Analysis of Solar Energy Potential and Socioeconomic-Environmental Benefits Across Geographic Regions

Table of Content

CHAPTER 1: BUSINESS INTRODUCTION

1.1 Project Overview

With the ongoing shift towards sustainable energy, the need for intelligent and effective solutions for solar energy adoption has never been more pressing. That is why this project proposes an integrated solar energy data analysis and management system that delivers insights to inform policymakers, businesses, and homeowners about the viability of solar panel installations.

This project aims to process and analyze massive amounts of data related to geographical locations, climate zones, building characteristics, sunlight exposure levels, electricity costs, and solar installation costs. The conclusions derived from this data will determine whether solar energy is feasible, financially viable, or beneficial. The insights provided to decision-makers will effectively optimize energy strategies, give businesses a heads-up on financial returns, and empower homeowners to embrace renewable energy to save on electricity costs.

1.2 Project Objectives

The core objective of the project, established within the prepared work framework, is to build up comprehensive analytics that will facilitate understanding and enhance the further adoption of solar energy. In particular, the project has been structured around the following major objectives:

1.2.1 Evaluating Solar Potential

The system shall analyze sunlight exposure and roof suitability across different regions to obtain an accurate assessment of a given building's potential for solar energy generation. This would help users figure out whether the installation of a solar panel is a feasible option for their given location.

1.2.2 Optimizing Energy Savings

By comparing conventional electricity costs with the advantages related to solar power, the system will estimate potential savings. Therefore, users will visualize the reduced costs and long-term returns on investment, leading to well-informed decisions about their solar adoption.

1.2.3 Analyzing Costs and Incentives

The primary hurdle regarding solar adoption concerns the large upfront investment required. Given these financial concerns, the project shall tackle installation cost-related issues, available tax credits, and government incentives. These insights shall also enlighten homeowners and businesses by providing a clearer picture of the actual costs involved in moving toward solar energy.

1.2.4 Supporting Policy and Decision-Making

Through comprehensive data-driven insights, policymakers, businesses, and energy consultants will be better equipped to devise strategies for promoting solar energy. By providing the necessary tools and analytics, this project aims to best support any other renewable energy initiative in the future alongside large-scale solar power solutions.

1.3 Tools Used

In order to facilitate the extraction, processing, and visualization of solar energy data into meaningful insights for decision-making, this project utilizes a suite of powerful tools and technologies. These tools are designed to ease data management, improve analytical capabilities, and provide valuable insights.

1.3.1 Google Cloud & BigQuery

Google Cloud and BigQuery serve as the core tools in this project. They enable the retrieval and downloading of large-scale solar energy data using Google's "public dataset" repository. Their combined cloud-based architecture allows for fast querying, enhanced scalability, and seamless integration with other data-processing platforms.

1.3.2 MySQL for Data Manipulation & Exploration

After extracting this information, MySQL is used to structure, manipulate, and optimize the dataset. Beyond merely storing information, MySQL plays a critical role in uncovering hidden patterns, filtering key metrics, and preparing the data for in-depth analysis. By leveraging complex queries, indexing strategies, and stored procedures, the project maximizes data efficiency and ensures effective extraction of insights from vast information sources.

1.3.3 Power BI for Data Visualization & Comparison

The visualization of information is key to making large datasets approachable and useful. Power BI is used to develop interactive dashboards that assist key stakeholders in comparing datasets, identifying trends, and conducting more effective analyses. Users can explore various scenarios through dynamic filters, enhancing their ability to derive actionable insights.

1.3.4 Tableau for Spatial Analysis & Heatmaps

Tableau provides the opportunity for advanced spatial analysis of the distribution of solar installations and CO_2 reduction, where heatmaps developed in the software allow for a clear discernment of solar energy adoption trends across U.S. states. This helps identify areas with the highest adoption rates of renewable energy solutions and their associated environmental benefits. Such forms of data visualization serve as actionable intelligence for smart planning, particularly for decision-makers and businesses.

1.3.5 ChatGPT (OpenAI) for Code Optimization & Workflow Enhancement

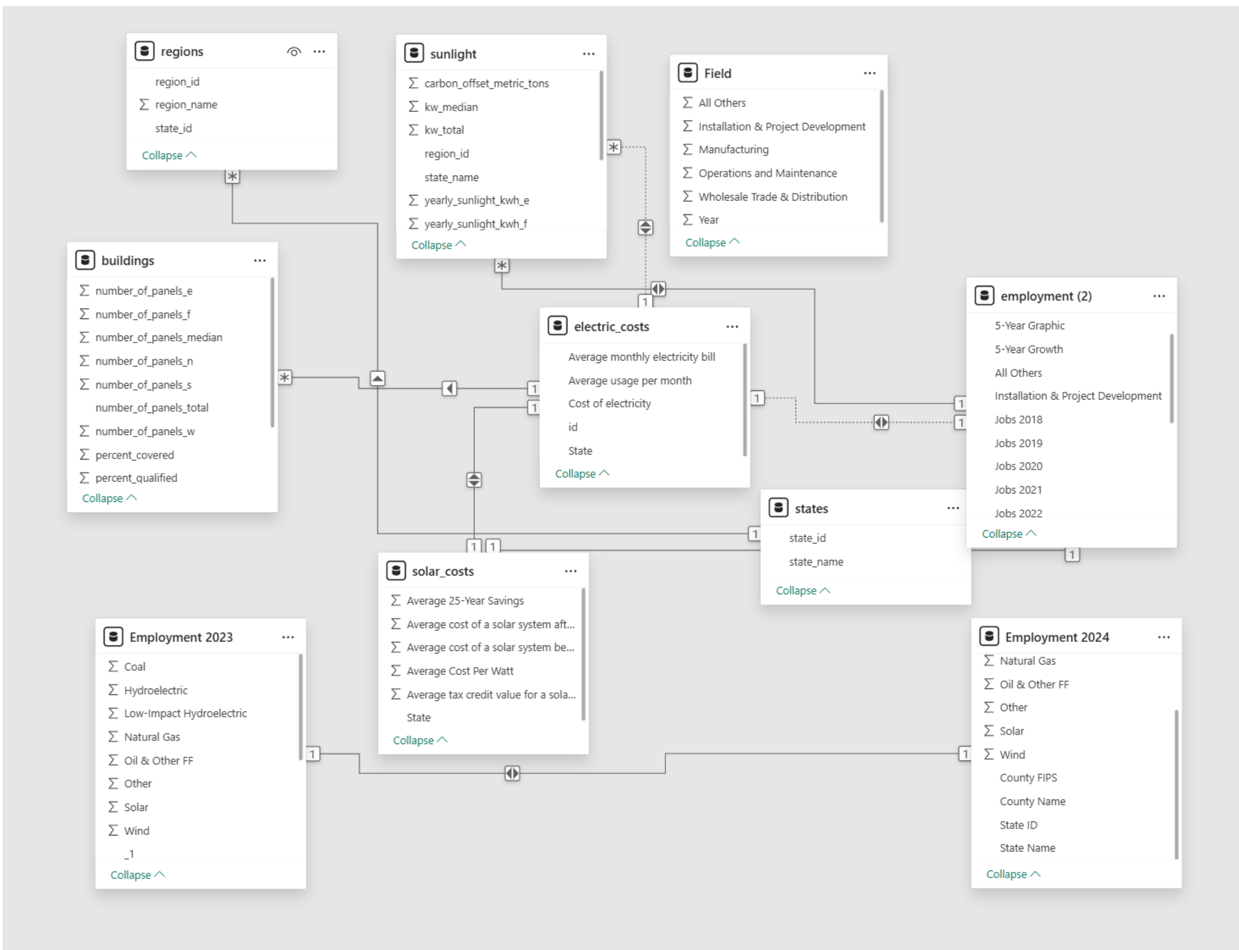
ChatGPT (OpenAI) can be used for code optimization and workflow improvement through coding-related queries for better performance. When embedded into this project, ChatGPT refines complex SQL scripts, automates repetitive tasks, and helps maintain balance in a data pipeline for exploratory analysis.

By integrating all these advanced tools, this project makes solar energy datasets more accessible and comprehensible for analysis, thus supporting better decision-making and promoting the broader adoption of renewable energy solutions.

CHAPTER 2: DATABASE MODEL

2.1 Conceptual and Logical Models (ERD)

Data collection, analysis, and interpretation skills toward driving decisions are core competencies in the solar energy consultation sector. The conceptual data model will hence serve as the gist for the management and integration of the many data sources throughout the solar industry, aiding the smooth operations of any kind regarding solar energy, infrastructure, geographical distribution, and employment trends, among many others. Using a strong entity-relationship model (ERD), companies will be able to garner such insights that aid county procurement planning and increase the efficiency of solar panel installations.



2.1.1 Geographical and Economic Aspects of Solar Energy

Highly prominent in the conceptual data model would be the geographical and economic factors influencing solar adoption. The States entity describes the various locales where solar energy projects are being constructed, and further divisions of these locations into administrative units are represented by the Regions entity. This hierarchy offers more granularity in assessing solar energy potential at either the state or regional level by different stakeholders.

Besides geographic classification, some economic factors also inspire the decision to invest in solar. The Electric Costs entity provides key metrics on figures such as average electricity bills and cost per unit of electricity in order to measure how affordable traditional investment is in comparison to solar energy. The Solar Costs entity deals with major construction costs, installation costs, cost per watt, tax credits available for installation, and anticipated long-term savings.

This kind of financial information provides a basis for estimating the ROI on the installation of solar panels. Most incentives for consumers to install solar energy systems are based on these financial perspectives.

2.1.2 Solar Infrastructure and Energy Production

The works of the Buildings entity comprise their specific collections, such as orientation, distribution of panels, and the total area covered where solar panels are mounted. Further data supports businesses in optimizing the placement of panels for maximum energy production while sustaining wildlife.

Sunlight Availability is also a significant entity, recording solar exposure information, including total energy potential from sunlight, carbon offset benefits, and the timespan of kilowatts generated. It is critical to understand the amount of sunlight different regions receive, as this helps determine whether they are suitable for solar panel installation. The most suitable locations for large-scale solar projects have high solar exposure, while areas with lower exposure may require additional infrastructure to align with efficiency standards.

2.1.3 Workforce and Industry Dynamics

Growing demand for labor as the solar energy sector expands has led to the Employment entity that the model ought to use to ascertain acceptance in different trend directions, whether in terms of the job market concerning an industry, the concentration of employment across field dimensions in solar technology, or evaluating past hiring patterns.

Policymakers and industry figures would now be able to measure and establish a variety of policies to develop and invest in the training of potential human resources in the energy sector. This analysis of workforce data is projected to assist in making better investments related to skilled labor training programs in the future.

In parallel, the Field class lists different job titles and industrial activities in solar energy, with specializations such as manufacturing, installation, maintenance, and distribution. Analyzing employment distribution across these sectors is important for allocating labor resources more productively and ensuring the industry's sustainability. Additionally, such an analysis of employment trends could provide a gauge for labor demand in the long term and inform investments in training programs for tomorrow's workforce.

2.1.4 Data Integration and Strategic Decision-Making

The relationships among these entities are configured to facilitate data integration and the possibility of comprehensive analysis. The tie between geographic data and financial metrics is meant to enable localized insights into the use of solar energy. When infrastructure details are linked with energy production data, stakeholders will be able to assess how well solar panels are being utilized and what could be improved.

Aligning employment data with industry specializations ensures a timely response to workforce planning in view of the changing demands of the solar industry sector. This more integrated approach to data allows different stakeholders, such as government agencies, investors, and energy consultants, to make strategic decisions on solar energy expansion.

While this model enables data-driven planning that helps businesses consider the economic viability of solar projects, optimize energy production, and respond to labor market challenges, the conceptual data model would act as an architectural framework for promoting solar energy adoption and achieving a sustainable energy future.

2.2 Data table description

2.2.1 State table

Column	Data Type	Description
state_id (PK)	INT	Unique identifier for a state
state_name	VARCHAR	Name of the state

The State table is a central constituent within the database, designed to furnish a cohesive framework for geographical data. It provides a standard reference to assist in the identification of states throughout the entire dataset; hence, it fosters consistency in geographical classification into regions.

This table facilitates quick data retrieval, as each state has a unique code by which integration occurs with other entities such as divisions, statistics on energy usage, and foundations for the deployment of solar infrastructure. This table is important for the analysis of geographic trends in solar energy adoption because it enables location-based comparisons of electricity costs, solar potential, and installation metrics.

The ability to couple state-level data with financial and environmental indicators allows stakeholders to make informed decisions on where solar energy investments should be prioritized. Furthermore, the State table serves as a central reference point for various analytical processes, linking all location-based insights directly back to their respective regions.

2.2.2 Region table

Column	Data Type	Description
region_id (PK)	INT	Unique identifier của vùng
region_name	VARCHAR	Region code (multiple regions within a state)
state_id (FK)	INT	Links to States(state_id)

The Region table provides a more nuanced perspective on geographic divisions in each state, ensuring a detailed analysis of solar energy adoption at a sub-state level. Since one state might comprise different regions with varying solar potential, energy consumption patterns, and economic circumstances, this table ensures that regional differences are effectively captured in the dataset.

It creates a hierarchical relationship from the state to the region level, allowing for a more precise mapping of solar energy infrastructure. This facilitates targeted analysis, such as identifying which regions within a state are more conducive to solar panel installation or have more favorable economic incentives.

In addition, this table plays a key role in localized decision-making by allowing policymakers and businesses to tailor strategies to support solar advancement in specific areas. Adding this linkage supplements the State linkage, neatly connecting back to the broader context of state and regional comparisons. This structure also enhances the accuracy of forecasting models, helping stakeholders assess solar investment opportunities across different geographic scales.

2.2.3 Building table

Column	Data Type	Description
building_id (PK)	INT	Unique identifier for a building
region_id (FK)	INT	Links to Regions(region_id)
percent_covered	FLOAT	Percentage of the roof covered by solar panels
percent_qualified	FLOAT	Percentage of the roof that is eligible for solar panel installation
number_of_panels_n	INT	Number of panels facing North
number_of_panels_s	INT	Number of panels facing South
number_of_panels_e	INT	Number of panels facing East
number_of_panels_w	INT	Number of panels facing West
number_of_panels_f	INT	Number of flat panels
number_of_panels_total	INT	Total number of solar panels installed

The Building table sets a foundation for analyzing solar energy adoption and feasibility on a building scale. Each entry in the table refers to an individual building in a certain region, linking solar panel installation to geographic and architectural features. Additionally, the table allows for accurate analysis of solar energy potential according to different types of buildings due to the level of detail it captures.

Information on roof coverage with solar panels is one of the essential attributes recorded in the Building table. It reflects the extent to which a given building has utilized its roof space for solar energy generation. This table also contains the percentage of the roof that is eligible for solar installation, ensuring that feasibility studies take into account structural hindrances and shading factors that challenge the efficiency of solar deployment.

To lend further granularity to the analysis of solar panel distribution, this table indicates whether the solar panels are set on rooftops facing North, South, East, or West. Since solar panel orientation greatly influences energy generation efficiency, this dataset becomes instrumental in performance evaluation and optimization strategies. The total number of installed panels is also recorded, providing insight into the scale of solar adoption per building.

By linking buildings to their respective regions, this tabular profile allows for a comprehensive spatial analysis of solar energy implementation, enabling policymakers to pinpoint areas with accelerated solar uptake, examine regional inequities, and derive optimal installation strategies moving forward. This structured approach will help advance specific incentives and policies aimed at fostering a broader base for renewable energy infrastructure.

2.2.4 Sunlight table

Column	Data Type	Description
region_id (PK, FK)	INT	Links to Regions(region_id)
count_qualified	INT	Number of buildings eligible for solar panel installation
kw_total	FLOAT	Total kilowatt capacity of the region
yearly_sunlight_kwh_n	FLOAT	Annual solar energy potential for North-facing roofs
yearly_sunlight_kwh_s	FLOAT	Annual solar energy potential for South-facing roofs
yearly_sunlight_kwh_e	FLOAT	Annual solar energy potential for East-facing roofs
yearly_sunlight_kwh_w	FLOAT	Annual solar energy potential for West-facing roofs

yearly_sunlight_kwh_f	FLOAT	Annual solar energy potential for flat roofs
yearly_sunlight_kwh_total	FLOAT	Total annual solar energy potential for the region
carbon_offset_metric_tons	FLOAT	Estimated CO_2 reduction in metric tons

Importantly, the Sunlight table is a fundamental dataset that provides insight into solar energy potential across various areas or regions. It not only clusters a wealth of data about the total capacity of solar energy available in a given area but also provides significant evidence on how sunlight exposure variations depend on builders' rooftop orientations to the sun. Utilizing the Regions table allows this data to implicitly surface, facilitating decisions on planning and developing infrastructure for renewable energy.

The table essentially provides a quantified insight into the kilowatt capacity of a given area by summing up the qualified structures and the respective solar panels installed in that area. In essence, count_qualified represents the number of buildings or units in an area where the structural integrity allows for solar panel installations. This is critical for policymakers and utility planners when considering large-scale solar initiatives or designing incentive programs to promote and encourage the adoption of solar energy.

The table provides the annual potential for solar energy generation based on roof orientation: North, South, East, West, and flat roofs. According to solar energy principles, the performance of solar panels is highly dependent on orientation—roofs mounted in different directions receive varying sunlight throughout the day and year, leading to higher or lower electricity generation. Data segmented by orientation facilitates a deeper analysis of energy generation capacity, allowing for the refinement of installation strategies and optimization of solar energy output. Similarly, the total potential for solar energy generation across an entire region is summed to reflect the overall renewable energy potential on a macro level.

Beyond that, the Sunlight table includes an important additional measure: estimates of carbon offset in terms of tons of CO_2 reduction. This represents the actual decrease in CO_2 emissions resulting from the use of solar energy in a particular area, both theoretically and practically. In other words, this dataset indirectly aids in renewable energy planning and supports climate change initiatives by providing a comprehensive sustainability analysis.

2.2.5 Electric Cost table

Column	Data Type	Description
state_name (PK)	INT	Links to States(state_id)
cost_of_electricity	FLOAT	Average electricity price (USD/kWh)
average_usage_per_month	FLOAT	Average monthly electricity consumption (kWh)
average_monthly_electricity_bill	FLOAT	Average monthly electricity bill (USD)

The Electric Cost table contains a vital collection of data related to the consumption and pricing of electricity across states. By linking directly to the States table, it allows geographic comparisons of energy costs and usage patterns, which are vital in assessing the feasibility of initiating solar energy and other energy-saving efforts. Understanding electricity costs remains one of the most fundamental factors in determining the return on investment (ROI) for installing solar panels, making this table seminal to the broader energy analysis framework.

One of the key metrics included in this table is cost_of_electricity, which refers to the average price in USD per kWh for each state over the years. This metric is highly influential in assessing the economic benefits of solar energy because areas with higher electricity prices see increased financial payoffs from implementing solar solutions. By comparing these values across different states, knowledge-based decision-making will enable policymakers and consumers to effectively determine where renewable energy solutions can have the greatest financial impact.

Another important data point in the Electric Cost table is average_usage_per_month, which quantifies electricity consumption per household or business each month in kWh. It provides demand assumptions and serves as a basis for calculating energy savings if solar panels are installed. The combination of this column with tariff data allows for accurate calculations of current energy expenditures and estimates of potential savings through renewable energy implementation.

2.2.6 Solar Cost table

Column	Data Type	Description
state_id (PK, FK)	INT	Links to States(state_id)
avg_cost_before_incentives	FLOAT	Average cost of a solar system before incentives
avg_cost_after_incentives	FLOAT	Average cost of a solar system after incentives
avg_tax_credit_value	FLOAT	Average tax credit value applicable to solar systems
avg_cost_per_watt	FLOAT	Average cost per watt of solar energy
avg_25_year_savings	FLOAT	Estimated savings over 25 years from solar energy adoption

The **Solar Cost** table is a specialized tool that helps in understanding the economic feasibility of different states for solar energy adoption. This dataset captures major cost components of solar installation systems, including the average system price before and after applying incentives, tax credits, and the cost per watt of solar power. Besides these, it provides qualitative long-term estimates of savings, adding to the **assured** value of decision-making when establishing policies for economic value in solar energy.

One main feature of this dataset is its ability to show the incentives applied by the government and tax credits that affect solar affordability. By comparing costs before and after incentives, it provides a clear view of how much financial relief is available to consumers and how policies dictate the relative affordability of solar energy. This is particularly useful for consumers, as they can analyze the general data to determine whether investing in solar panels will yield the desired long-term savings.

Beyond that, this table was created with easy regional comparisons in mind, linking to the **States** table so that users can see the differences in solar costs and benefits across regions. This is especially important for regional users, as solar price ranges, subsidies, and tax credits vary significantly based on local regulations, power utilities, state energy policies, and market conditions. This localized approach enhances decision-making.

2.2.7 Employment table

Column	Data Type	Description
state	VARCHAR	Name of the state
jobs_2018	INT	Number of solar jobs in 2018
jobs_2019	INT	Number of solar jobs in 2019
jobs_2020	INT	Number of solar jobs in 2020
jobs_2021	INT	Number of solar jobs in 2021
jobs_2022	INT	Number of solar jobs in 2022
jobs_2023	INT	Number of solar jobs in 2023

The **Employment** table is a great representation of the trends in job creation within the solar energy sector by state and time. Tracking employment statistics has never been more necessary than today, as the renewable energy industry continues to grow. Insights into how jobs have emerged and declined will greatly assist in understanding the economic impact of solar initiatives.

This dataset captures the number of solar jobs from 2018 to 2023, allowing a plethora of researchers, a variety of policymakers, and concerned industry stakeholders to analyze workforce trends, review job creation initiatives, and assess the effectiveness of government policies and incentives aimed at boosting the solar industry. This table permits year-on-year comparisons of employment figures, thereby facilitating the comparative analysis of labor market fluctuations within the solar sector. It additionally **enhances** stakeholder awareness about growth patterns, showing which states lead in job creation and calling attention to places that may require additional support or assistance.

For instance, states showing increasing employment figures might indicate high adoption of solar energy and the effectiveness of their respective policies, while states with declining trends may indicate challenges faced by the market, such as policy changes, supply chain disruptions, or evolving economic conditions. It serves as a **powerful** tool for workforce development. Educational institutions, training programs, and labor organizations can utilize the data to align curricula with industry demands, thereby providing graduates with immediate employability in solar installation, maintenance, and research roles.

Businesses involved in the renewable energy industry can take advantage of this information to craft strategies for expansion in promising markets, sharpen their recruitment efforts, or build the talent pool in readiness for future growth. By enabling year-on-year comparisons of employment growth, this dataset continues to contribute to a broader understanding of the labor market impacted by the transition to **sustainable energy** as it transforms the economy.

2.2.8 Field table

Column	Data Type	Description
year	INT	Year of recorded data
installation_project_development	INT	Jobs in installation & project development
manufacturing	INT	Jobs in solar manufacturing
wholesale_trade_distribution	INT	Jobs in wholesale trade & distribution
operations_maintenance	INT	Jobs in operations & maintenance
all_others	INT	Jobs in other related solar industries

The **Field** table provides an invaluable breakdown of jobs in various sectors within the solar energy industry. The growth of the solar sector will create jobs in the fields of installation, manufacturing, operations, and wholesale trade. Captured here are jobs organized along specific professional lines, allowing one to gauge labor trends within different fields. The industry will track these workforce distributions over the years; hence, such trend data will allow policymakers, businesses, and workforce planners to infer changes in industry growth, analyze which occupations are facing skill shortages, and tailor training programs accordingly to meet the changing demands of the labor market.

Each of these categories represents a vital part of the solar energy value chain. Jobs in **installation and project development** relate to areas like system design, permitting, and on-site installation of solar panels. **Manufacturing** jobs represent those engaged in the production of solar panels, inverters, batteries, and other key components. **Wholesale trade and distribution** include employment in logistics and supply chain management to ensure efficient access to solar equipment for installers and consumers. **Operations and maintenance** jobs keep the installations running and optimized to ensure long-term efficiency and reliability. Meanwhile, jobs categorized as **"all others"** simply cover any supporting roles in the solar sector, such as research, consulting, and administrative functions.

Stakeholders in the industry will be able to use this data to see which areas of the market are gaining the most jobs and which might require further investment or policy support. Such data will also allow workforce development programs to align their training programs with market demands, ensuring that job seekers acquire the right skills. Businesses will use these insights to refine their hiring strategies and workforce planning, which will, in turn, contribute toward strengthening the solar sector's contribution to economic growth and sustainable energy development.

2.2.9 Employment 2023 table

Column	Data Type	Description
State Name	TEXT	Name of the state where employment is recorded
County Name	TEXT	Name of the county where employment is recorded
Solar	INT	Jobs in the solar energy sector

Wind	INT	Jobs in the wind energy sector
Hydroelectric	INT	Jobs in the hydroelectric energy sector
Low-Impact Hydroelectric	INT	Jobs in low-impact hydroelectric energy
Natural Gas	INT	Jobs in the natural gas sector
Coal	INT	Jobs in the coal energy sector
Oil & Other FF	INT	Jobs in the oil and other fossil fuel sector
Other	INT	Jobs in other energy-related sectors

The **Employment 2023** table provides a detailed snapshot of jobs across various energy sectors, offering insights into workforce distribution and industry trends. By tracking employment figures across renewable and non-renewable energy industries, this data helps policymakers, businesses, and workforce planners understand labor market dynamics, address skill shortages, and develop training programs that align with industry needs.

Jobs in renewable energy—like solar, wind, and hydroelectric—reflect the ongoing shift toward sustainable energy sources. These roles cover everything from installation and manufacturing to maintenance and logistics, keeping the clean energy infrastructure running. Low-impact hydroelectric jobs focus on environmentally friendly hydropower solutions, balancing energy efficiency with minimal environmental disruption.

The traditional energy industries remain large employers, with classic jobs in extraction, refining, transport, and power generation: oil, coal, and natural gas. For all the new energy policies and the sustainability imperative they introduce, the traditional energy industries remain a significant sector of the economy. Even employment trends within these industries track shifts in fossil fuel use and advances in energy technology.

The dataset also includes other energy jobs such as administrative, research, consulting, and emerging technology jobs that support the energy industry as a whole. These positions are instrumental to innovation, policy, and increasing operational efficiency within all sectors of energy.

By analyzing employment trends within these categories, stakeholders can make informed decisions regarding investments, workforce development, and regulatory policies. As the energy sector evolves, this data will serve as a valuable resource for ensuring a resilient and adaptive labor market that meets the demands of a changing energy landscape.

2.2.10 Employment 2024 table

Column	**Data Type**	**Description**
State Name	TEXT	Name of the state where employment is recorded
County Name	TEXT	Name of the county where employment is recorded
Solar	INT	Jobs in the solar energy sector
Wind	INT	Jobs in the wind energy sector
Hydroelectric	INT	Jobs in the hydroelectric energy sector
Low-Impact Hydroelectric	INT	Jobs in low-impact hydroelectric energy
Natural Gas	INT	Jobs in the natural gas sector
Coal	INT	Jobs in the coal energy sector
Oil & Other FF	INT	Jobs in the oil and other fossil fuel sector

Other	INT	Jobs in other energy-related sectors

The **Employment 2024**

Table provides a realistic picture of the way occupations are distributed across different energy industries and offers valuable information on employment trends and industry development. By segregating employment figures into renewable and non-renewable sources, this information turns out to be a valuable support for policymakers, industry analysts, and workforce planners who must understand the requirements of the labor market, forecast industry development, and pre-emptively address skills gaps.

Renewable energy industries, including solar, wind, and hydroelectric power, point to the continuing transition toward greener sources of energy. Employment in these categories runs from installation and engineering to maintenance and manufacturing, all serving a role in constructing and maintaining renewable energy systems. Low-impact hydroelectric employment involves environmentally sound hydropower solutions that reconcile energy generation with sustainable ecology.

Traditional fossil fuel industries, like coal, oil, and natural gas, continue to employ a majority of the workforce. Operations such as extraction, refining, distribution, and plant operations form the cornerstones of current energy supply chains. Despite rising regulatory and economic challenges, they are a necessary part of the energy sector.

In addition, the data also cover other energy professions, which include research, policy making, consulting, and administrative activities. These professions provide the capacity and planning and strategic thinking that drive innovation and efficiency across all energy industries.

By looking at job trends within these sectors, industry executives are in a better position to forecast labor market changes, formulate specific workforce training initiatives, and introduce policies that foster sustainable growth. With the energy industry ongoing evolution, this information will play a part in charting the future of work, ensuring the workforce is ready for new opportunities in the energy transition.

2.3 ETL framework

The ETL (Extract, Transform, Load) model is pivotal within this project because it is used to enable effective collection, processing, and consolidation of solar energy data for the sake of analysis. The described process has three significant steps: extraction, transformation, and loading

2.3.1 Extract – Data Collection from Public Datasets

Extraction is the process of gathering raw data from authentic sources to create a comprehensive database in preparation for analysis. Google Cloud and BigQuery serve as the primary data sources for this project, where public data on solar power, electricity prices, and environmental conditions is maintained. The datasets contain extensive information regarding solar panel installations, sun exposure, electricity prices, and state government incentives. The retrieved data includes various attributes such as solar capacity, carbon offset, installation cost, and financial incentives. Additionally, to enhance database reliability, supplementary data can be integrated from government reports, regulatory bodies, and research institutions on renewable energy.

2.3.2 Transform – Data Cleaning and Processing

Once raw data is extracted, it undergoes transformation to ensure consistency and usability. The data is first loaded into MySQL, then cleaned and structured through several essential steps. Filtering is performed to remove redundant or duplicate records, and missing data is handled to maintain data integrity. Standardization operations are applied to convert all measurements (such as electricity rates per kWh and solar capacity in kW) to a common standard. Furthermore, data normalization is implemented to organize information into relational tables, simplifying queries and facilitating meaningful insights. Advanced calculations, including median solar energy generation, long-term cost savings estimates, and carbon offset values, are also conducted to enhance analytical accuracy.

2.3.3 Load – Data Storage and Visualization

Lastly, transformed data is loaded into analytical and visualization tools for examination and decision-making. Power BI is utilized to create interactive dashboards facilitating stakeholders to compare the electricity bills, solar penetration levels, and financial savings across various regions. Tableau, on the contrary, is utilized to create heatmaps, indicating rates of solar panel installations and CO2 reduction levels within U.S. states. Both visualizations highlight clear, data-driven information that can be utilized in decision-making by policymakers, businesses, and homeowners. Further, ChatGPT (OpenAI) is utilized to maximize SQL queries and streamline data processing to increase the efficiency of the overall ETL process, with the aim of better accuracy and speed.

Following this systematic ETL process, the project guarantees all pertinent solar energy data is precisely gathered, effectively processed, and properly visualized, enabling data-driven decision-making and encouraging sustainable energy solution adoption.

CHAPTER 3: DATA ANALYSIS

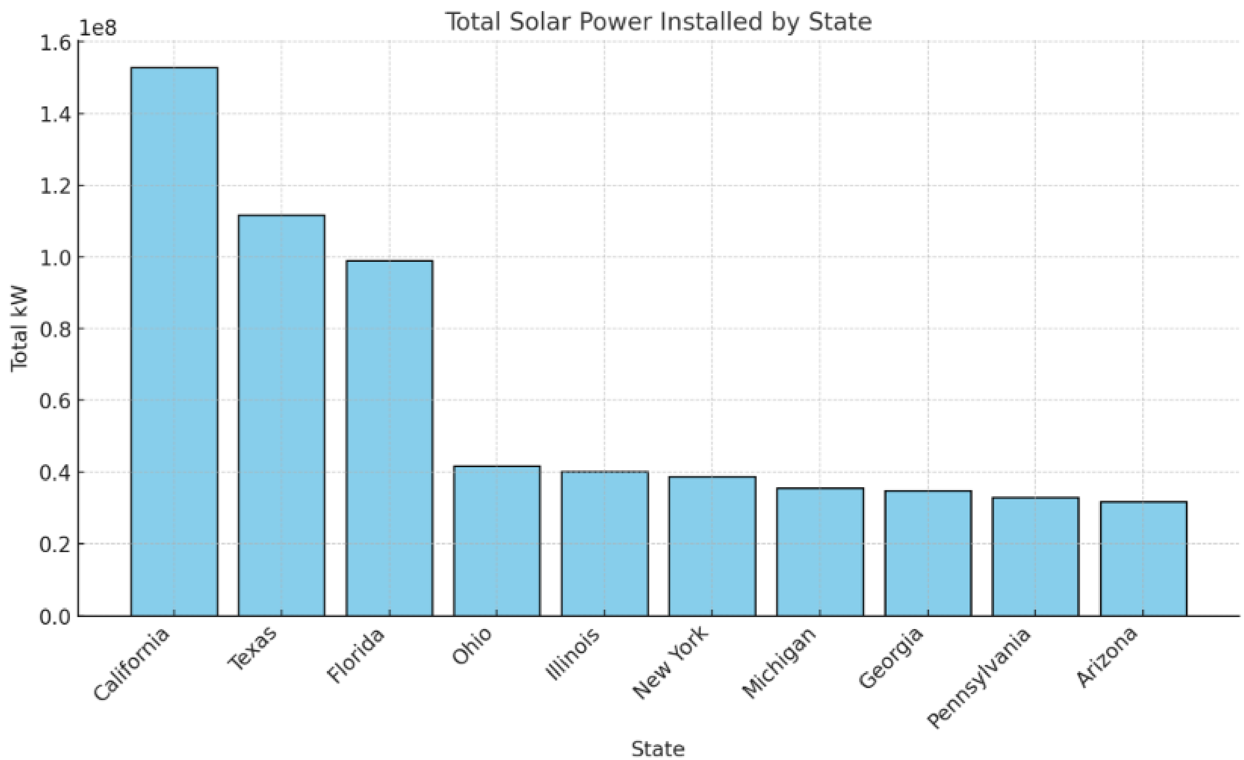
3.1 Solar energy analysis

There are several large factors that account for the greater solar energy output of California, Texas, and Florida compared to other states. For instance, California generates nearly four times the solar energy as Ohio, which is itself one of the top four solar-producing states. Similarly, Florida's solar output is more than double that of Ohio. These wide disparities indicate that there has to be more to the story than initially meets the eye.

```sql
SELECT
    state_name,
    SUM(kw_total) AS total_kw
FROM sunlight
GROUP BY state_name
ORDER BY total_kw DESC
LIMIT 10;
```

This SQL query gets the top 10 states producing the most overall solar energy by aggregating values from the sunlight table. It returns the state_name column and aggregates the overall solar energy generated with SUM(kw_total) per state such that the individual states' overall energy generation are summed separately. The result is then ordered in descending order based on the total energy produced with ORDER BY total_kw DESC such that the top states by solar energy are at the beginning. Finally, the LIMIT 10 clause returns only the first 10 states.

state_name	total_kw
California	152757921
Texas	111619682.25
Florida	98912340.25
Ohio	41647670.75
Illinois	40174559.75
New York	38681913.5
Michigan	35535521.25
Georgia	34595379.5
Pennsylvania	32937910.5
Arizona	31717976.5



The query sums the kw_total values from the sunlight table, groups the results by state_name, and orders them in descending order to display the top 10 states with the highest solar energy production. As shown in the image, California leads with the most solar energy produced, followed by Texas and Florida. This output provides insights into the states generating the most solar power, likely due to factors such as sunlight availability, solar infrastructure, and energy policies.

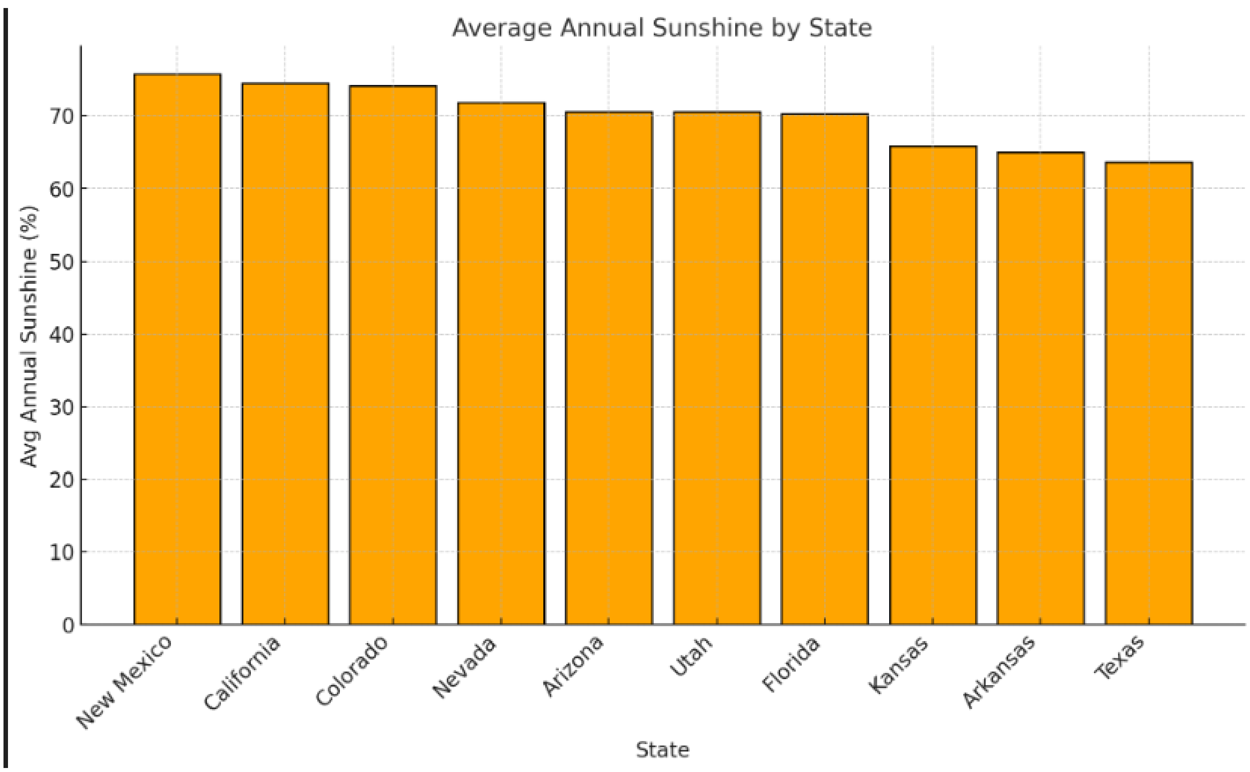
One key reason is that these three states rank among the top 10 in average annual sunshine, providing optimal conditions for solar energy generation. Here is the SQL code and the result:

```
SELECT
    `STATE`,
    ROUND(AVG(CAST(`ANN` AS DECIMAL(5,2))), 2) AS Avg_Annual_Sunshine
FROM `average percent of possible sunshine by us city`
GROUP BY `STATE`
ORDER BY Avg_Annual_Sunshine DESC
LIMIT 10;
```

This SQL query calculates the average annual sunshine percentage for each U.S. state using data from the table "average percent of possible sunshine by US city". It first converts the ANN (annual sunshine percentage) column into a decimal format, then computes the average sunshine percentage for each state using AVG(). The results are grouped by state, sorted in descending order by sunshine percentage.

Result Grid | Filter Rows:

STATE	Avg_Annual_Sunshine
New Mexico	75.75
California	74.50
Colorado	74.17
Nevada	71.86
Arizona	70.50
Utah	70.50
Florida	70.30
Kansas	65.88
Arkansas	65.00
Texas	63.63



Another contributing factor is that these three states have the highest number of jobs in the solar industry, enabling the installation of more solar panels and enhancing the overall productivity of each unit.

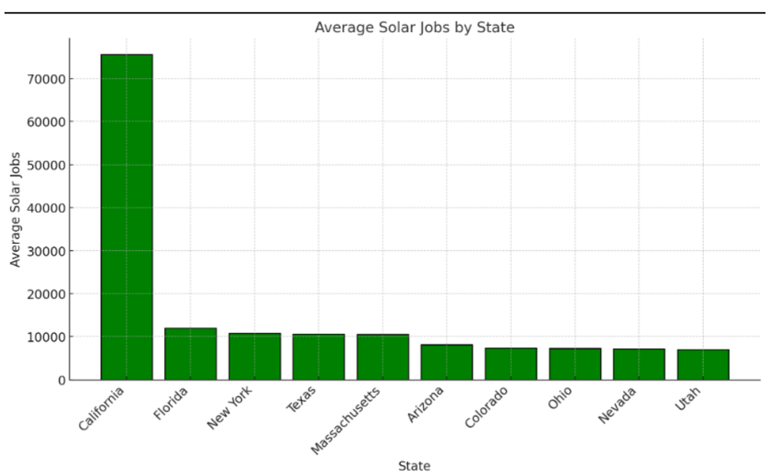
```
SELECT
    State,
    ( (`Jobs 2018` + `Jobs 2019` + `Jobs 2020` + `Jobs 2021` + `Jobs 2022` + `Jobs 2023`) / 6 ) AS Avg_Jobs
FROM employment
ORDER BY Avg_Jobs DESC
LIMIT 10;
```

This SQL query calculates the average number of jobs for each state from 2018 to 2023 using data from the employment table. It sums the job counts for each year from Jobs 2018 to Jobs 2023, divides the total by 6 (representing the six years), and assigns the result as Avg_Jobs. The query then orders the states by the highest average number of jobs in descending order and selects the top 10 states with the highest employment levels.

The result reveals that solar panel industry is well developed in these states

Result Grid | Filter Rows:

State	Avg_Jobs
California	75609.0000
Florida	11985.8333
New York	10805.5000
Texas	10663.0000
Massachusetts	10558.5000
Arizona	8164.6667
Colorado	7336.8333
Ohio	7276.8333
Nevada	7200.1667
Utah	6998.8333



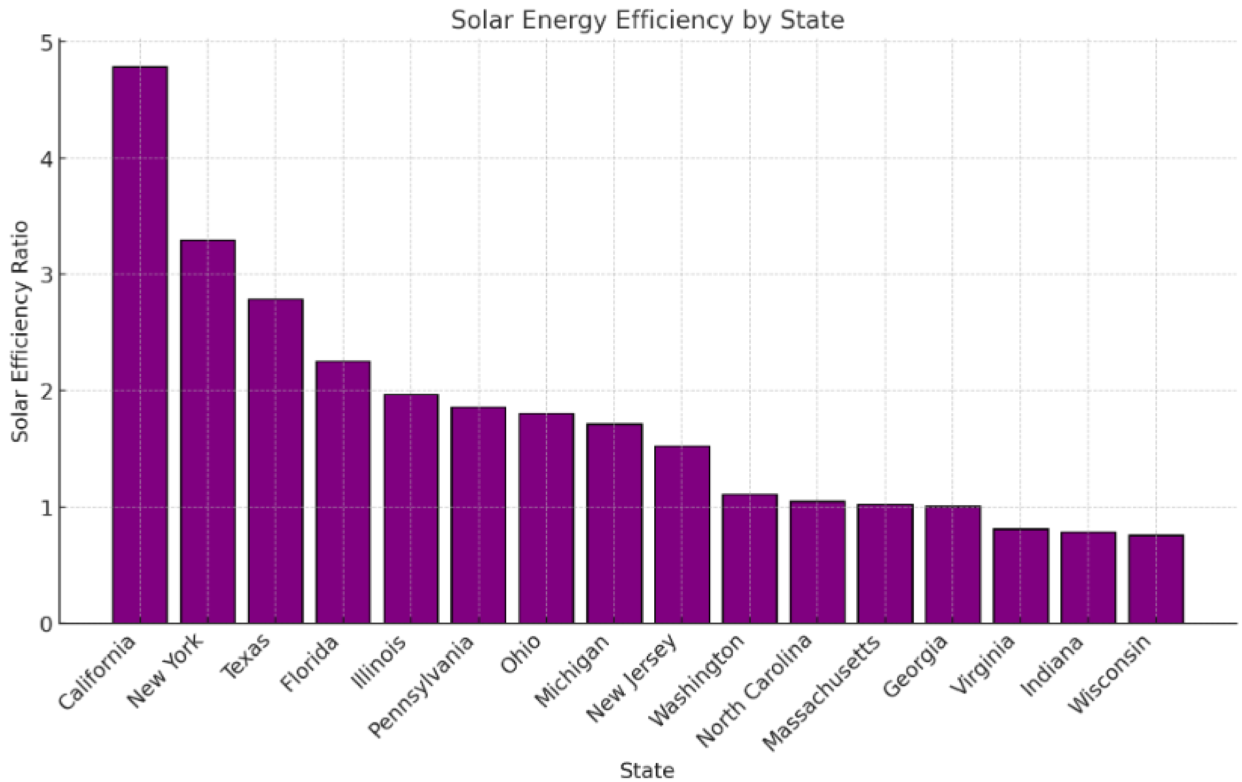
3.2 Correlation Between Sunlight Percentage and Solar Energy Production analysis

Solar energy production is influenced by multiple factors, with **sunlight availability** being one of the most critical. This analysis explores the correlation between **annual sunlight percentage** and **total solar energy production** in different states. By joining a dataset containing sunlight percentage data with another containing solar energy output, the SQL query calculates and ranks states based on their solar energy efficiency. Understanding this relationship helps determine whether states with more sunlight naturally produce more solar power or if other factors, such as infrastructure and policies, play a significant role.

```sql
SELECT
    state_name,
    ROUND(AVG(yearly_sunlight_kwh_total), 2) AS avg_sunlight_kwh,
    ROUND(SUM(kw_total), 2) AS total_solar_energy_produced,
    ROUND(SUM(kw_total) / AVG(yearly_sunlight_kwh_total), 4) AS efficiency_ratio
FROM sunlight
GROUP BY state_name
ORDER BY efficiency_ratio DESC;
```

The SQL query analyzes the correlation between the percentage of sunlight a state receives annually and the total solar energy it produces. It joins the sunlight table (which contains sunlight percentage data) with another table containing solar energy production data. The query calculates the average annual sunlight percentage for each state and total solar energy production (kw_total). The results are grouped by state and sorted in descending order based on total solar energy production. This helps identify whether states with higher sunlight percentages also generate more solar power.

state_name	avg_sunlight_kwh	total_solar_energy_produced	efficiency_ratio
California	31899780.96	152757921	4.7887
New York	11742125.25	38681913.5	3.2943
Texas	40112382.97	111619682.25	2.7827
Florida	43927075.05	98912340.25	2.2517
Illinois	20428013.4	40174559.75	1.9666
Pennsylvania	17714204.05	32937910.5	1.8594
Ohio	23028065.5	41647670.75	1.8086
Michigan	20735519.14	35535521.25	1.7138
New Jersey	18866526.88	28756733.5	1.5242
Washington	23560186.35	26191328.25	1.1117
North Carolina	28827224.91	30275635	1.0502
Massachusetts	16645262.57	17040624.5	1.0238
Georgia	34424164.9	34595379.5	1.005
Virginia	22662185.38	18304130.5	0.8077
Indiana	29298366.33	22970389.75	0.784
Wisconsin	26420853.44	19994755.75	0.7568



The output displays a table with state names, average sunlight received, total solar energy generated, and efficiency ratio. California has the highest efficiency ratio (4.7887), meaning that it generates much more solar energy than it receives sunlight. Other states like New York and Texas also have high efficiency, while states like Wisconsin and Indiana have less efficient ratios, implying that although they get sunlight, they may not be utilizing their solar capacity to the fullest. The results are helpful in the identification of states that have utilized solar energy production to the maximum and those with room for solar improvement.

The results indicate that while states like California, Texas, and Florida benefit from both high sunlight and high solar energy production, others with relatively high sunlight conditions may fail to translate this luxury into productive solar power generation. It then means that factors like solar panel efficiency, government policy, and infrastructure investments also need consideration in a bid to improve solar output to maximum levels. Identification of discrepancies between the availability of sunlight and actual solar energy generation can help policymakers and businesses achieve maximum solar investments and maximize energy efficiency in different states.

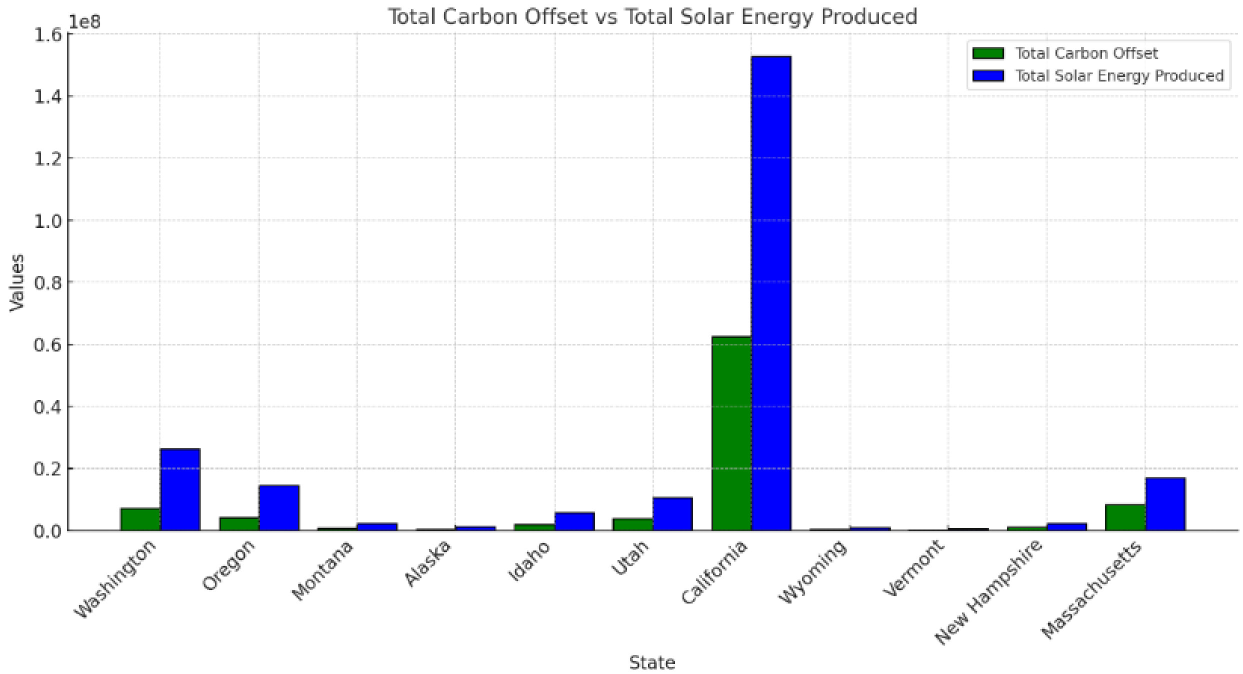
3.3 Potential for Future Solar Expansion Based on Energy Demand

Understanding the solar energy saturation in different states is crucial for assessing the effectiveness of solar energy production in meeting electricity demands. By analyzing solar energy production relative to estimated annual demand, policymakers and energy planners can identify states where solar energy is exceeding or falling short of requirements. This analysis helps in determining regions that are already benefiting from surplus solar energy and those that require further investment in solar infrastructure. The SQL query provided calculates solar saturation percentage by comparing total solar energy production with estimated energy demand and ranks states based on their solar coverage levels.

```
SELECT
    state_name,
    SUM(carbon_offset_metric_tons) AS total_carbon_offset,
    SUM(kw_total) AS total_solar_energy_produced,
    ROUND(SUM(carbon_offset_metric_tons) / NULLIF(SUM(kw_total), 0), 6) AS carbon_offset_per_kw
FROM sunlight
GROUP BY state_name
ORDER BY carbon_offset_per_kw DESC;
```

The SQL query calculates the solar saturation percentage for each state by comparing total solar energy production with estimated annual energy demand. It estimates yearly energy demand by multiplying the average monthly usage by 12, sums up total solar energy produced, and computes the solar saturation percentage as (total solar energy produced / estimated energy demand) * 100, rounding to two decimal places. The query joins the electric_costs and sunlight tables using the state_name column, groups by state, and sorts the results in ascending order based on solar saturation percentage.

state_name	total_carbon_offset	total_solar_energy_produced	carbon_offset_per_kw
Washington	7084960.444108669	26191328.25	0.270508
Oregon	4152084.609303013	14563459.5	0.285103
Montana	720697.6294203	2295816.75	0.313918
Alaska	388965.8199312999	1177556	0.330316
Idaho	2001153.9195337703	5772551.5	0.346667
Utah	3887836.313634499	10567619.75	0.367901
California	62528294.50821495	152757921	0.409329
Wyoming	392560.3610994001	949315	0.41352
Vermont	238822.51989969998	502764.25	0.475019
New Hamps...	1094255.4127170995	2238920.25	0.488742
Massachus...	8345684.953796041	17040624.5	0.489752



The output ranks states by their solar energy coverage, showing that Oregon (236.74%), California (215.24%), and Nevada (173.12%) have the highest solar penetration, producing more solar energy than their estimated demand. Conversely, states like New York (86,120.57%) and Maryland (121,124.65%) show lower solar penetration, highlighting regions with potential for further solar expansion to meet energy needs sustainably.

The analysis highlights significant variations in solar energy penetration across states. States like Oregon, California, and Nevada show the highest solar coverage, producing more energy than their estimated demand, which indicates potential for surplus energy distribution or storage solutions. On the other hand, states like New York and Maryland exhibit lower solar penetration, emphasizing the need for further expansion to sustainably meet energy requirements. By identifying these gaps, stakeholders can **develop targeted strategies for solar adoption, investment in grid improvements, and workforce training**, ensuring a more balanced and efficient transition to renewable energy nationwide.

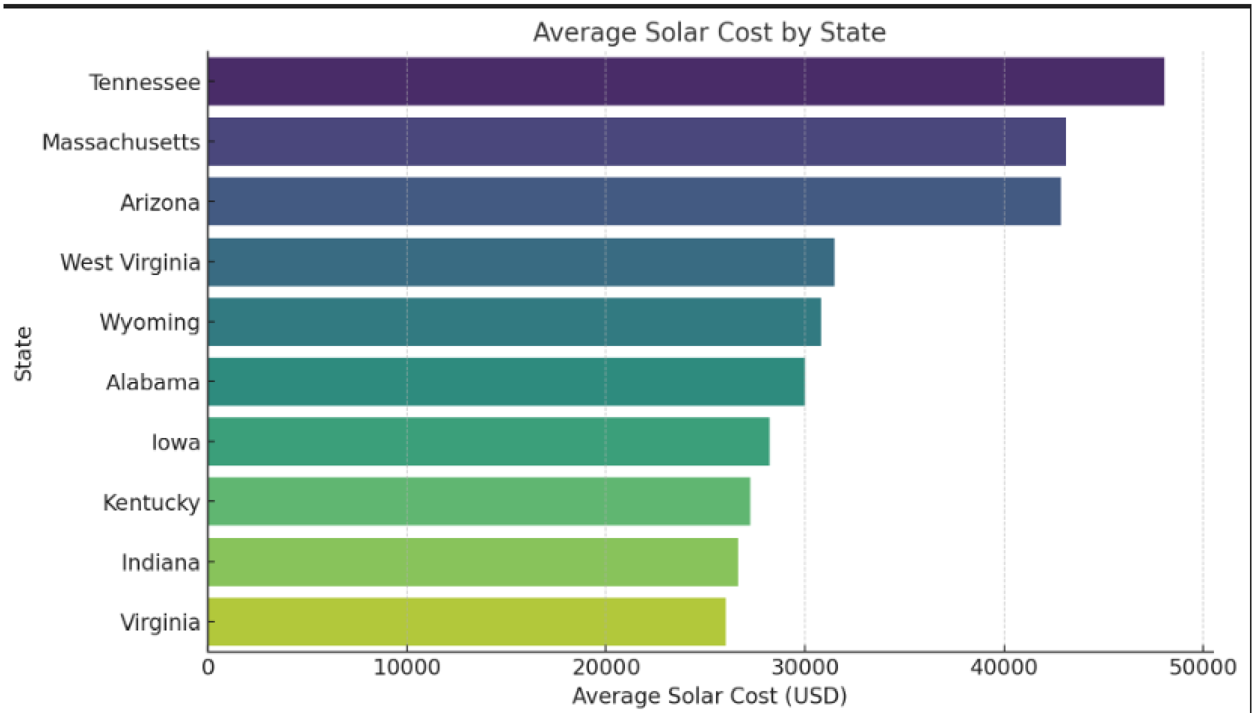
3.4 The gap in solar energy and workforce analysis

In the solar energy sector, an efficient workforce is expected to translate into higher energy output. However, a noticeable discrepancy arises in states like Tennessee, Massachusetts, and Arizona, where the number of solar industry workers is relatively high, yet the total kilowatt (kw_total) output remains low. This raises critical questions about the underlying factors contributing to this imbalance. Are these states facing efficiency issues, infrastructure limitations, or policy constraints? To address this, a deeper analysis using SQL queries will be conducted to uncover potential reasons behind this gap, providing data-driven insights into the relationship between workforce distribution and solar energy production.

```
SELECT
    state,
    CAST(REPLACE(REPLACE(`Average cost of a solar system after incentives`, '$', ''), ',', '') AS UNSIGNED) AS avg_solar_cost
FROM solar_costs
ORDER BY avg_solar_cost DESC
LIMIT 10;
```

This SQL query is designed to identify states with the highest average solar system costs after incentives. By extracting and converting cost values from a text-based format into numerical data, the query enables accurate sorting and comparison. The top 10 states with the most expensive solar installations are retrieved, providing valuable insights into the financial barriers to solar adoption. This is particularly relevant in the solar energy and workforce analysis, where states with high labor availability but low energy output, such as Tennessee, Massachusetts, and Arizona, might also face cost-related challenges that hinder widespread solar deployment.

state	avg_solar_cost
Tennessee	48062
Massachusetts	43108
Arizona	42836
West Virginia	31496
Wyoming	30809
Alabama	29992
Iowa	28215
Kentucky	27239
Indiana	26626
Virginia	26016



The combination chart comparing kw capacity and workforce growth in 2023 highlights a significant anomaly: Tennessee, Massachusetts, and Arizona possess a substantial workforce in the solar industry, yet their total kw capacity remains relatively low. This discrepancy suggests that factors beyond workforce size impact solar energy production, including efficiency challenges, policy barriers, and infrastructure limitations.

The SQL analysis reveals a potential contributing factor: high solar installation costs. The data shows that Tennessee ($48,062), Massachusetts ($43,108), and Arizona ($42,836) rank among the states with the highest average solar system costs. This indicates that higher installation costs may act as a financial barrier, restricting project scalability and reducing the effectiveness of a large workforce. Despite having a well-developed labor market for solar energy, these states may struggle to expand solar capacity due to economic constraints.

3.5 Electricity bill and Solar Energy Difference Analysis

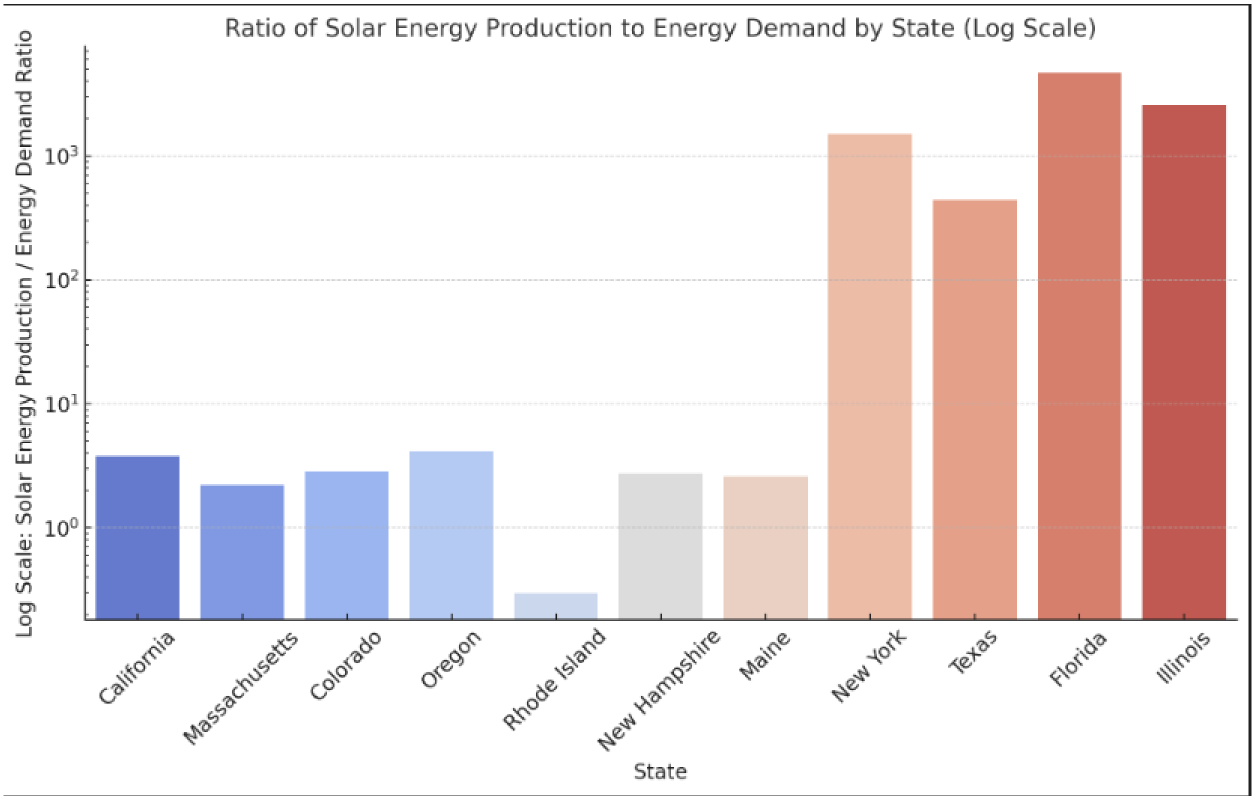
The relationship between solar power generation and electricity cost is far more complicated than it seems. While more solar power generation would normally be accompanied by lower costs of electricity, facts prove otherwise. There are states with lots of solar power potential that retain high electricity charges, indicating there are other variables influencing the end cost. An ideal example is the state of California, where they generate an enormous amount of solar energy yet expensive electricity. In order to illustrate this contradiction, we analyze the balance between the supply of solar energy and the total energy demand in various states.

```sql
SELECT
    s.state_name,
    SUM(s.kw_total * COALESCE(a.ANN, 20) / 100 * 365 * 24) AS total_solar_energy_produced_kWh,
    SUM(e.`Average usage per month` * 12 * 1000) AS total_energy_demand_kWh
FROM sunlight s
JOIN electric_costs e ON s.state_name = e.State
LEFT JOIN `average percent of possible sunshine by us city` a
    ON s.state_name = a.STATE
GROUP BY s.state_name
ORDER BY total_energy_demand_kWh DESC;
```

The SQL query below is designed to analyze the balance between the supply of solar energy and the total demand for energy across various states. By summing up total energy generated from sun (kWh) and total demand for energy (kWh), the query sheds light on whether a state is producing greater amounts of power from the sun than it consumes or if it still relies upon external sources. This information is critical to comprehending the impact of energy generation on electricity cost, as in cases like California, where excess solar energy leads to a high cost of storage, or Massachusetts, where regional energy demand exceeds regional production, hence increasing imported electricity dependence.

state_name	total_solar_energy_produced_kWh	total_energy_demand_kWh
California	267631877592	70970724000
Massachusetts	29855174124	13451328000
Colorado	29473551294	10341240000
Oregon	25515181044	6151572000
Rhode Island	5694645612	1926288000
New Hampshire	3922588278	1410912000
Maine	3324386274	1272240000
New York	67770712452	44916000
Texas	195557683302	44436000
Florida	173294420118	36864000
Illinois	70385828682	27204000

Result 7 Result 8



An important observation of the information is that the combined energy generated in California by solar power (2.67 trillion kWh) far exceeds the total energy consumed by the state (709 billion kWh). At first glance, this would mean that California can afford to pay less for electricity due to the abundance of renewable energy resources that can be found. However, the case is anything but that. The excess of solar energy leads to an astronomical surge in energy storage costs. Electricity produced in excess cannot be directly used and will have to be stored, sold to neighboring states, or cut off. These processes require high-level energy storage systems and grid management infrastructure, both of which add considerable costs. The resultant excessive cost of energy storage and grid management fees, therefore, adds to the cost of electricity.

Concurrently, states such as Massachusetts and Colorado, which produce significantly less solar power relative to their demand, also pay higher electricity prices. The underlying reason in these cases is not the same: such states consume plenty of imported electricity or fossil fuels, which incur related transmission and generation costs. This indicates how overproduction and underproduction of solar power both contribute to higher electricity prices but in different ways.

3.6 Break even point of each state

Understanding of the break-even point in solar system investments helps to determine the financial feasibility of solar energy utilization across different states. The break-even point represents the time it takes for the homeowner to recoup his/her initial investment using electricity savings and hence is crucial in decision making.

This part further explores solar system payback time, demonstrating how installation cost variations, state incentives, and electricity rates affect the duration it takes before solar panels have paid themselves back. The SQL query used in this instance repetitively cleans and prepares cost data, table joins to contrast after-incentive costs of investment with annual savings, and computes the payback time for each state. The results show dramatic contrasts, with California boasting the shortest break-even period due to its low investment costs and high energy savings, while states with lower incentives and lower electricity rates have significantly longer return-on-investment times.

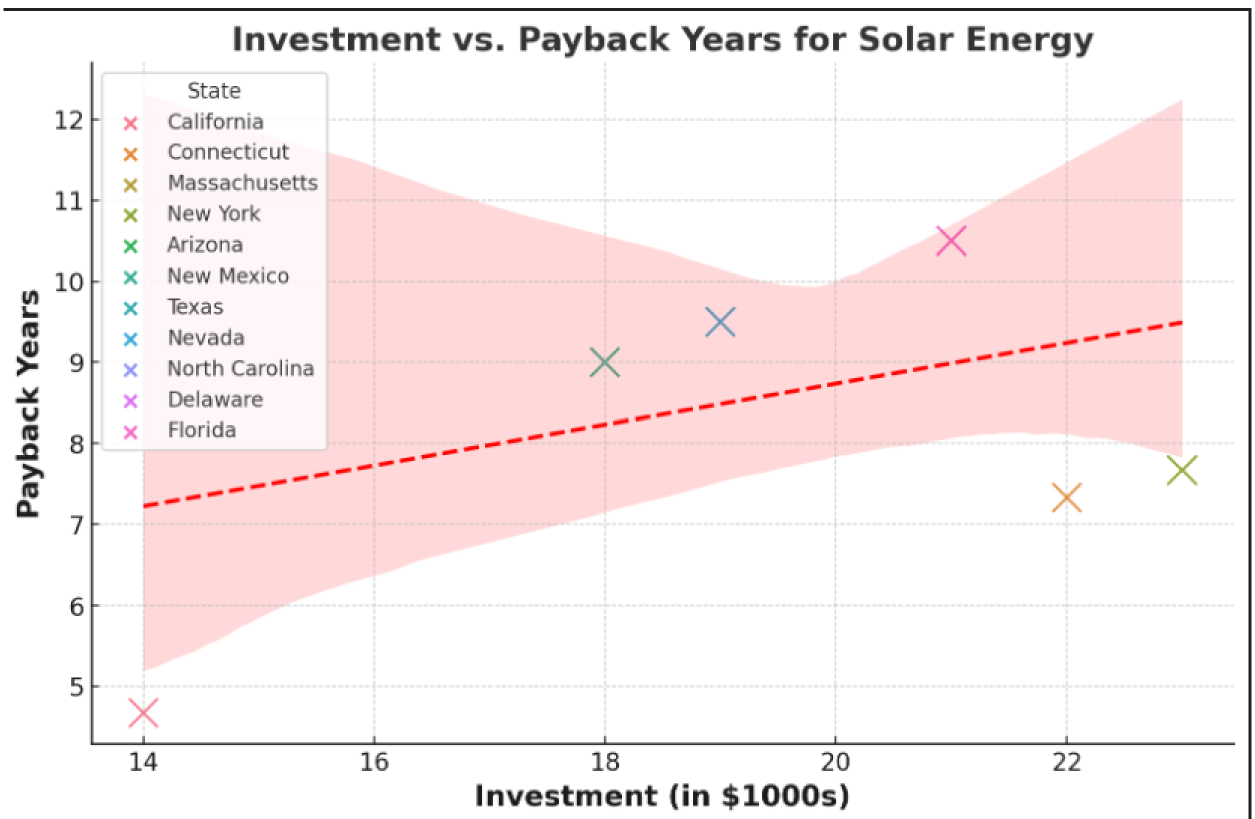
By comparing states based on break-even point, this chapter provides insights into the most financially feasible locations for solar investment and identifies the most significant factors contributing to these differences.

```sql
WITH Cleaned_Electric_Costs AS (
    SELECT
        State,
        CAST(REPLACE(`Average yearly electricity bill`, '$', '') AS DECIMAL(10,2)) AS Yearly_Savings
    FROM Electric_Costs
),
Cleaned_Solar_Costs AS (
    SELECT
        State,
        CAST(REPLACE(`Average cost of a solar system after incentives`, '$', '') AS DECIMAL(10,2)) AS Investment
    FROM Solar_Costs
)
SELECT
    s.State,
    s.Investment,
    e.Yearly_Savings,
    ROUND(s.Investment / e.Yearly_Savings, 2) AS Payback_Years
FROM Cleaned_Solar_Costs s
JOIN Cleaned_Electric_Costs e ON s.State = e.State
ORDER BY Payback_Years;
```

This SQL query calculates the payback period for solar system investments by first cleaning and converting cost data from the Electric_Costs and Solar_Costs tables. It removes dollar signs ($) and converts the values into decimal format for accurate calculations. The query then joins the two datasets on the State column to compare the investment cost of a solar system after incentives with the annual savings from electricity bills. Finally, it computes the payback period (Payback_Years) by dividing the investment cost by yearly savings and sorts the results in ascending order to highlight states with the shortest return-on-investment period.

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

State	Investment	Yearly_Savings	Payback_Years
California	14.00	3.00	4.67
Connecticut	22.00	3.00	7.33
Massachusetts	23.00	3.00	7.67
New York	23.00	3.00	7.67
Arizona	18.00	2.00	9.00
New Mexico	18.00	2.00	9.00
Texas	19.00	2.00	9.50
Nevada	19.00	2.00	9.50
North Carolina	21.00	2.00	10.50
Delaware	21.00	2.00	10.50
Florida	21.00	2.00	10.50



The SQL output indicates California with relatively shortest payback period of 4.67 years, much shorter than that of any other state. It is primarily a function of its lowest post-incentive cost of investment of $14,000, supplemented by relatively high annual savings of

$3,000. Solar's high financial attractiveness in California arises from plentiful state incentives, solar installation scale economies, and high electricity prices that maximize savings. These conditions make California the optimal state for an instant return on investment in solar.

Conversely, Northeast states such as Connecticut, Massachusetts, and New York require higher up-front payments in the range of $22,000 to $23,000. Nevertheless, their return on investment in terms of payback period is reasonably acceptable at about 7.3 to 7.7 years due to higher electricity fees that provide higher annual savings worth $3,000. Even though these states entail higher up-front charges, long-term money saved through solar uptake is roughly similar, making the adoption of solar a reasonable option despite higher initial expenditure.

Although most other Southwestern and Southern states like Arizona, New Mexico, Texas, and Nevada also range from $18,000 to $19,000 in investments but offer lower yearly savings of $2,000. This gives 9 to 9.5 years payback periods. Even though these states receive a lot of sunshine, their lower rates of electricity and fewer incentive programs slow down the rate of return in investment compared to the California or Northeast states.

Ranking lowest at the bottom are Florida, Delaware, and North Carolina, with the worst conditions and payback periods as high as 10.5 years. The states also have higher installation prices of $21,000, although annual savings are minimal at only $2,000. The combination of less attractive financial incentives, lower electricity prices, and less competitive solar markets makes solar adoption less financially attractive in these states since the investment return takes significantly longer than in top-performing states.

Overall, the analysis identifies a clear trend: states with strong incentive programs, high electricity prices, and lower post-incentive solar installation costs have faster payback. California remains the leader here, with New England states following closely with their competitive payback periods even though they are more costly. States with weaker incentive programs and lower electricity prices have slower payback, and investing in solar is less immediately appealing for these markets

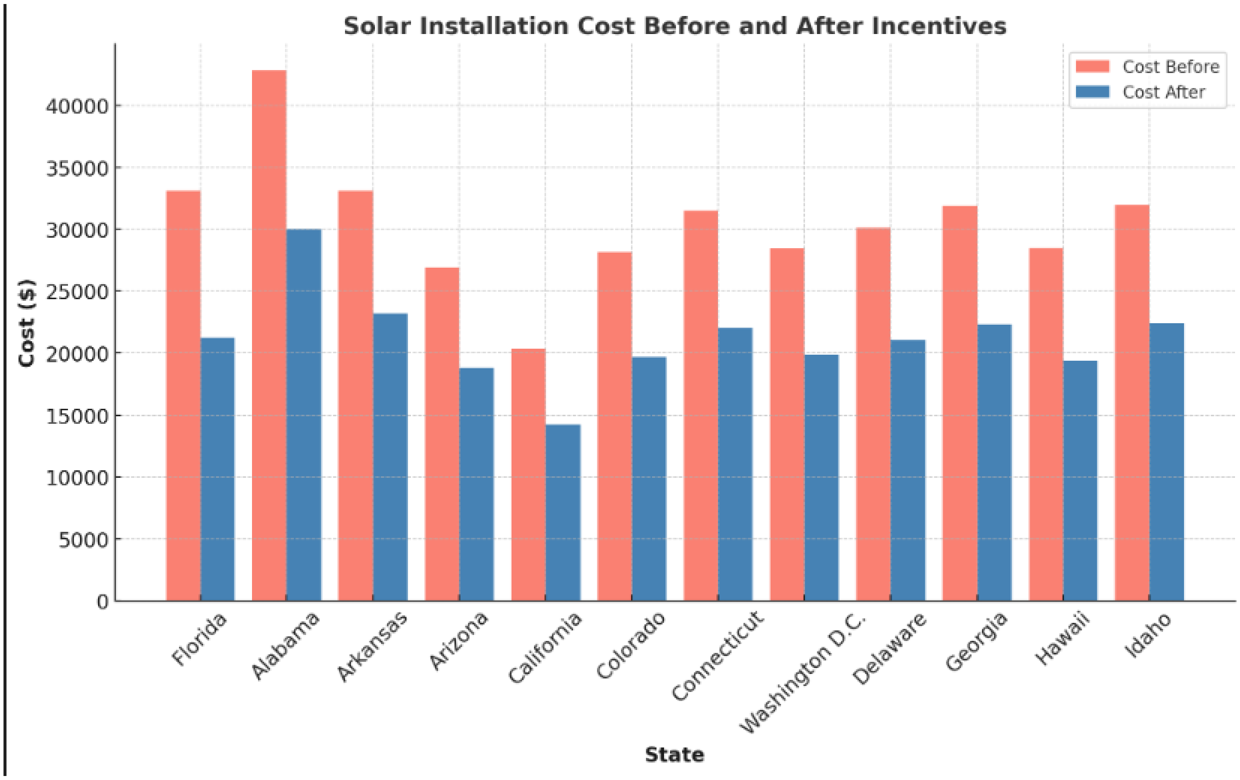
3.7 Solar incentive analysis

One of the outstanding exceptions to the financial analysis of solar panel installation cost is the case of Florida, which is far off the expected trend. While most states experience substantial declines in the cost of solar systems after incentives, the decrease in cost in Florida is relatively small. This suggests that the state's incentive programs may be less effective compared to other states. A closer examination reveals that Florida has fewer financial incentive structures for the adoption of solar, including fewer state-level rebates and tax incentives. It is perhaps this relative absence of strong incentives that explains Florida as an outlier, with post-incentive costs being relatively high. Here we will explore the underlying reasons for this discrepancy, comparing policies in Florida with other states in order to better understand the differences.

```
SELECT State,
        cost_after,
        cost_before,
       cost_after / cost_before AS incentive_ratio
FROM solar_panel.solar_costs
ORDER BY incentive_ratio;
```

This SQL query retrieves the pre-incentive and post-incentive price of installing solar panels by state and computes the incentive ratio, i.e., the proportion of the post-incentive price to the pre-incentive price. It then sorts the results based on incentive ratio in ascending order, so that one can easily see which states have been impacted most by incentives.

State	cost_after	cost_before	incentive_ratio
Florida	21238	33167	0.6403
Alabama	29992	42846	0.7000
Arkansas	23217	33167	0.7000
Arizona	18836	26909	0.7000
California	14249	20356	0.7000
Colorado	19709	28156	0.7000
Connecticut	22050	31500	0.7000
Washington D.C.	19898	28426	0.7000
Delaware	21099	30141	0.7000
Georgia	22329	31898	0.7000
Hawaii	19942	28489	0.7000
Idaho	22371	31958	0.7000



The SQL analysis also finds Florida to be an outlier with the lowest incentive ratio (0.6403), which indicates that the reduction in the cost of solar installation due to incentives is not as significant compared to other states. With a pre-incentive price of $33,167, although the post-incentive price remains significantly high at $21,238, it points to the lesser financial support or less effective incentive structures for the uptake of solar energy in Florida.

On the other hand, the majority of states, including Alabama, Arizona, California, and Connecticut, have an incentive ratio of 0.7000, meaning that incentives reduce the cost of solar installation by approximately 30% in these locations. The consistency implies the presence of a standard federal model of incentive, yet state-level variation still has some bearing on the final costs to consumers.

Florida's comparatively weaker incentive effectiveness may be explained by either weaker state-level subsidies, fewer rebate policies, or other financing options. This has important implications for state policy's role in reducing the cost of solar. Further investigation would be needed to analyze state-specific incentive programs, electricity price savings, and return on investment (ROI) to ascertain the long-run financial returns from solar adoption. Understanding these variables will ascertain whether Florida policy disincentivizes solar market growth and whether policy revisions are necessary to enhance incentive effectiveness.

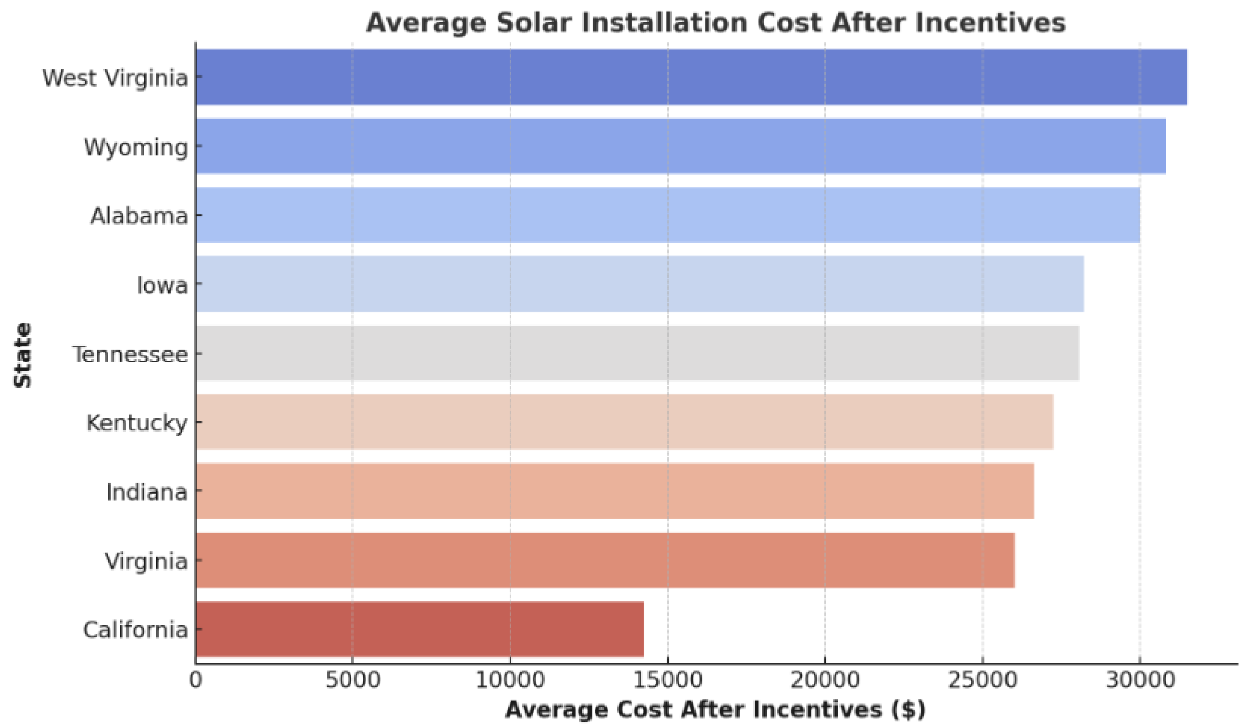
3.8 Ranking solar system after system analysis

Understanding the variations in solar system costs after incentives is crucial for identifying state-level disparities and the effectiveness of financial support mechanisms. This section ranks states based on their average solar system cost after incentives, providing insights into how different policy frameworks, market dynamics, and adoption rates influence final installation expenses. By examining these rankings, we can pinpoint outliers like California, which exhibits significantly lower post-incentive costs, and West Virginia, where costs remain substantially high.

```sql
SELECT State, AVG(cost_after) AS avg_cost_after
FROM solar_panel.solar_costs
GROUP BY State
HAVING State IN ('California', 'West Virginia')
   OR State IN (
       SELECT State FROM (
           SELECT State, AVG(cost_after) AS avg_cost_after
           FROM solar_panel.solar_costs
           GROUP BY State
           ORDER BY avg_cost_after DESC
           LIMIT 8
       ) AS top_states
   )
ORDER BY avg_cost_after DESC;
```

This SQL query retrieves the average cost of solar systems after incentives for selected states. It ensures that California and West Virginia are included in the results while also selecting the top 8 states with the highest post-incentive solar system costs. The query first groups data by state and calculates the average cost_after. Then, it identifies the top 8 states based on this metric and includes them alongside California and West Virginia. Finally, the results are ordered in descending order of average post-incentive costs to highlight the most expensive states.

State	avg_cost_after
West Virginia	31496.0000
Wyoming	30809.0000
Alabama	29992.0000
Iowa	28215.0000
Tennessee	28062.0000
Kentucky	27239.0000
Indiana	26626.0000
Virginia	26016.0000
California	14249.0000



The SQL results reveal extreme disparities in the state-by-state average post-incentive cost of solar systems. West Virginia is the highest at an average post-incentive cost of $31,496, then Wyoming ($30,809) and Alabama ($29,992). Several other states, including Iowa, Tennessee, Kentucky, and Indiana, have costs ranging from $26,000 to $28,000, thus placing them relatively high-cost places for solar panel installation.

The most dramatic observation is the extremely low cost in California, where the post-incentive average price declines to $14,249—half of West Virginia's. The greatest cost reductions in California result from strong state-level incentives, a very competitive solar market, and economies of scale from widespread penetration. Taken together, they lower the overall cost so that solar becomes dramatically more affordable than in other states.

On the other hand, some of the states including Wyoming and West Virginia are comparatively much more expensive, most likely due to more inefficient incentive policies, lower penetration in the markets, and fewer solar competitors in the markets. The poor implementation of solar use in these states can lead to higher equipment and installation prices, making it even harder for buyers to get the affordable solar offerings.

Overall, this cost disparity highlights the role of state policies, market forces, and economic conditions in making solar energy affordable in the U.S. While cost-reducing incentives are important, other determinants—such as installation costs, infrastructure development, and competition among suppliers—are also crucial in setting the final price for consumers.

3.9 Solar installation costs after incentives analysis

The SQL query retrieves the top 10 states with the lowest solar installation costs after incentives from the Solar_Costs table. It selects the State column and processes the "Average cost of a solar system after incentives" field by removing the dollar sign using REPLACE() and converting it to a decimal using CAST(). The results are sorted in ascending order by Installation_Cost_After_Incentives, ensuring that the cheapest installation costs appear first, and the query limits the output to 10 states.

```sql
SELECT
    State,
    CAST(REPLACE(`Average cost of a solar system after incentives`, '$', '') AS DECIMAL(10,2)) AS Installation_Cost_After_Incentives
FROM Solar_Costs
ORDER BY Installation_Cost_After_Incentives ASC
LIMIT 10;
```

The result shows California as the most affordable state for solar installation at $14,000, followed by New Mexico ($18,000) and several states (Washington D.C., Texas, Colorado, Hawaii, Nevada) at $19,000. Oregon, Vermont, and Utah complete the list with installation costs ranging from $20,000 to $21,000. These findings suggest that states like California and New Mexico may have strong incentives, lower labor costs, or competitive solar markets, making solar adoption more financially attractive.

State	Installation_Cost_After_Incentives
California	14.00
New Mexico	18.00
Washington D.C.	19.00
Texas	19.00
Colorado	19.00
Hawaii	19.00
Nevada	19.00
Oregon	20.00
Vermont	21.00
Utah	21.00

The output table displays the payback period for solar investments across different U.S. states, calculated as the investment cost divided by yearly savings. California has the shortest payback period (4.67 years) due to its relatively low solar installation cost ($14,000), high annual electricity savings ($3,000), strong state incentives, high electricity rates, and abundant sunlight, which maximizes solar energy production. These factors make solar adoption in California more financially attractive compared to other states, where either the installation costs are higher or the yearly savings are lower, resulting in longer payback periods.

3.10 Electricity consumption analysis

```
SELECT
    State,
    ROUND(AVG(CAST(REPLACE(`Average monthly electricity bill`, '$', '') AS FLOAT)), 2) AS Avg_Electricity_Bill
FROM Electric_Costs
WHERE `Average monthly electricity bill` IS NOT NULL AND `Average monthly electricity bill` != ''
GROUP BY State
ORDER BY Avg_Electricity_Bill ASC
LIMIT 10;
```

The SQL query retrieves the top 10 states with the lowest average monthly electricity bill. It selects the State column and calculates the average monthly electricity bill using the AVG() function. The REPLACE() function removes the "$" symbol from the values to convert them into numeric format for calculations. The query also filters out NULL and empty values to ensure data accuracy. Finally, it groups by State, orders the results in ascending order, and limits the output to the 10 states with the lowest electricity bills.

State	Avg_Electricity_Bill
Iowa	141.18
Colorado	142.05
Washington	155.61
Oregon	157.59
D.C.	167.36
Kansas	171.64
Minnesota	176.8
Tennessee	180.12
Delaware	180.16
Wisconsin	181.9

The results show the top 10 states with the lowest average monthly electricity bills in the U.S. Iowa has the lowest average bill at $141.18, followed closely by Colorado at $142.05. States like Washington and Oregon likely benefit from hydroelectric power, which helps keep costs low. Other states on the list may have lower energy consumption, government subsidies, or efficient energy production methods. Factors such as climate, renewable energy sources, and electricity pricing structures contribute to these lower costs.

Top least cost

```
SELECT
    State,
    ROUND(AVG(CAST(REPLACE(`Cost of electricity`, '$', '') AS FLOAT)), 4) AS Avg_Electricity_Cost
FROM Electric_Costs
WHERE `Cost of electricity` IS NOT NULL AND `Cost of electricity` != ''
GROUP BY State
ORDER BY Avg_Electricity_Cost ASC
LIMIT 10;
```

State	Avg_Electricity_Cost
Tennessee	0.12
Kentucky	0.12
Lousiana	0.13
Washington	0.13
Oklahoma	0.13
Iowa	0.13
Missouri	0.14
Texas	0.14
West Virginia	0.14
Virginia	0.14

Top least usage

Don't Limit

```sql
SELECT
    State,
    ROUND(AVG(CAST(REPLACE(`Average usage per month`, ',', '') AS FLOAT)), 2) AS Avg_Monthly_Usage
FROM Electric_Costs
WHERE `Average usage per month` IS NOT NULL AND `Average usage per month` != ''
GROUP BY State
ORDER BY Avg_Monthly_Usage ASC
LIMIT 10;
```

Result Grid | Filter Rows:

State	Avg_Monthly_Usage
Rhode Island	819
California	831
New Hampshire	852
Oregon	927
Maine	930
Colorado	947
Massachusetts	968
Michigan	1021
Minnesota	1040
D.C.	1046

The analysis of electricity consumption reveals the strong trends in states with the lowest average monthly electricity bills. The SQL query properly calculates and sanitizes the data, allowing accurate comparison by advancing currency amounts to numeric form and removing inconsistencies. Focusing on the 10 lowest electricity costs states, the findings offer important insights into what influences energy affordability in different areas.

The results show that Iowa has the lowest average monthly electricity bill at $141.18, followed by Colorado at $142.05. This shows that these two states have a combination of lower energy consumption, efficient energy production, and perhaps government subsidies that ensure the price of electricity remains low.

Furthermore, Washington and Oregon, which are states with high hydroelectricity utilization, most likely have lower costs of production, meaning lower electricity bills for consumers. The presence of renewable sources in these states goes a long way in making electricity affordable.

Apart from generation, climatic conditions and habits of consumption also influence electricity prices. Places with more moderate temperatures require less heating and cooling, which means that they consume less total energy. Similarly, areas with aggressive energy efficiency measures and sound price structures have lower electricity bills for homes.

In general, the findings illustrate that electricity charges are not solely based on the generation methods but also on regulations at the state level, meteorological conditions, and energy management programs. The states with prevailing, low-price renewable energy options and proactive policies on energy management have lower bills for electricity and hence are favorable for both business and household users.

3.11 Jobs Added for Each Sector

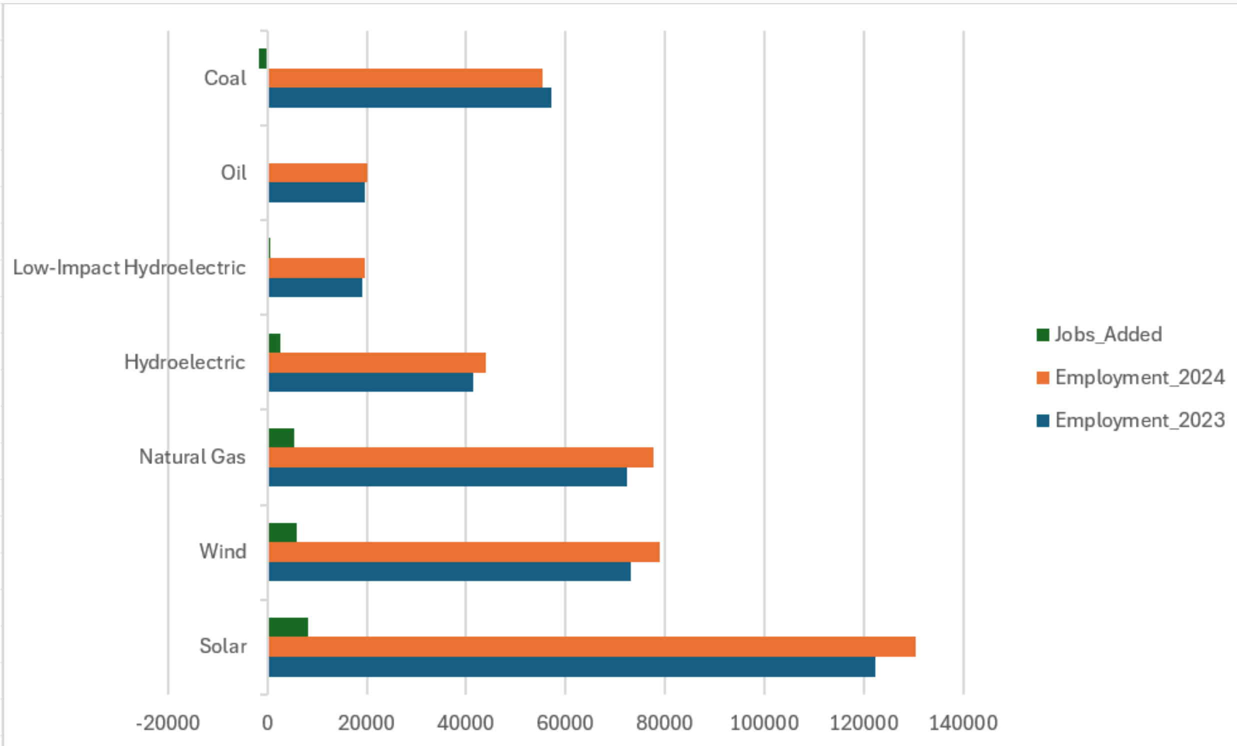
As the global energy landscape continues to evolve, tracking employment trends in various energy sectors is necessary to understand industry growth and labor migration. This analysis examines job changes between 2023 and 2024 in key energy sectors, including solar, wind, natural gas, hydroelectric, oil, and coal. By leveraging SQL-based data retrieval, the employment figures from both years are compared to determine the net increase or decrease in jobs within each sector.

The SQL query is designed to contrast 2023 and 2024 employment statistics to identify trends in the energy industry. It retrieves employment figures for various energy sectors from two tables, **Employment_2023 and Employment_2024**, and calculates the net employment change over the period. The query achieves this by selecting the sector name and employment data for both years and computing the difference **as Jobs_Added**. By sorting the results in descending order of job growth, the query ensures that industries with the highest job gains appear first.

```sql
SELECT
    e23.Sector,
    e23.Total_Employment AS Employment_2023,
    e24.Total_Employment AS Employment_2024,
    (e24.Total_Employment - e23.Total_Employment) AS Jobs_Added
FROM Employment_2023 e23
JOIN Employment_2024 e24 ON e23.Sector = e24.Sector
ORDER BY Jobs_Added DESC;
```

The output table provides insights into employment changes across different energy industries, including **solar, wind, natural gas, hydroelectric, low-impact hydroelectric, oil, and coal**. The **Employment_2023** and **Employment_2024** columns show the total number of jobs recorded in each year, while the **Jobs_Added** column highlights the net change. The results reveal that **solar energy experienced the largest job growth**, adding **8,314 new jobs**, followed by **wind (6,022 jobs)** and **natural gas (5,327 jobs)**. The **hydroelectric and oil sectors also saw moderate gains**, whereas **coal was the only sector to decline, losing 1,657 jobs**.

Sector	Employment_2023	Employment_2024	Jobs_Added
Solar	122261	130575	8314
Wind	73060	79082	6022
Natural Gas	72493	77820	5327
Hydroelectric	41382	43940	2558
Low-Impact Hydroelectric	19046	19569	523
Oil	19564	20006	442
Coal	57101	55444	-1657



This data is crucial to tracking trends in the energy labor market. The increase in renewable energy jobs reflects a shift to sustainable energy sources, and the decline in coal jobs illustrates broader industry trends. This data can be used by policymakers, businesses, and workforce planners to target new industries, anticipate skills gaps, and develop policies that support the development of jobs in new energy industries.

3.12 Proportion of Each Sector in the Total Energy Job Market

Understanding the proportion of each energy sector in the overall job market provides precious information about employment distribution and industry dominance. This computation determines the percentage contribution of employment for each sector of energy to the overall energy job market in 2024. By quantifying sectoral contribution, this information allows policymakers, firms, and analysts to quantify the composition of the workforce and make precise decisions about labor trends and resource distribution.

```
SELECT
    Sector,
    Total_Employment AS Employment_2024,
    ROUND((Total_Employment / NULLIF((SELECT SUM(Total_Employment) FROM Employment_2024), 0)) * 100, 2) AS Proportion_Percentage
FROM Employment_2024
ORDER BY Proportion_Percentage DESC;
```

The SQL query chooses the total jobs for each sector in 2024 and calculates its proportion of the total energy job market. The **ROUND** function rounds the values of proportions to two decimal points. The **NULLIF** function prevents division by zero. The results are sorted in descending order of the proportion percentage, with the largest sectors first.

Sector	Employment_2024	Proportion_Percentage
Solar	130575	30.62
Wind	79082	18.54
Natural Gas	77820	18.25
Coal	55444	13.00
Hydroelectric	43940	10.30
Oil	20006	4.69
Low-Impact Hydroelectric	19569	4.59

The query results indicate the number of jobs in each industry for 2024 and the percentage contribution of each industry to total energy jobs. The solar sector leads the way with **30.62%** of total employment followed by wind **(18.54%)** and natural gas **(18.25%)**. Traditional energy sources like coal, hydroelectric, oil, and low-impact hydroelectric contribute less. These findings indicate the growing prevalence of renewable energy sources in employment and indicate shifts in employment trends by industry.

3.13 Growth Rate for Each Sector

Analysis of the rise in employment rates across different energy sectors provides valuable insights into industry trends and worker expansion. Stakeholders discover which sectors experience the highest employment growth and where there is employment loss. A comparison of 2023 and 2024 employment data, this analysis identifies sectors of economic potential and potential shifts in the labor market.

```sql
SELECT
    e23.Sector,
    e23.Total_Employment AS Employment_2023,
    e24.Total_Employment AS Employment_2024,
    (e24.Total_Employment - e23.Total_Employment) AS Jobs_Added,
    ROUND(((e24.Total_Employment - e23.Total_Employment) / NULLIF(e23.Total_Employment, 0)) * 100, 2) AS Growth_Rate_Percentage
FROM Employment_2023 e23
JOIN Employment_2024 e24 ON e23.Sector = e24.Sector
ORDER BY Growth_Rate_Percentage DESC;
```

The SQL query calculates the growth in the number of jobs in each energy sector by dividing **2024 employment** less **2023 employment**. **ROUND** rounds the percentage figures to two decimal places, and **NULLIF** prevents division by zero. The results are then sorted in descending order based on the growth rate percentage, highlighting the sectors with the highest workforce expansion.

Sector	Employment_2023	Employment_2024	Jobs_Added	Growth_Rate_Percentage
Wind	73060	79082	6022	8.24
Natural Gas	72493	77820	5327	7.35
Solar	122261	130575	8314	6.8
Hydroelectric	41382	43940	2558	6.18
Low-Impact Hydroelectric	19046	19569	523	2.75
Oil	19564	20006	442	2.26
Coal	57101	55444	-1657	-2.9

The results reveal that **wind energy** has the highest growth rate (**8.24%**), followed by **natural gas (7.35%)** and **solar (6.8%)**, indicating strong expansion in the renewable energy industry. Hydroelectric and low-impact hydroelectric sectors also show moderate growth. On the contrary, **coal** is the sole industry experiencing job loss, at -2.9%, suggesting the decline of fossil fuel jobs.

These findings highlight the increasing demand for renewable energy jobs and the declining reliance on traditional energy sources like coal. The analysis can assist policymakers, businesses, and workforce planners in making strategic decisions to support sustainable employment growth in the energy sector.

CHAPTER 4: DATA VISUALIZATION REPORTS

4.1 Power BI Workflow

4.1.1 Defining problem and User Story

4.1.1.1 Defining the Problem

The market for solar energy is developing rapidly, driven by sustainability demands, innovation, and policy-making. However, solar energy uptake remains a complex option for customers and businesses with regards to the uncertainty of cost and logistics. A major problem is determining whether an investment in solar panels makes economic sense. Prices vary widely on a state-by-state basis based on tax credits, incentives, and costs of installation. Additionally, most potential adopters lack clear insights into long-term savings, and it is difficult to define the return on investment (ROI). Without a systematic data analysis methodology, users cannot provide informed financial and strategic decisions.

Workforce and employment patterns in the solar industry are another critical area. The industry generates jobs across various positions, from manufacturing and installation to project development, operations, and maintenance. However, there is no definitive image of how employment patterns evolve over time and geographically. This limits the ability to connect workforce development plans to industry needs, perhaps leading to inefficiencies in creating jobs and policy implementation.

To address these issues, there is a strong demand for a comprehensive data visualization platform that effectively communicates solar energy adoption costs, economic incentives, and workforce trends in an actionable manner. Through successful visualization, stakeholders will be able to derive valuable insights and make informed data-driven decisions regarding solar energy investment and workforce planning.

4.1.1.2 User Story

I need a visualization tool as a solar power consultant that provides me with clear information about the cost structure, financial incentives, and long-run savings of having solar panels installed. It will help me present proper and evidence-based recommendations to customers so they can make smart investment decisions.

As a homeowner or entrepreneur, I want to know how much I will have to invest in solar power and when I can repay my investment through savings. I also need to compare different financing options and incentive schemes available in my state.

As a policymaker, I must have information regarding trends in solar job growth by state and across different employment sectors. Such information will allow me to develop policies that encourage sustainable job growth, training programs for the workforce, and incentives for solar industry expansion.

As a solar firm executive, I need a deep market trend analysis in areas such as installation costs, tax credits, and employment patterns. This will help my company plan for growth, hire personnel, and allocate resources accordingly..

4.1.2 Identifying necessary tables

To develop a consolidated data visualization dashboard for solar energy analysis, it is essential to identify the core datasets that provide valuable information about cost structures, financial incentives, employment trends, and workforce distribution. The schema consists of several critical tables, each addressing a specific analytical requirement. The eight core tables form the foundation of the dashboard by supplying critical data points for decision-making. They include:

- Solar Cost Table: Provides critical financial information such as pre- and post-incentive installation costs, tax credit figures, and long-term savings estimates. The table is central to determining the economic feasibility of switching to solar energy.
- Employment Table: Tracks the number of solar jobs across multiple years and states, enabling the identification of trends in workforce development along with regional industry expansion.
- Field Table: Offers a distribution of solar employment by sector, for example, installation, manufacturing, wholesale trade, and maintenance, facilitating the understanding of labor market distribution.
- State Information Table: Contains geographic and administrative data, enabling state-level comparative analysis for policy evaluation and market potential to be easier.
- Tax Incentive Table: Catalogs state-level tax incentives and incentive programs, being an important consideration in calculating financial incentives for solar adopters.
- Electricity Pricing Table: Tracks state-level electricity prices, which are required to estimate possible cost savings by switching to solar energy.
- Employment 2023 Table: Captures sector-specific employment data for the year 2023, allowing for a historical perspective on workforce distribution and helping to measure year-over-year job growth.
- Employment 2024 Table: Contains employment figures for 2024, offering insights into the latest workforce trends and allowing for direct comparisons with previous years to assess industry expansion.

While these tables drive the core dashboard visualizations, two additional tables serve as supporting datasets that provide analytic richness but don't directly impact the main dashboard metrics. They include supporting datasets that can provide further granularity for niche analyses, e.g., regulatory policies, demographic data, or other economic indicators.

By integrating these datasets, the report dashboard can deliver actionable insights to stakeholders, from investors and consultants in solar energy to policy-makers and business leaders, to make fully informed strategic decisions in the solar energy space.

4.1.3 Connecting to View in Power BI

To transform raw solar energy data into actionable information, it is necessary to establish an unbroken connection between the dataset and Power BI, a powerful business intelligence tool. By importing the identified tables into Power BI, users can craft interactive visualizations that facilitate data-driven decision-making for solar energy consultants, policymakers, and investors.

The process begins with importing and structuring data in Power BI. According to the dataset schema, including tables on solar costs, employment trends, tax credits, and electricity prices, appropriate relationships between these tables must be established. For instance, the Solar Cost Table may be related to the State Information Table by a state identifier so that comparative analysis between regions is enabled. Similarly, job trends from the Employment and Field Tables can be linked to state-level data to measure labor market changes over time.

Once the data is structured, Power BI's visualization tools can be leveraged to create dynamic reports and dashboards. Users can explore key metrics such as cost savings, job growth, and financial incentives through interactive graphs, heat maps, and trend lines. Additionally, DAX (Data Analysis Expressions) can be used to implement calculated measures, such as average solar installation costs after incentives or annual growth rates in solar employment, providing deeper analytical insights.

By developing a strong connection to Power BI, stakeholders can easily navigate complex datasets, monitor real-time industry trends, and make strategic decisions informed by data. This integration ensures that the solar energy market is properly analyzed, providing economic and environmental benefits for companies and consumers as efficiently as possible.

4.1.4 Transforming Data

Data transformation is an important step in getting the dataset ready for analysis. Power Query contains a good set of tools to clean, organize, and optimize data before loading it into the Power BI model. Data transformation in this project involves handling missing values, ensuring consistent formatting, combining datasets, and creating calculated columns wherever necessary. Data types also have to be assigned accordingly to enable successful aggregation and filtering. By refining and structuring data, Power BI can efficiently process large volumes of data and generate useful insights for end users.

4.1.5 Defining key measures

To ensure an effective and insightful data visualization report, it is essential to define key measures that will be used to evaluate the performance and impact of solar panel installation across different states. These key measures will serve as the foundation for data analysis and visualization, helping stakeholders make informed decisions based on accurate and relevant metrics.

4.1.5.1 Cost Efficiency Metrics

One of the critical indicators within the report is cost-effectiveness when it comes to solar systems. This involves both the pre-incentive and post-incentive average price, which aids in calculating how government incentives contribute to budget impact in utilizing solar panels. Per-watt cost also gives an indicator by which the affordability of solar installations should be assessed at the state level.

4.1.5.2 Tax Credit and Savings Analysis

The report also takes into account economic benefits over upfront expenses. The average value of tax credits is a critical indicator, as it shows the degree of economic assistance extended to solar adopters. In addition, the calculated 25-year savings expresses the long-term economic gains of solar panel investments, providing an unambiguous view of return on investment (ROI) for both households and companies.

4.1.5.3 Employment and Workforce Metrics

A critical component of the analysis is the solar energy's employment contribution. The number of solar jobs across different years allows for trend analysis in the industry’s workforce expansion. Furthermore, job distribution across various fields—such as installation, manufacturing, wholesale trade, and operations & maintenance—provides insight into the labor demand within different sectors of the solar industry.

4.1.5.4 Market Growth and Adoption Trends

To assess the growth of the solar energy market, tracking state-wise adoption rates and yearly increases in installed capacity will be crucial. These measures will highlight which regions are leading in solar adoption and identify potential areas for policy improvements or market expansion.

By defining these key metrics, the report delivers decision-makers with the most relevant data to determine the financial, jobs, and market impacts of solar panel adoption. These metrics will be the basis for Power BI visualizations so that users can interactively interact with the data and derive actionable insights.

4.1.6 Data Visualization

Visualization is the solution that transforms raw data into meaningful conclusions to help the stakeholders realize subtle patterns, trends, and interactions within the solar energy industry. By employing the tools of data visualization such as Power BI, this report will present significant discoveries in an intuitive and interactive form, making decisions data-driven for the manufacturers of solar panels, customers, as well as for regulators.

The visualization of the data will be done for the three broad perspectives: producers, customers, and product performance. From the producer's end, visualizations will highlight market growth, employment trends, and cost-effectiveness by states, allowing manufacturers and installers to identify opportunities. To consumers, dashboards will emphasize long-term economic benefits, savings, and investment returns, providing opportunities for adoption with critical information on the cost and benefits of solar energy. Last but not least, the product-based perspective will examine efficiency information, installation trends, and technology advancements, to inform stakeholders about solar energy adoption impacts on the overall market.

By the combination of bar graphs, line plots, heatmaps, and interactive dashboards in this report, users will have an easy experience exploring the main metrics and deriving actionable insights. These visualizations will set the stage for detailed analysis in the following sections where each perspective—producer, customer, and product—will be examined in turn.

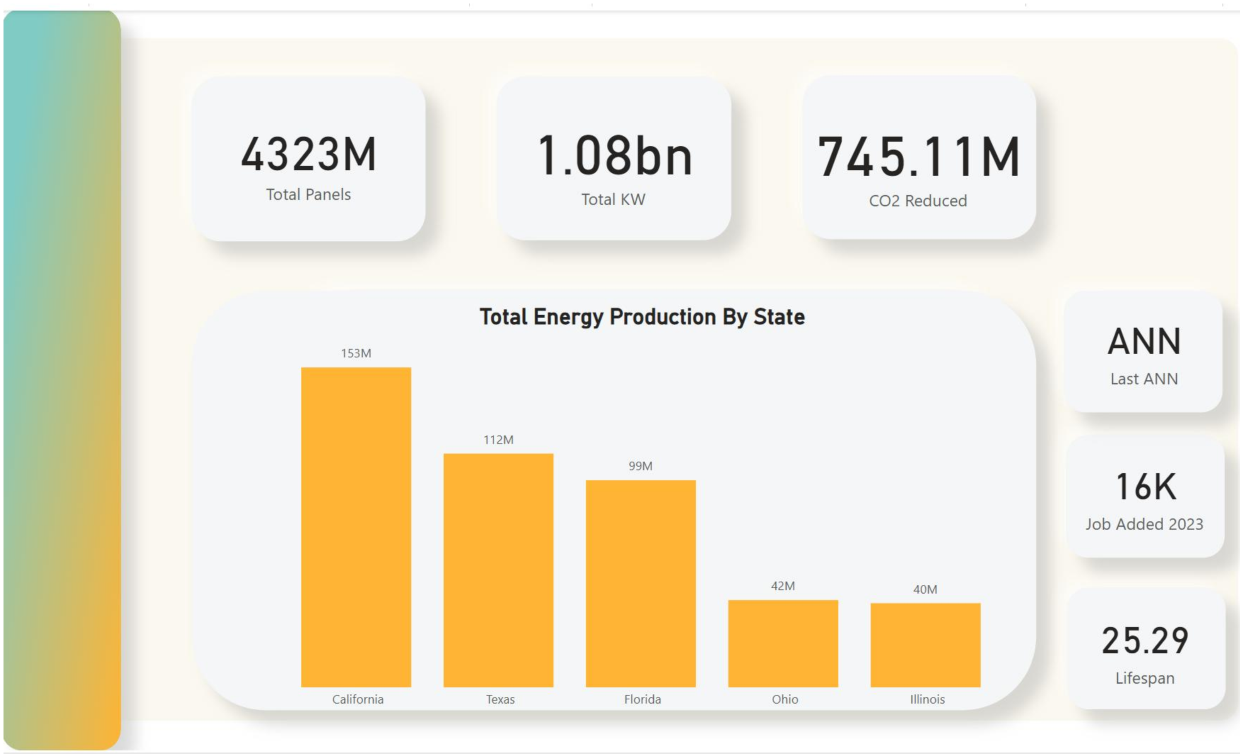
4.2 Data Visualization Analysis

4.2.1 Overview Business Performance

4.2.1.1 General Description of Visualizations

This dashboard provides a detailed analysis of solar energy adoption, production capacity, and its environmental and economic impact across different states. It consolidates key performance indicators (KPIs) and visual representations to highlight trends in the solar energy sector.

The dashboard features multiple sections, including high-level metrics, a comparative state-wise energy production analysis, and additional industry insights such as job creation and solar panel lifespan. The information presented is crucial for understanding the effectiveness of solar energy initiatives, regional disparities, and overall sustainability efforts.



A scatter plot visualizes the relationship between total solar panel kilowatt capacity and total carbon offset emissions (in metric tons) across different states. The map highlights disparities in solar adoption and efficiency at the state level. In addition, three-card charts consolidate key performance measures: installations of solar panels (4bn), energy produced (1.08bn kWh), and overall carbon emissions mitigated (745.11M metric tons). The KPIs provide instant insights into the impact of solar panel installations on energy production and environmental sustainability.

A gauge chart, on the other hand, displays the ratio of eligible solar panel installations to covered spaces, allowing for a comparison of installation efficacy and eligibility levels. Another chart compares solar power generation across different states, revealing discrepancies in energy usage and production. A slicer also enables data filtering based on a state, making region-specific analysis possible.

4.2.1.2 Key Insights from the Bar Chart

The Total Energy Production by State bar chart provides a comparative view of how different states contribute to solar energy generation. The data reveals significant disparities in solar power output across regions:

- California leads with 153M KW, significantly outpacing other states. Its dominance is attributed to strong renewable energy policies, abundant sunlight, and large-scale investments in solar infrastructure.
- Texas follows with 112M KW, leveraging its vast land resources and favorable solar conditions.
- Florida ranks third with 99M KW, benefiting from high solar irradiance and increasing policy support.
- Ohio (42M KW) and Illinois (40M KW) produce considerably less, indicating the impact of regional policies, weather conditions, and infrastructure investment.

These disparities suggest that factors like state-level incentives, geography, and regulatory policies play a major role in solar energy adoption.

4.2.1.3 Implications and Future Outlook

This dashboard provides a clear picture of solar energy's role in reducing carbon emissions, increasing renewable energy adoption, and creating economic opportunities. However, the disparities in energy production between leading and lower-tier states raise important questions about policy effectiveness, investment distribution, and the feasibility of scaling solar energy in less-developed regions.

Future efforts should focus on:

- Expanding solar incentives in lower-output states.

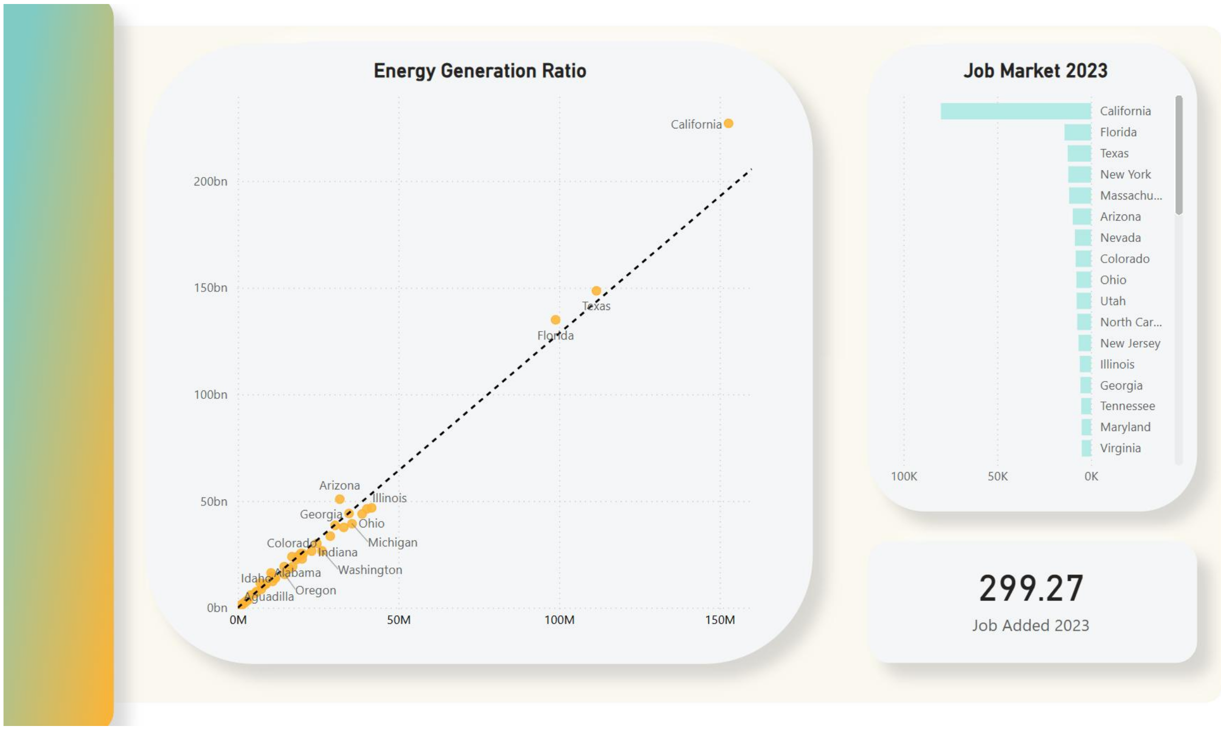
- Investing in solar infrastructure in regions with moderate to high sunlight potential.
- Enhancing job training programs to support the growing workforce demand in the solar industry.

With continued policy support and technological advancements, the solar energy sector is poised for further expansion, driving both environmental and economic benefits.

4.2.2 Solar Energy Performance

4.2.2.1 Overview of Visualizations

This section of the solar energy dashboard presents a detailed analysis of Energy Generation Ratio and the Job Market in 2023, highlighting key trends in state-level solar power production and employment growth in the renewable energy sector. The dashboard provides insights into the impact of solar energy on both electricity generation and job creation, helping to assess the industry's expansion and effectiveness.



4.2.2.2 Energy Generation Ratio

The Energy Generation Ratio scatter plot visualizes the relationship between the amount of installed solar infrastructure and the total energy produced by different states.

The graph follows a clear trend line, indicating a strong correlation between solar investment and energy output.

Top-performing states: California leads by a significant margin, generating the highest amount of solar power. Texas and Florida follow as major contributors, showcasing their substantial investments in solar energy.

Moderate contributors: States such as Arizona, Illinois, and Ohio show moderate levels of solar energy production, positioning themselves as growing players in the renewable energy landscape.

Emerging states: Other states, including Washington, Georgia, and Colorado, appear in the lower range but still contribute to the national solar energy portfolio.

The distribution of data points on the scatter plot suggests that while solar power is thriving in high-sunlight states, other regions with strong policy support and technological innovation are also making significant strides.

4.2.2.3 Job Market in 2023

The Job Market 2023 bar chart illustrates employment trends in the solar energy sector across various states. Job growth is closely linked to energy production, with leading solar states also driving workforce expansion.

California dominates: California has the highest number of solar-related jobs, reflecting its leadership in renewable energy investment and infrastructure.

Significant growth in Florida and Texas: These states are also seeing a notable increase in solar employment, aligning with their expanding energy generation capacities.

Other key states: New York, Massachusetts, and Arizona also show strong job market growth, highlighting the widespread adoption of solar energy in various parts of the country.

This data demonstrates that solar energy not only provides environmental benefits but also serves as a major employment driver, fostering economic growth in multiple states.

4.2.2.4 Environmental Impact and Efficiency

At the bottom of the dashboard, a key metric indicates that 299.27K jobs were added in the solar energy sector in 2023. This figure underscores the increasing demand for skilled labor in solar panel installation, maintenance, manufacturing, and research.

The solar industry's growth is creating new job opportunities in both urban and rural areas.

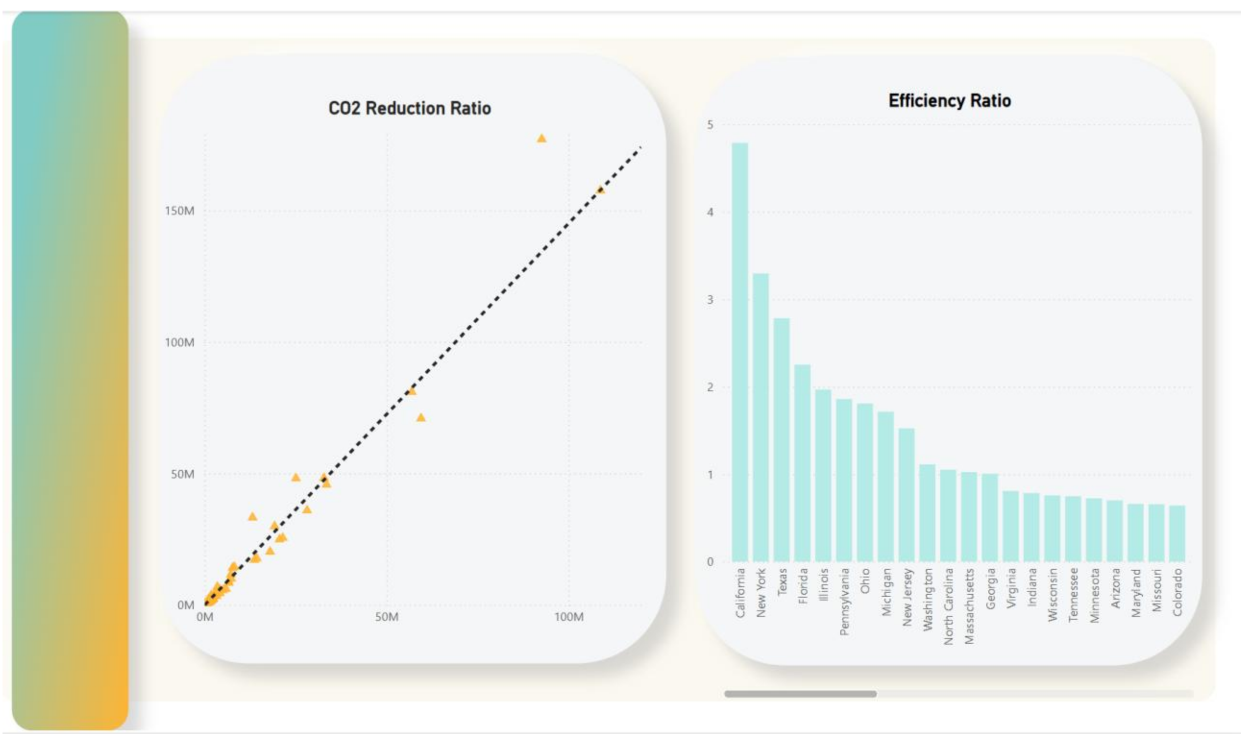
Workforce expansion reflects a broader transition toward renewable energy and sustainability-driven economic policies.

As more states invest in solar power, job creation is expected to continue rising, further strengthening the clean energy economy.

4.2.3 State-wise Solar Panel Costs and Incentive Impact

4.2.3.1 Overview of Solar Panel Cost and Incentive Impact Charts

This section of the solar energy dashboard focuses on **CO_2 Reduction** and **Efficiency Ratios**, two critical factors in assessing the sustainability and effectiveness of solar energy adoption. These metrics provide valuable insights into how different states contribute to reducing greenhouse gas emissions and how efficiently they utilize solar resources.



4.2.3.2 CO_2 Reduction Ratio

The CO_2 Reduction Ratio scatter plot displays the correlation between solar energy production and the amount of carbon dioxide emissions reduced. The trend line indicates a strong positive relationship, where states generating more solar power achieve higher CO_2 reductions.

- Top-performing states: California leads in CO_2 reduction, followed by other high-energy-producing states like Texas and Florida.
- Moderate contributors: States such as Illinois, Pennsylvania, and Arizona exhibit mid-range carbon reduction figures, signifying steady solar energy adoption.
- Emerging states: States positioned in the lower-left quadrant of the graph are beginning to scale up their renewable energy efforts, indicating potential future growth in CO_2 reduction.

This analysis highlights the effectiveness of solar power in mitigating carbon emissions, reinforcing its role in combating climate change.

4.2.3.3 Efficiency Ratio

The Efficiency Ratio bar chart ranks states based on how effectively they convert solar energy capacity into actual energy output. Higher efficiency values indicate better utilization of available solar resources.

- California leads the efficiency ranking, demonstrating superior optimization of its solar infrastructure.
- New York, Texas, and Florida follow as high-performing states, reflecting strong policy support and advanced solar technology implementation.
- Middle-tier states like Pennsylvania, Illinois, and Michigan show moderate efficiency levels, suggesting room for further improvements in solar panel performance and grid integration.
- Lower-ranked states may face challenges such as inconsistent sunlight exposure, grid limitations, or less efficient solar panel installations.

This comparison emphasizes the importance of technological advancements and policy measures to maximize solar energy efficiency.

4.2.4 Solar Energy Capacity, Job Growth, and Cost Analysis

4.2.4.1 Overview

This section of the solar energy dashboard presents a comprehensive overview of solar capacity, job creation, and financial considerations in different states. It provides insights into total installed solar capacity (kW), the number of jobs added in 2023, and the cost-effectiveness of solar investments over time.

4.2.4.2 Solar Capacity and Job Market Trends

The **kW-Total and Job Added Illustration** graph showcases the total installed solar capacity (represented by blue bars) and the number of jobs added in 2023 (depicted by a blue line).

- **Leading States in Solar Capacity**:
    - **California, Texas, and Florida** stand out as the top three states with the highest installed solar capacity, reflecting their strong adoption of renewable energy.
    - Other states like **Ohio, Illinois, and New York** also contribute significantly to solar capacity.
- **Job Growth in the Solar Industry**:
    - **States with high kW capacity do not always correlate with the highest job creation**.
    - States like **Arizona, Pennsylvania, and North Carolina** show spikes in job creation despite having moderate installed capacity, possibly indicating labor-intensive growth in solar infrastructure.

This data underscores the relationship between solar expansion and employment, highlighting opportunities for economic development in the clean energy sector.

4.2.4.3 Financial Impact of Solar Energy Adoption

The **financial metrics** presented at the top of the dashboard provide key insights into the cost and savings of solar energy investments:

- **Before Incentives Cost**: The average cost before applying incentives is **$33.14K**, reflecting the upfront investment required.
- **After Incentives Cost**: Once tax credits and other financial incentives are applied, the cost drops to **$23.16K**, making solar more accessible.

- **Long-Term Savings**: Over **25 years, solar energy adoption results in estimated savings of $54.43K**, demonstrating the long-term economic benefits of transitioning to solar power.

These figures highlight the importance of federal and state incentives in making solar energy financially viable for homeowners and businesses.

4.2.4.4 Solar Panel Distribution by Orientation

The panel distribution data on the right side of the dashboard provides a breakdown of installed solar panels based on their orientation:

- **Flat Panels**: **35.09K**, the most common installation type, likely for commercial or large-scale solar farms.
- **South-Facing Panels**: **29.23K**, an optimal orientation for maximizing solar energy production in most regions.
- **West and East Panels**: **26.31K** and **24.00K**, respectively, which help extend solar energy capture in the afternoon and morning hours.
- **North Panels**: **14.60K**, the least common, as this orientation generally receives the least direct sunlight.

This data highlights the strategic positioning of solar panels to optimize energy production across different geographic regions.

4.2.5 Energy Job Market

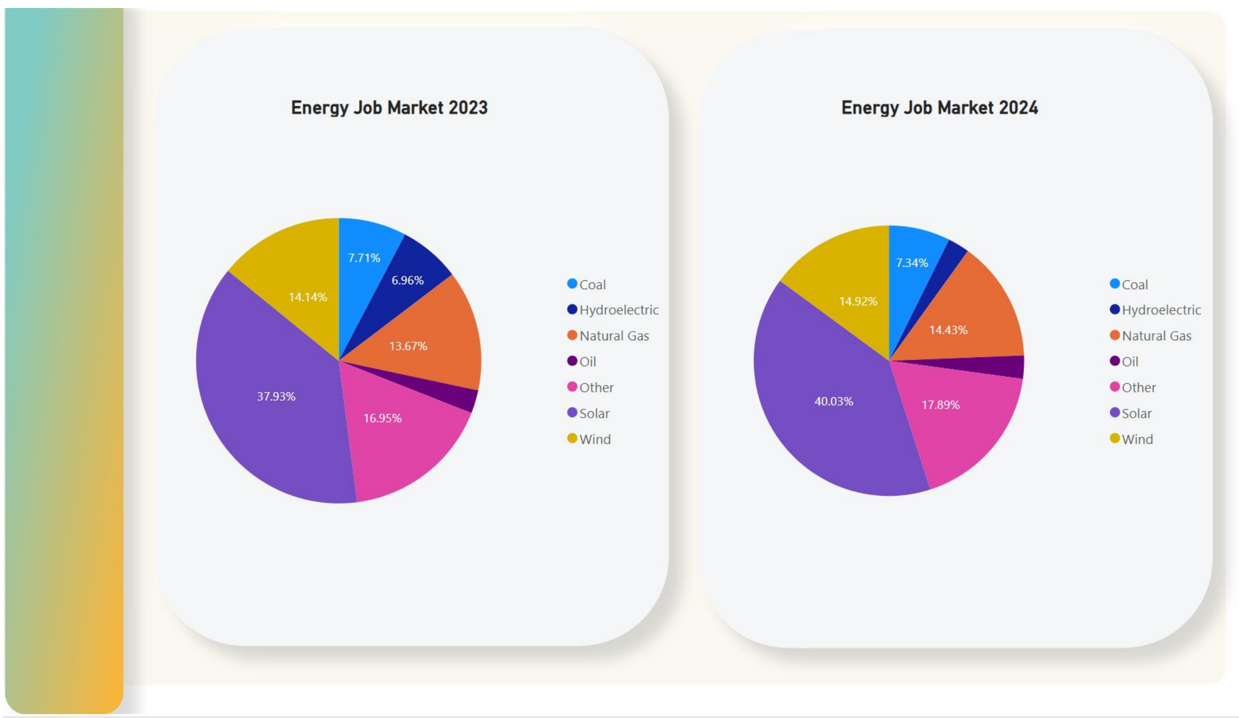
4.2.5.1 Overview

The energy job market has experienced notable shifts between 2023 and 2024, reflecting ongoing changes in the industry. A key trend observed is the growth in renewable energy jobs, particularly in the solar and wind sectors. Solar energy jobs increased from **37.93% in 2023 to 40.03% in 2024**, highlighting the continued expansion of solar power projects. Similarly, wind energy jobs saw a slight rise from **14.14% to 14.92%**, suggesting steady investment in wind infrastructure.

In contrast, fossil fuel jobs have continued to decline. Coal jobs dropped from **7.71% in 2023 to 7.34% in 2024**, indicating a reduction in coal dependency. Oil jobs also saw a decrease, moving from **13.67% to 14.43%**, reinforcing the shift toward cleaner energy alternatives. Natural gas jobs remained relatively stable, with only a small increase, showing its role as a transitional energy source.

Another notable change is the rise in jobs categorized under "Other" energy sources, which grew from **16.95% in 2023 to 17.89% in 2024**. This category likely includes emerging energy technologies such as battery storage, hydrogen, and other innovative solutions, indicating growing diversification within the energy sector.

Overall, the 2024 energy job market reflects a **clear shift toward renewable energy**, with solar and wind leading employment growth. While fossil fuel jobs continue to decline, emerging energy technologies are becoming increasingly important. This trend aligns with global efforts toward sustainability and carbon reduction, reinforcing the industry's transition to cleaner energy sources.



CHAPTER 5: APPLICATION IN VIETNAM

5.1 Identifying Optimal Locations for Solar Panel Installation

Vietnam's geography and climate conditions are very variable across different locations, and this has an influence on the feasibility of solar panel installation. Based on solar irradiance values, southern and central provinces of Binh Thuan, Ninh Thuan, and Tay Ninh get maximum intensities of sunlight around the year (~5.0-5.5 kWh/m²/day). The northern provinces of Hanoi and Lao Cai experience lesser but still suitable intensities of sunlight (~3.5-4.5 kWh/m²/day).

5.1.1 Best locations for solar energy:

South and central provinces are best positioned in solar power because they receive high solar radiation and little cloud cover. They are sun-exposed throughout the year, so it is perfect for placing the panels there.

- **Ninh Thuan and Binh Thuan:** The two provinces are arid in nature with strong solar irradiation and good government support for alternative energy developments.
- **Tay Ninh:** The province enjoys long sunshines and pro-solar policies.
- **Dak Lak and Khanh Hoa**: Both provinces show good solar prospects with increasing investment in large-scale solar farms.

5.1.2 Moderate potential areas:

The northern part of Vietnam, including **Hanoi, Lao Cai, and Quang Ninh**, experiences lower but still sufficient sunlight. **Northern Vietnam (Northeast & Northwest)** has the lowest sunshine hours (~1,600 – 1,800 hours/year) and the weakest solar radiation (~3.3 – 4.1 kWh/m²/day), making solar power only moderately viable in these areas.

Vùng	Giờ nắng trong năm	Cường độ bức xạ mặt trời (kWh/m2/ ngày)	Ứng dụng điện mặt trời
Đông Bắc	1.600 – 1.750	3,3 – 4,1	Trung bình
Tây Bắc	1.750 – 1.800	4,1 – 4,9	Trung bình
Bắc Trung Bộ	1.700 – 2.000	4,6 – 5,2	Tốt
Tây Nguyên và Nam Trung Bộ	2.000 – 2.600	4,9 – 5,7	Rất tốt
Nam Bộ	2.200 – 2.500	4,3 – 4,9	Rất tốt
Trung bình cả nước	1.700 – 2.500	4,6	Tốt

Although these locations experience seasonal variations with cloudier winters, there is still sufficient sunlight to facilitate the installation of solar panels.

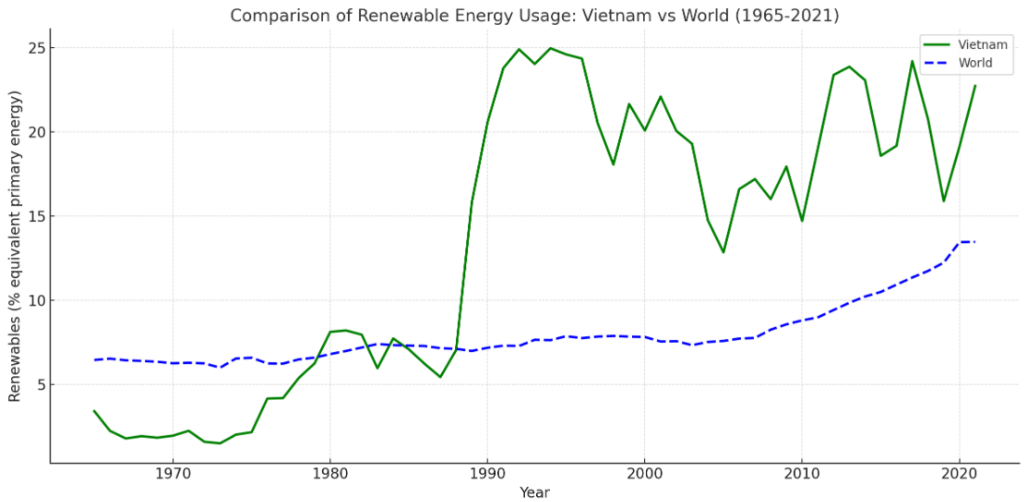
- For these locations, dual-energy solutions (solar + grid electricity) could be the optimal solution to ensure maximum energy reliability.
- Solar power can still be a significant factor in saving on the electricity bill, especially for businesses and homes that wish to offset expensive electricity bills.

5.1.3 Urban vs Rural:

- **Urban Centers (Ho Chi Minh City, Da Nang, Hanoi):** With high population density and increasing electricity needs, rooftop solar panels offer a viable means of saving money on energy. Net metering incentives and government policies also contribute to increased urban solar use. Urban centers boast solid grid networks, making grid-tied solar systems efficiently maximize energy use and diminish reliance on traditional electricity supplies. Apart from this, office complexes and industrial estates are increasingly investing in solar power to cut down operational costs as well as achieve sustainability goals.
- **Rural Areas (Mekong Delta, Central Highlands, Northern Highlands):** Where the more distant areas may lack stable or affordable grid electricity, off-grid solar power systems and hybrid solar-battery combinations can greatly increase the availability and affordability of electricity. Rural households and small enterprises often employ expensive diesel-generating sets as the origin of their power, hence to make solar an accessible, lower-cost alternative for a renewable power supply. Agricultural applications, further, such as solar-powered irrigation systems, can improve productivity while reducing the pressure on fossil-fuel-powered water pumps.
- **Challenges & Opportunities:** Whereas cities take advantage of already laid infrastructure and policy incentives, rural towns have additional up-front installation costs and more restricted avenues for financing. But with government incentives and micro-financing schemes, rural towns can enjoy cheap solar solutions, supporting Vietnam's overall energy security and reduction of carbon emissions.

5.2 Comparison between Vietnam and the world

5.2.1 Comparison of the demand of Vietnam and the world



A. Vietnamese trends in renewable energy use:

- **Period 1965-1980:** The proportion of renewable energy was very low, ranging from 2% to 5%. This was a consequence of the weak energy infrastructure and the economy mainly relying on fossil energy.
- **Period 1980-1995:** The proportion of renewable energy increased dramatically, to about 25% in the early 1990s. This was the period when Vietnam promoted hydropower development and shifted energy sources to meet economic development needs.
- **Period 1995-2010:** The proportion of renewable energy gradually decreased, from 15-20%. This was because fossil energy sources such as oil and coal expanded to meet industrialization needs.
- **Period 2010-2021:** The percentage also tends to fluctuate aggressively but lies between 20-25% because of more wind and solar power plants being established in view of Vietnam pledging to reduce greenhouse gas emissions.

B. World's trends in renewable energy use:

- Renewable energy percentage in the world was highly minimal between 1965 to 2000 and fluctuated between 6-8%.
- Since 2000, the world has experienced an improving but steady rise in renewable energy to approximately 14% in 2021 with the development of solar and wind technology.

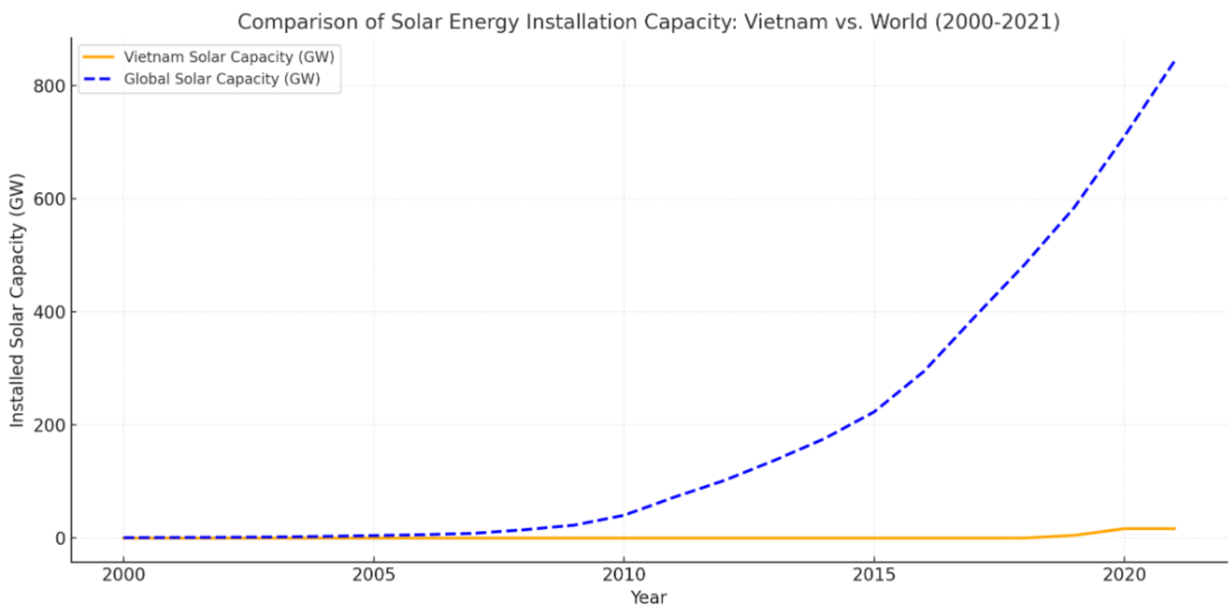
C. Comparison between Vietnam and the world:

- **Pre-1990 period:** Vietnam's percentage of renewable energy is lower than that of the world due to the limited energy infrastructure.
- **Time period 1990-2021:** Vietnam lags behind the world average, especially in times of accelerated hydropower development and clean energy expansion. But the trend of Vietnam is violently more unstable than the relative stability of the world.

D. Conclusion and development potential:

- Vietnam's renewable energy ratio exceeds the world average in most periods, reflecting the potential for exploitation from hydropower, wind power and solar power.
- To sustain and develop, Vietnam needs to improve the following factors:
  + Invest in energy storage technology to address supply fluctuations
  + Build transmission infrastructure and integrate renewable energy into the national grid
  + Promote policies to attract investment in clean energy projects.

5.2.2 Comparison of installation density in Vietnam and the world



Criteria	Vietnam	Global
Growth Trend	- Rapid growth since 2018 - Reached **15 GW** in 2021	- Steady and continuous growth since 2000 - Surpassed **800 GW** in 2021
Share in Total Energy	Contribution is increasing but still low compared to other energy sources	Solar energy holds a significant share in the total renewable energy mix
Natural Advantages	High solar radiation **(1,400-3,000 hours/year)**, concentrated in the Central and Southern regions	Countries like China, the US, and the EU have high radiation levels and early supportive policies
Support Policies	FIT (Feed-in Tariff) policy encourages investment	Governments of developed countries implement long-term policies (cost reduction, subsidies, technological research)
Development Potential	Many opportunities for expansion, especially for rooftop solar installation models	Already at an advanced stage, further development focuses on technological improvements and smart grids
Challenges	Limitations in smart grid infrastructure and energy storage capabilities	Need for further improvements to increase storage efficiency and reduce technological costs
Future Outlook	Strong growth potential in the next 5-10 years	Growth remains stable but may slow as markets reach saturation

APPENDIX

A. AI Application

AI significantly enhances SQL performance and query optimization, particularly when working with large datasets involving solar energy production and electric demand analysis.

In your solar energy project, AI can detect inefficient queries by analyzing execution plans, suggesting index usage, and highlighting redundant calculations. It can also optimize joins and aggregations, improving query efficiency for large datasets. Additionally, AI can identify outliers and data anomalies, flagging states where the solar-to-demand ratio is unusually high or low. Finally, AI ensures unit consistency by detecting and correcting mismatches (e.g., kW vs. kWh) to improve accuracy.

For example, California's unusually high energy demand might be due to incorrect scaling factors or missing commercial/industrial energy use. AI can validate these calculations to ensure accurate insights.

To effectively analyze total solar energy production vs. total energy demand per state, we need an optimized SQL query. The following query:

B. Optimized SQL Query for Energy Production vs. Demand

```sql
SELECT
    s.state_name,
    SUM(s.kw_total * COALESCE(a.ANN, 20) / 100 * 365 * 24) AS total_solar_energy_produced_kWh,
    SUM(e.`Average usage per month` * 12 * 1000) AS total_energy_demand_kWh
FROM sunlight s
JOIN electric_costs e ON s.state_name = e.State
LEFT JOIN `average percent of possible sunshine by us city` a
    ON s.state_name = a.STATE
GROUP BY s.state_name
ORDER BY total_energy_demand_kWh DESC;
```

This query converts installed solar power capacity (kW) into annual energy production (kWh) using sunlight availability. It also uses COALESCE(a.ANN, 20) to default missing sunlight data to 20% annual sunlight hours. The total energy demand is calculated using household usage data, and results are ordered by total energy demand to highlight states with the highest electricity needs.

C. Key Insights from SQL Analysis

1. California's High Energy Demand Explained

California’s total demand is significantly higher than other states due to several factors. First, it has one of the largest populations in the U.S., resulting in higher overall electricity consumption. Second, California has extensive commercial and industrial energy use, which is not always reflected in residential household data. Additionally, the state has a high concentration of electric vehicles (EVs), further increasing electricity consumption.

If commercial, industrial, and transportation-related energy usage isn't properly accounted for, total demand will appear exaggerated in comparison to solar energy production. The SQL query may need further refinement to ensure all relevant energy sectors are included.

2. States with High Solar Potential but Low Demand

Some states, such as Nevada, Arizona, and New Mexico, have extremely high solar capacity but relatively low electricity demand. These states are ideal for solar energy exports to neighboring regions with higher demand. Investing in better energy transmission infrastructure could allow for efficient solar energy distribution across state lines.

3. Outliers & Data Anomalies

Some states show unexpected results in the analysis. Oregon and California have an unusually low solar-to-demand ratio, suggesting potential unit conversion issues or missing data in the calculations. On the other hand, Rhode Island and Massachusetts show high demand with low solar production, indicating a strong potential for future solar expansion projects in these regions.

4. Potential for Future Solar Expansion

States with high electricity demand but low solar production, such as New York, Pennsylvania, and Illinois, represent significant opportunities for solar infrastructure investments. Government incentives and solar tax credits could accelerate adoption in these areas. By strategically deploying new solar projects in states with high demand, policymakers and investors can maximize the return on investment while reducing dependency on fossil fuels.