maps created successfully.
judonato@i535-instance:~/project/Big_Data_Final_Project$ cd output
judonato@i535-instance:~/project/Big_Data_Final_Project/output$ ls
 China_ghi.csv                          'Indonesia_ghi_2019-12-31 12:00:00.png'
'China_ghi_2019-12-31 12:00:00.png'     'South Africa_ghi.csv'
 India_ghi.csv                          'South Africa_ghi_2019-12-31 12:00:00.png'
'India_ghi_2019-12-31 12:00:00.png'      time_index.csv
 Indonesia_ghi.csv
```

Fig (3): Successful Data Extraction from API.
  - Used the `gsutil cp` command to transfer the processed data from the VM to a designated bucket in Google Cloud Storage.

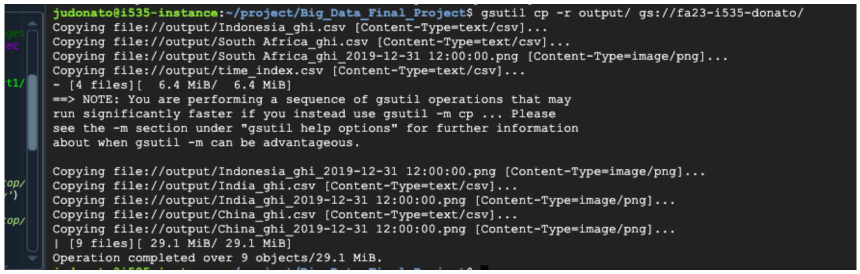


Fig (4): Data Successfully copied to bucket

Part 2: NoSQL Database Implementation for World Bank Data

Technological Setup:
- Google Bigtable: Chosen for its NoSQL database capabilities, ideal for handling large volumes of data.
- Google SDK: Provides tools for interacting with Google Cloud services, crucial for database management.
- Python Script (nosql_preprocess.py): Custom script for data preprocessing and formatting tailored for Bigtable.

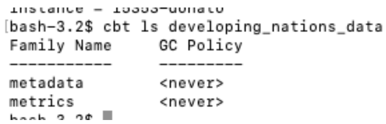
Steps:

1. Data Preprocessing and Preparation:
   - The `nosql_preprocess.py` script processes World Bank Open Data metrics, converting them into a format suitable for Bigtable (wide-column store).
   - It involves creating a unique row key for each data entry based on country code, series code, and year, ensuring efficient data retrieval.
   - The script systematically organizes the data and prepares it for insertion into the Bigtable database.

2. Bigtable Configuration and Setup:
   - Executed Google Cloud SDK setup and configuration on the local machine, including setting up authentication credentials.
   - Created a Bigtable instance using GCP console, defining the necessary column families and table structures.
   - The Bigtable schema was designed to efficiently store and query the large datasets involved in the project.

3. Inserting Data into Bigtable:
   - The preprocessed data was then written to the Bigtable using a Python script.
   - This step included converting data into the appropriate format, creating rows in the table, and populating the cells with the preprocessed data.

- Ensured error handling and verification of successful data insertion, with a focus on maintaining data integrity and accuracy.



```
bash-3.2$ cbt ls developing_nations_data
Family Name     GC Policy
-----------     ---------
metadata        <never>
metrics         <never>
```

Fig (5): Successful BigTable table creation

4. Data Analysis and Querying:
- Utilized the Jupyter notebook `No_SQL_Database_Analysis.ipynb` to perform in-depth analysis by querying the BigTable database.
- The analysis was geared towards uncovering insights into the socio-economic factors influencing renewable energy consumption in the target countries.

Results

<u>Results Part 1: Data Extraction and Storage Using GCP- National Radiation Dataset</u>

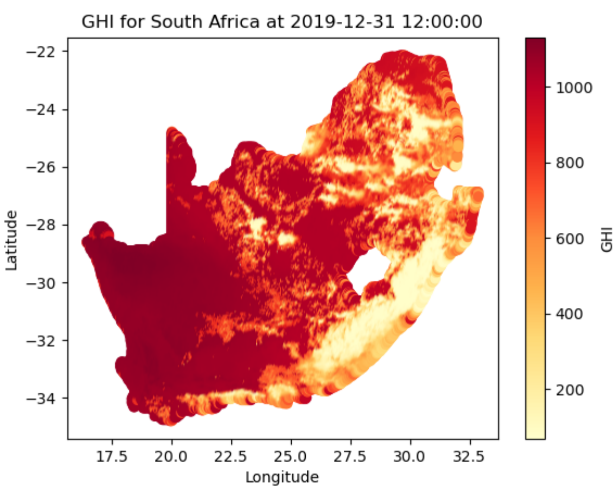


Fig (6): GHI for South Africa

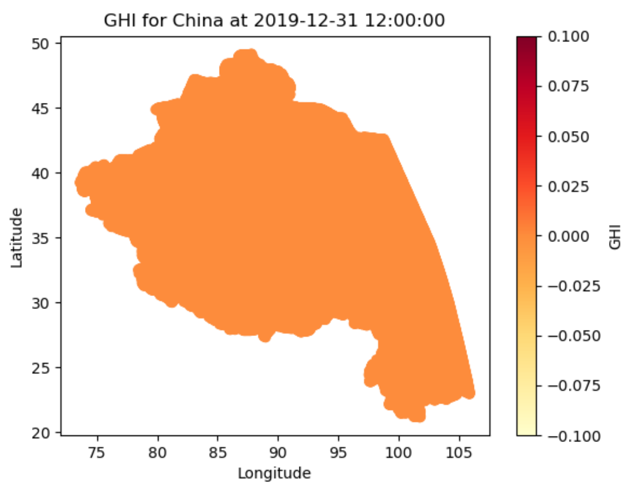


Fig (7): GHI for China

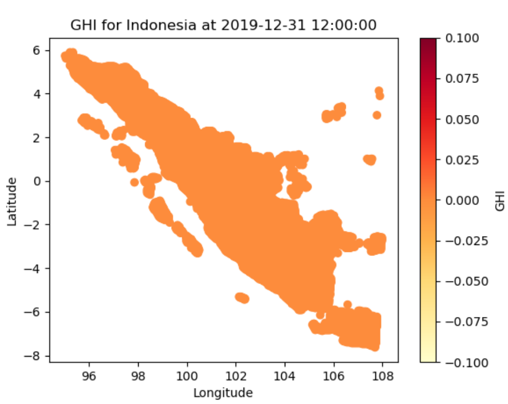


Fig (8): GHI for Indonesia

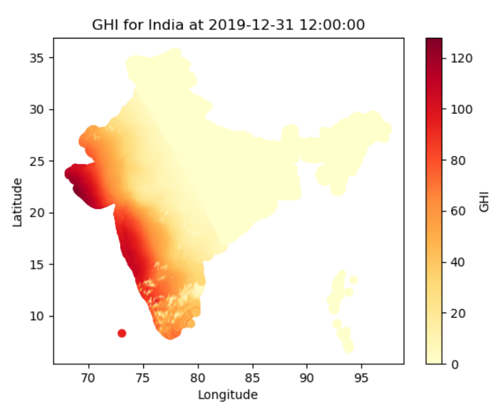


Fig (9): GHI for India

The analysis of the Global Horizontal Irradiance (GHI) data, sourced from the National Solar Radiation Database (NSRDB), provided insightful results for Indonesia, China, India, and South Africa. The NSRDB's dataset, specifically the METEOSAT IODC Region: Physical Solar Model (PSM), offers detailed solar radiation values with a 30-minute resolution, averaged over surface cells approximately 4 km² in size.

Results Part 2: NoSQL Database Implementation- World Bank Data

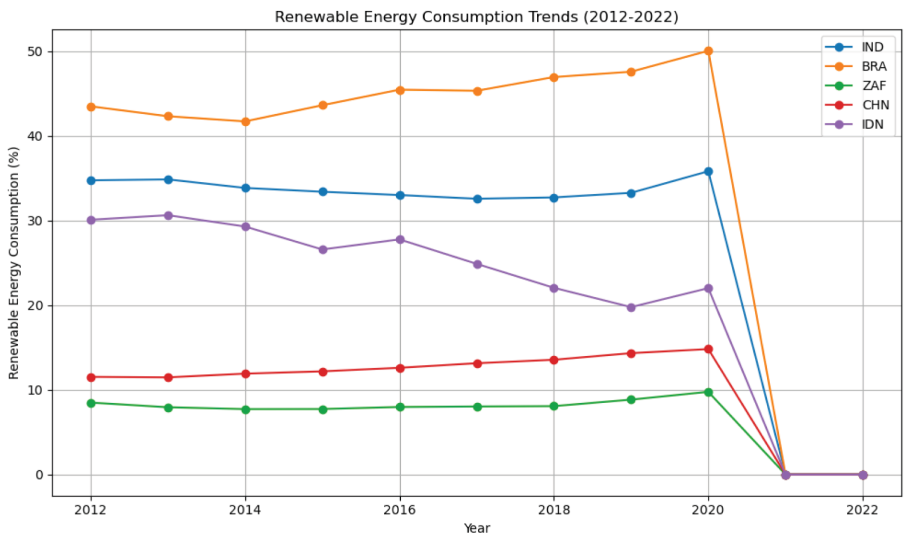


Fig (10): Renewable Energy Consumption Trends (2012-2022)

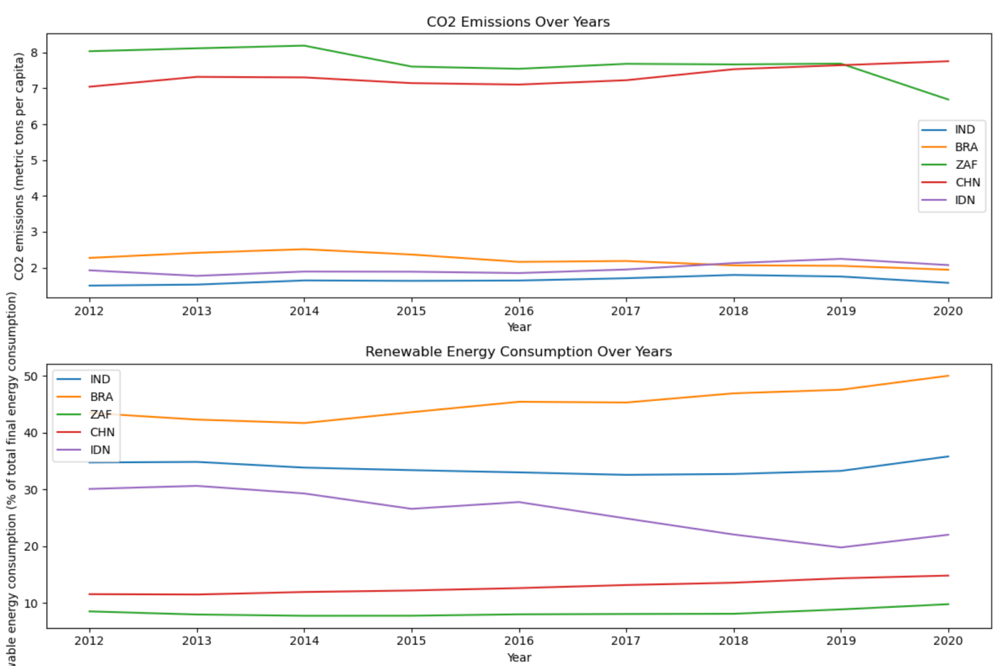


Fig (11): CO2 Emissions and Renewable Energy Consumption Trends (2012-2020)

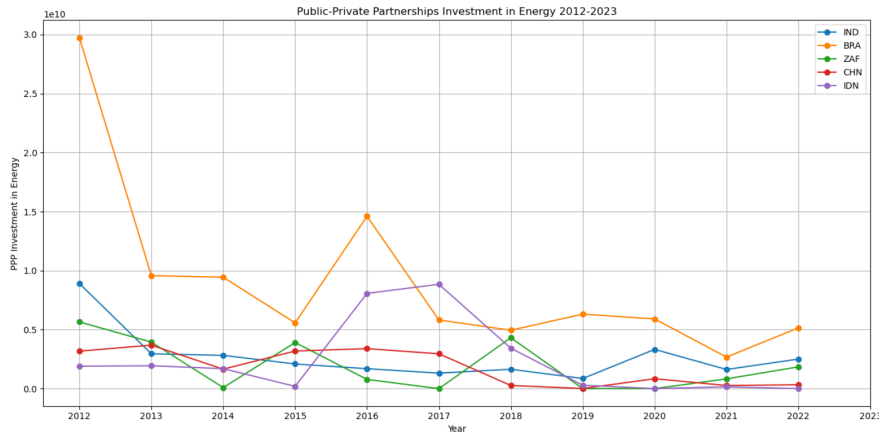


Fig (12): Public-Private Partnerships Investment in Energy (2012-2022)

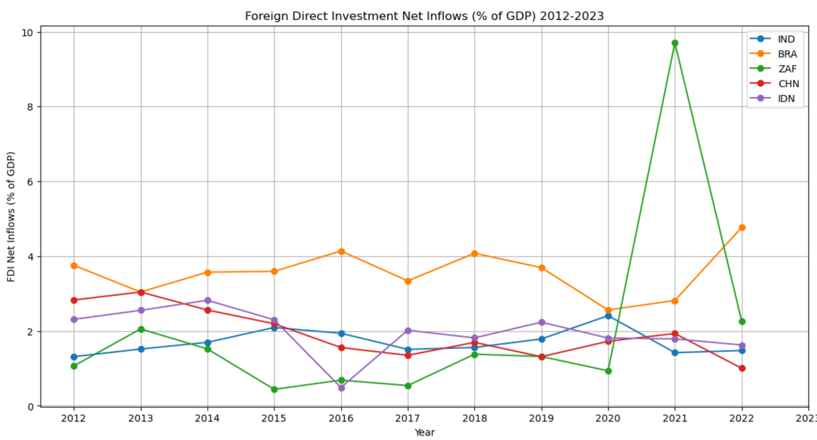


Fig (13): Foreign Direct Investment Net Inflows (% of GDP) (2012-2022)

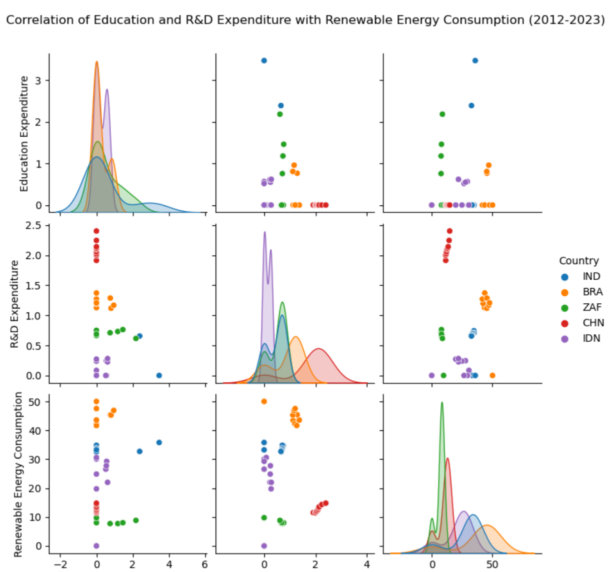


Fig (14): Correlation of Education and R&D Expenditure with Renewable Energy Consumption

The analysis of World Bank Open Data metrics using Google Bigtable provided a comprehensive overview of various socio-economic and environmental factors in selected emerging economies. The data, encompassing a range of indicators, was crucial in understanding the dynamics of renewable energy consumption and its interaction with other socio-economic variables.

Key Metrics Analyzed:

1. Renewable Energy Consumption: Assessed as a percentage of total final energy consumption, this metric provided insights into the extent of renewable energy integration in national energy mixes.

2. CO2 Emissions: Analysis of carbon dioxide emissions per capita helped in understanding the environmental impact of energy consumption patterns.

3. Electricity Production from Renewable Sources: This metric, excluding hydroelectric power, highlighted the contributions of alternative renewable sources like solar, wind, and biomass in electricity generation.

4. Foreign Direct Investment (FDI) in Energy: The role of FDI, as a percentage of GDP, in shaping energy infrastructure and renewable energy projects was examined.

5. Public-Private Partnerships in Energy: The investment figures in current US dollars offered insights into collaborative efforts between the public and private sectors in energy projects.

6. Urban Population Growth: The annual percentage growth of urban populations was analyzed to understand the increasing energy demands in urban settings.

7. Research and Development Expenditure: Investigated as a percentage of GDP, this metric provided a view into the commitment towards innovation in energy technologies.

8. Government Expenditure on Education: This analysis linked the total percentage of GDP spent on education to the potential for developing human capital in the energy sector.

9. Fossil Fuel Energy Consumption: Examining this as a percentage of total energy consumption helped in understanding the dependency on traditional energy sources.

10. Educational Attainment: The percentage of the population aged 25 and above with at least a Master's degree or equivalent offered insights into the educational landscape and its potential influence on energy sector developments.

Five distinct queries were executed in a Jupyter notebook to delve deeper into these metrics. The multifaceted analysis, supported by Google Bigtable's robust data handling capabilities, underscored the complex interplay between economic growth, environmental sustainability, and social factors in shaping the energy landscape of emerging economies.

Discussion

Results Discussion: Part 1

The Global Horizontal Irradiance (GHI) data analysis for the last day of 2019 at 12:00:00, sourced from the National Solar Radiation Database (NSRDB), revealed distinct solar energy potentials for India, Indonesia, China, and South Africa. The GHI, indicative of the available solar energy resource, was estimated using models like the AVHRR Pathfinder Atmospheres-Extended (PATMOS-x) and the Fast All-sky Radiation Model for Solar applications (FARMS), considering factors such as cloud properties, aerosol optical depth (AOD), and precipitable water vapor (PWV).

Notably, the analysis for Indonesia and China (Fig (7-8)) showed GHI values of 0, which could be attributed to specific weather conditions like heavy cloud cover or technical anomalies in data recording at the selected time. This underscores the importance of considering temporal variability and potential data limitations in solar energy assessments.

In contrast, India and South Africa (Fig (6) & Fig (9)) exhibited measurable GHI values, reflecting their solar energy potential under the prevailing climatic conditions at the time of data capture. These variations highlight the diverse solar radiation climates across these regions and emphasize the need for region-specific solar energy strategies.

This analysis, facilitated by the NSRDB's detailed data and advanced computational models, is vital for understanding the renewable energy landscapes in these countries, especially for solar energy planning and optimization.

Results Discussion Part 2:

Figure (10) illustrates the trends in renewable energy consumption as a percentage of total energy consumption for the five countries over a decade from 2012 to 2022, with results until 2020. India, Brazil, South Africa and China have an overall increase in renewable energy consumption, with Brazil leading at over 40%, and South Africa trailing at under 10%. Indonesia on the other hand has seen an overall decrease in renewable consumption. This figure emphasizes the varying commitment and transition towards renewable energy across different nations.

Figure (11) compares CO_2 emissions per capita with renewable energy use from 2012 to 2020 for five countries. Emissions are stable for India, Brazil, and Indonesia, while China shows a decrease and South Africa a notable decline. Renewable energy consumption rises steadily in India and China, whereas it remains more constant in Brazil, South Africa, and Indonesia. The data indicates a shift towards renewables in some countries, as evidenced by stable or falling emissions alongside rising renewable usage, highlighting the influence of green policies.

Figure (12) displays the trends in Public-Private Partnerships (PPP) investment in energy across the five countries from 2012 to 2023. The graph shows investment variability, with notable peaks and troughs indicative of economic cycles and policy shifts within these countries. Brazil's investment spikes around 2016, suggesting a period of significant engagement in energy projects between the public and private sectors, potentially including renewable energy initiatives. Meanwhile, China's investments peak dramatically in 2021, which may signal a substantial push towards developing its renewable energy infrastructure in alignment with global sustainability goals.

Figure (13) illustrates the FDI net inflows as a percentage of GDP for the five countries from 2012 to 2023. The trends show variability, with South Africa displaying a sharp peak in 2021, possibly indicating targeted investments that year, potentially in renewable sectors given global trends. The stability in FDI for India, Brazil, and Indonesia suggests consistent investment climates, which may correlate with steady developments in renewable energy infrastructure. China's varying FDI could reflect its evolving renewable energy policies and market conditions.

Figure (14) explores the relationship between education and R&D spending with renewable energy consumption from 2012-2023. Education spending shows a mixed correlation with renewable usage, while R&D investment's impact is less clear. A direct comparison of renewable consumption among the countries suggests varying energy adoption levels, possibly reflecting the strength of their renewable energy sectors and supporting policies.

Discussion of Course Skills/Technology:

Throughout the course of this project, I utilized a comprehensive set of technologies and skills from the curriculum to analyze the impact of socio-economic factors on renewable energy consumption in emerging economies. Google Cloud Platform (GCP) served as the foundation for the project, where I employed virtual machines for robust data processing and Google Cloud Storage to manage the data lifecycle. Data and code can be found int the Github repository for this project with a detailed readme and metadata.

I established a data pipeline that encompassed downloading, transforming, summarizing, and visualizing the datasets. By implementing Google Bigtable, a NoSQL database, I was able to handle the distributed storage and processing of large-scale data efficiently, showcasing the practical application of cloud computing principles learned in the course.

A pivotal aspect of the project was ensuring data integrity, for which I applied data cleaning to the datasets. The project faced several challenges, particularly in navigating the intricacies of the NoSQL database. These obstacles provided a valuable learning experience in system design and the necessity for modularity in computing.

In conclusion, this project was a testament to the skills gained from the course, emphasizing the importance of a structured approach to managing and analyzing big data within a cloud environment and the practical challenges encountered in real-world data analysis scenarios.

Conclusion

The project "Impact Analysis of Socio-economic Factors on Renewable Energy Consumption & Solar Radiation Potential in Emerging Economies" successfully harnessed cloud computing, data analytics, and storage models to explore the nexus between socio-economic indicators and renewable energy trends. Using GCP, Bigtable, and an established data pipeline, the study dissected a decade of data, revealing critical insights into renewable energy adoption across five key economies. Despite challenges like interpreting zeroed GHI values or navigating NoSQL database complexities, the project's findings offer a valuable contribution to understanding how investment in education, R&D, and public-private partnerships influence renewable energy consumption. This research underscores the diverse paths countries take towards sustainability, shaped by unique socio-economic fabrics and policy landscapes. It demonstrates the intricate dance between development and sustainability, providing a template for data-driven policy formulation and a beacon for future research in the realm of green energy transition.

References

IEA. (2021). Financing clean energy transitions in emerging and developing economies. Paris: International Energy Agency. Retrieved from https://www.iea.org/reports/financing-clean-energy-transitions-in-emerging-and-developing-economies. License: CC BY 4.0.National solar radiation database. Retrieved from https://nsrdb.nrel.gov/

Salim, R. A., & Rafiq, S. (2012). Why do some emerging economies proactively accelerate the adoption of renewable energy? Energy Economics, 34(4), 1051-1057. https://doi.org/10.1016/j.eneco.2011.08.015

Sart, G., Bayar, Y., Sezgin, F. H., & Danilina, M. (2022). Impact of Educational Attainment on Renewable Energy Use: Evidence from Emerging Market Economies. Energies (19961073), 15(7), 2695–N.PAG. https://doi-org.proxyiub.uits.iu.edu/10.3390/en15072695

World development indicators. Washington, D.C.: The World Bank. Retrieved from https://data.worldbank.org/

Xu, Q., Dhaundiyal, S., & Guan, C. (2020). Structural conflict under the new green dilemma: Inequalities in development of renewable energy for emerging economies. Journal of Environmental Management, 273, N.PAG. https://doi-org.proxyiub.uits.iu.edu/10.1016/j.jenvman.2020.111117