**INDIA SOLAR POWER GENERATION ANALYSIS DASHBOARD**

**A PROJECT REPORT**

Submitted by

ANIRUDH G V(22ADR006)

BAIRAVI E (22ADR013)

KABILAN C (22ADR045)

For

**20ADC33 DATA ANALYSIS**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE**

****

**KONGU ENGINEERING COLLEGE**

**(Autonomous)**

**PERUNDURAI, ERODE - 638 060**

**OCTOBER 2024**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE**

**KONGU ENGINEERING COLLEGE**

**(Autonomous)**

**PERUNDURAI, ERODE - 638 060**

**OCTOBER 2024**

Department of Artificial Intelligence

**20ADC33 - Data Analysis Project Report**

Signature of course in-charge Signature of HOD

Submitted for the Continuous Assessment viva voce examination held on\_\_\_\_\_\_\_

**EXAMINER I EXAMINER II**

**ABSTRACT**

This research provides a thorough analysis of data on solar power generation collected over a 34-day period from two solar plants in India. The purpose of the study is to look at how environmental conditions, specifically temperature and sun radiation, affect the efficiency of power generation. Power output and sensor readings, such as module and ambient temperatures, were among the data gathered at the inverter and plant levels.

In order to investigate the connections between temperature, radiation, and electricity generation, the study required cleaning, converting, and modeling the data. By contrasting the two plants' performances, important conclusions were reached. To demonstrate the results, visual dashboards that show how temperature variations affect energy output were created. Higher temperatures have a negative impact on solar power output, according to the report.

Recommendations were given to improve plant operations in light of these findings, including methods for controlling temperature and arranging panels to maximize energy efficiency. In order to maximize energy output and operating efficiency, the project provides practical insights for enhancing solar power generation under various climatic situations.

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | 3 |
| **1** | **INTRODUCTION** | 5 |
| 1.1 | INTRODUCTION | 5 |
| 1.2 | DATA COLLECTION | 5 |
| 1.3 | PROBLEM STATEMENT | 6 |
| 1.4 | BUSINESS OBJECTIVE | 6 |
| **2** | **DATA PREPARATION AND MODELING** | 7 |
| 2.1 | DATA CLEANING | 7 |
| 2.2 | DATA TRANSFORMATION | 7 |
| 2.3 | DATA MODELING | 13 |
| 2.4 | DAX FUNCTIONS | 15 |
| **3** | **DATA ANALYSIS AND INTERPRETATION** | 22 |
| 3.1 | DATA ANALYSIS | 22 |
| 3.2 | PUBLISHING DASHBOARD | 31 |
| 3.3 | INFERENCE | 32 |
| **4** | **CONCLUSION** | 35 |
| 4.1 | RECOMMENDATIONS | 35 |
| **5** | **REFERENCES** | 36 |

**CHAPTER 1**

**INTRODUCTION**

**1.1. INTRODUCTION**

* The project's main goal is to analyze data on solar power generation that was gathered from two Indian solar installations.
* To investigate the connection between environmental elements including temperature and their impact on electricity generation, data was collected over a period of 34 days.
* The initiative seeks to maximize the generation of solar energy by comprehending these factors.
* PowerBi offers a wide range of tools for preprocessing and analyzing the data in an easy way.
* The dataset collected was from “data.world” which is preprocessed before analyzing.
* The preprocessing includes data cleaning, transformation, modeling and usage of DAX functions.
* Then the data is analyzed and the analysis is published in the dashboard.
* "Power BI Services" refers to the combination of PowerBI's desktop-based PowerBI Desktop and cloud-based BI (business intelligence) services that provides data warehouses with interactive dashboards, data preparation tools, and data discovery tools.
* In March 2016, Microsoft launched Power BI Embedded, a new service on its Azure cloud platform.

**1.2. DATA COLLECTION**

* This dataset was collected from two solar power plants in India over a period of 34 days.
* It includes two main components for each plant: power generation data and sensor data. Power generation metrics were captured at the inverter level, which means each inverter was responsible for collecting data from multiple lines of solar panels.
* On the other hand, environmental sensor data was collected at the plant level using an optimally positioned sensor array.
* The purpose of combining two datasets is to get a comprehensive view of environmental conditions affecting solar power production.
* **Source:** [data.world](https://data.world/gymprathap/india-solar-power-generation-dataset/workspace/file?filename=India-Solar-Power-Generation-Dataset.zip)
* **Key features:**
  + DATA OF POWER GENERATED: Date and Time, AC power, DC power, Daily Yield, Total Yield, Source key, Plant ID.
  + DATA OF TEMPERATURE: Date and Time, Ambient temperature, Module temperature, Irradiation level, Source key, Plant ID.

**1.3. PROBLEM STATEMENT**

* The primary issue this study attempts to solve is figuring out how different environmental elements, especially temperature, affect the production of solar power.
* It aims to determine whether temperature fluctuations result in errors in the plants' electricity output.

**1.4. BUSINESS OBJECTIVE**

* Finding the ideal environmental conditions for solar energy generation is the company's goal.
* The project's goal is to increase solar power plants' efficiency, which will boost energy output and increase solar power firms' profitability.

**OVERVIEW OF POWER BI**

* Power BI is a business analytics tool from Microsoft that offers interactive data visualization.
* It lets users see and share information throughout their organization.
* Business users can use it to centralize measurements and significant company goals in order to track their progress.
* It promotes cooperation and interaction on the site while being simple to use and aesthetically pleasing.
* Users can easily visualize dynamic and interactive Reports/Dashboards by utilizing its Business Intelligence Capabilities.

**CHAPTER 2**

**DATA PREPARATION AND MODELING**

**2.1. DATA CLEANING**

* The process of filling in missing values, smearing noisy data, identifying and eliminating outliers, and smoothing noisy data is known as data cleaning.
* To make the data easier to work with, the datasets from the two plants were combined and elements such as SOURCE\_KEY, which indicates the data's source, were labeled for this project.
* The two solar plants' data were first divided. A unified analysis was made possible by the merging of these datasets. This combination made it easier to compare plants (identified by PLANT\_ID) and made analysis easier overall.
* There may have been errors or gaps in the data because it included real-time readings from several sensors. These problems were resolved by data cleaning, which included standardizing units where needed, eliminating noisy data, or filling in missing values.

**2.2. DATA TRANSFORMATION**

* Converting data from one format or structure to another is known as data transformation, and it is crucial for data management tasks like data warehousing, data wrangling, and application integration.
* Data profiling, visualization, and purification are some of the tools used in this process, which can be basic or sophisticated and involve both automatic and manual stages.
* Businesses today frequently transform massive datasets for apps, warehouses, and repositories because of developments in these tools.
* In order to facilitate efficient data analysis and the development of DAX functions to reveal insights, data transformation is an essential preprocessing step for datasets.
* The date and time data were kept apart since time-series analysis was necessary to look at energy output over different time periods (daily, weekly, etc.). Particular time-based analysis, including peak solar irradiation hours, daily and weekly patterns, and seasonal trends, was made possible by this separation.
* In order to guarantee compatibility and prevent mistakes, temperature and irradiation values were standardized. This standardization made it possible to apply modeling functions consistently and compare plants accurately.
* To improve the data, more features were extracted. For example:
  + Calculating the difference between the module temperature—the temperature of the solar panels—and the ambient temperature—the temperature of the outside world—helped ascertain how temperature affected efficiency.
  + Irradiation Values: To evaluate the effects of solar exposure on power generation, irradiation data was converted to provide cumulative values, peak values, and averages across time.

**PROCEDURE**

## **STEP 1**

1. Go to HOME tab in ribbon.
2. Click on GET DATA and select data from the system or from any platform where it resides.
3. Here select 4 different tables of CSV format from system and load it to POWER BI.

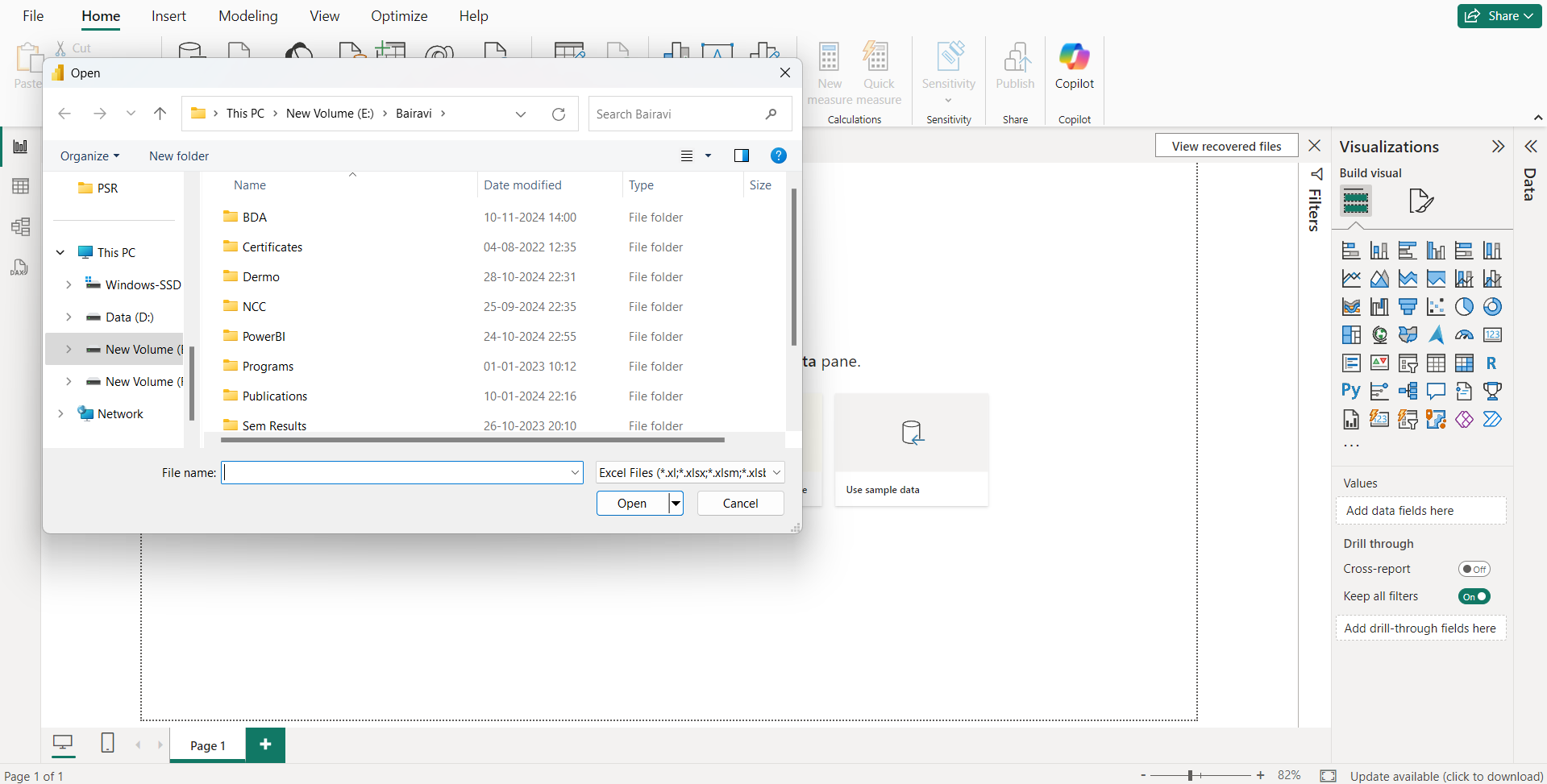


Figure 2.2.1 Select Dataset

## **STEP 2**

1. From the ribbon of HOME tan select TRANSFORM DATA in order to clean and transform data.

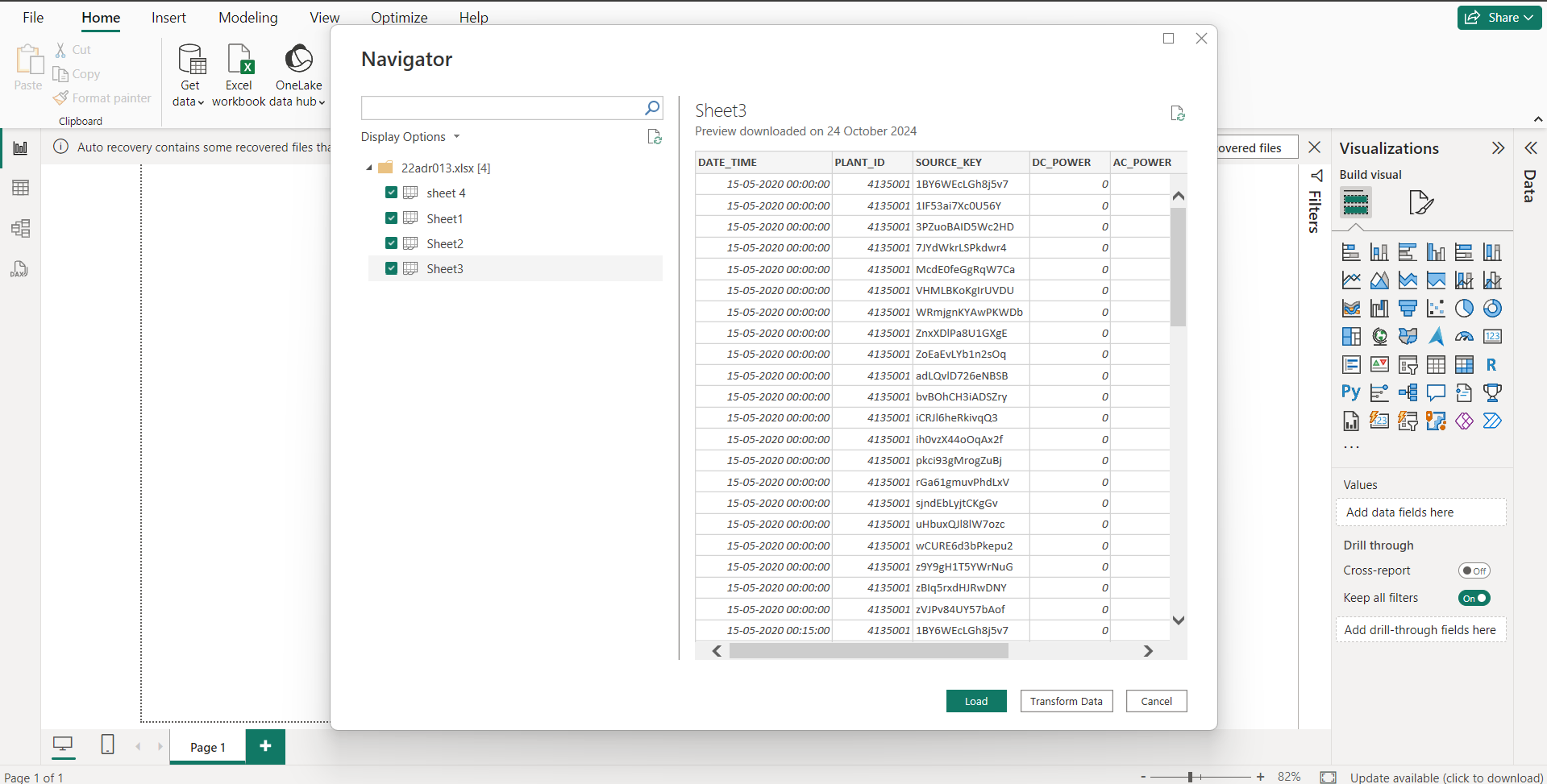
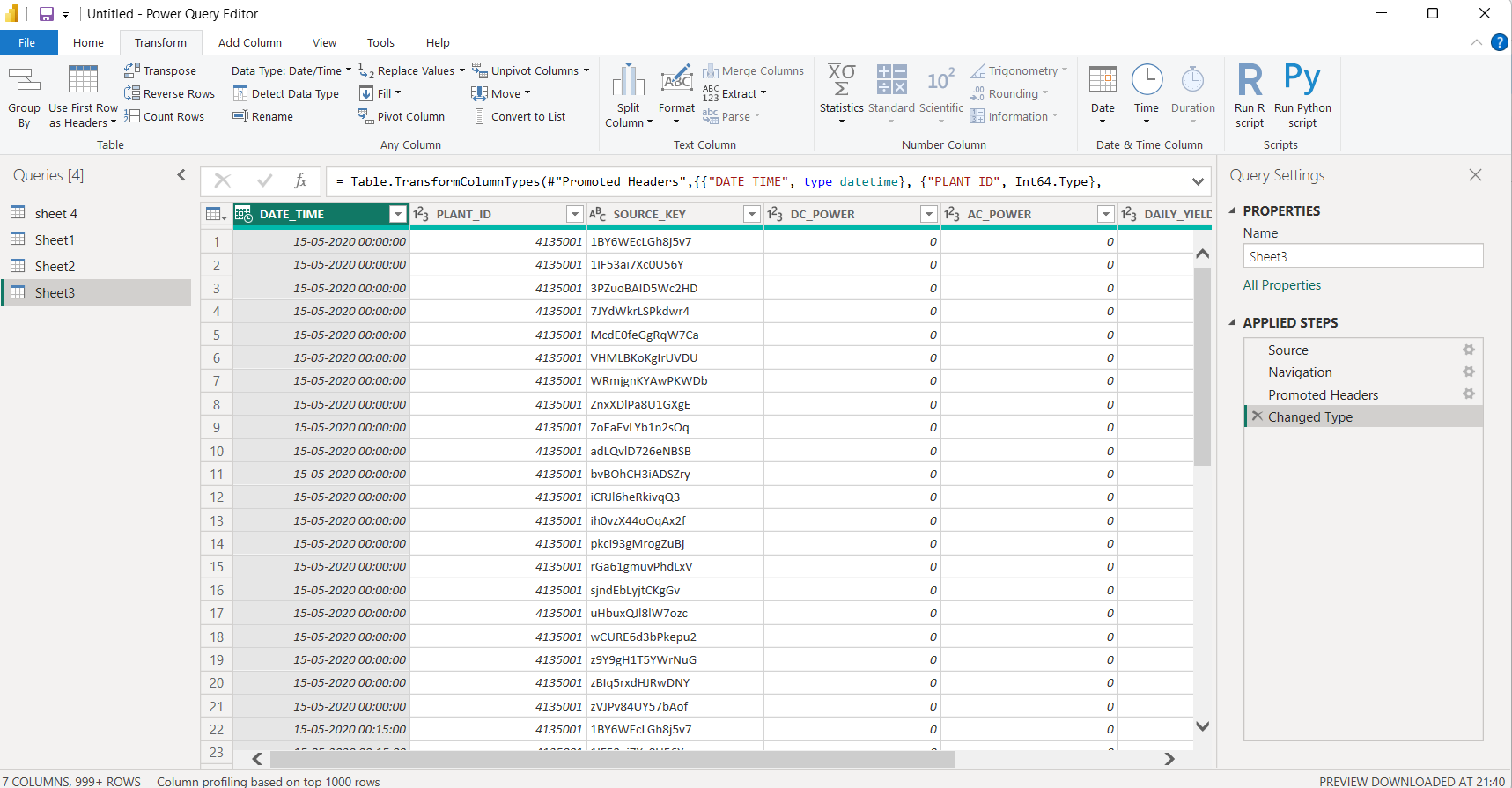


Figure 2.2.2 Transform Data

## **STEP 3**

1. After choosing transforming data all the loaded tables and opened in POWER QUERY EDITOR, so that we can make any changes as per our wish.
2. Then open the table and replace the values which are blank.
3. Then try to add NULL values to the rows with no value.

Figure 2.2.3 Power Query Editor

## **STEP 4**

1. Then on the same table apply REPLACE VALUES.
2. In this select any column that needs new values to be replaced for further processing.

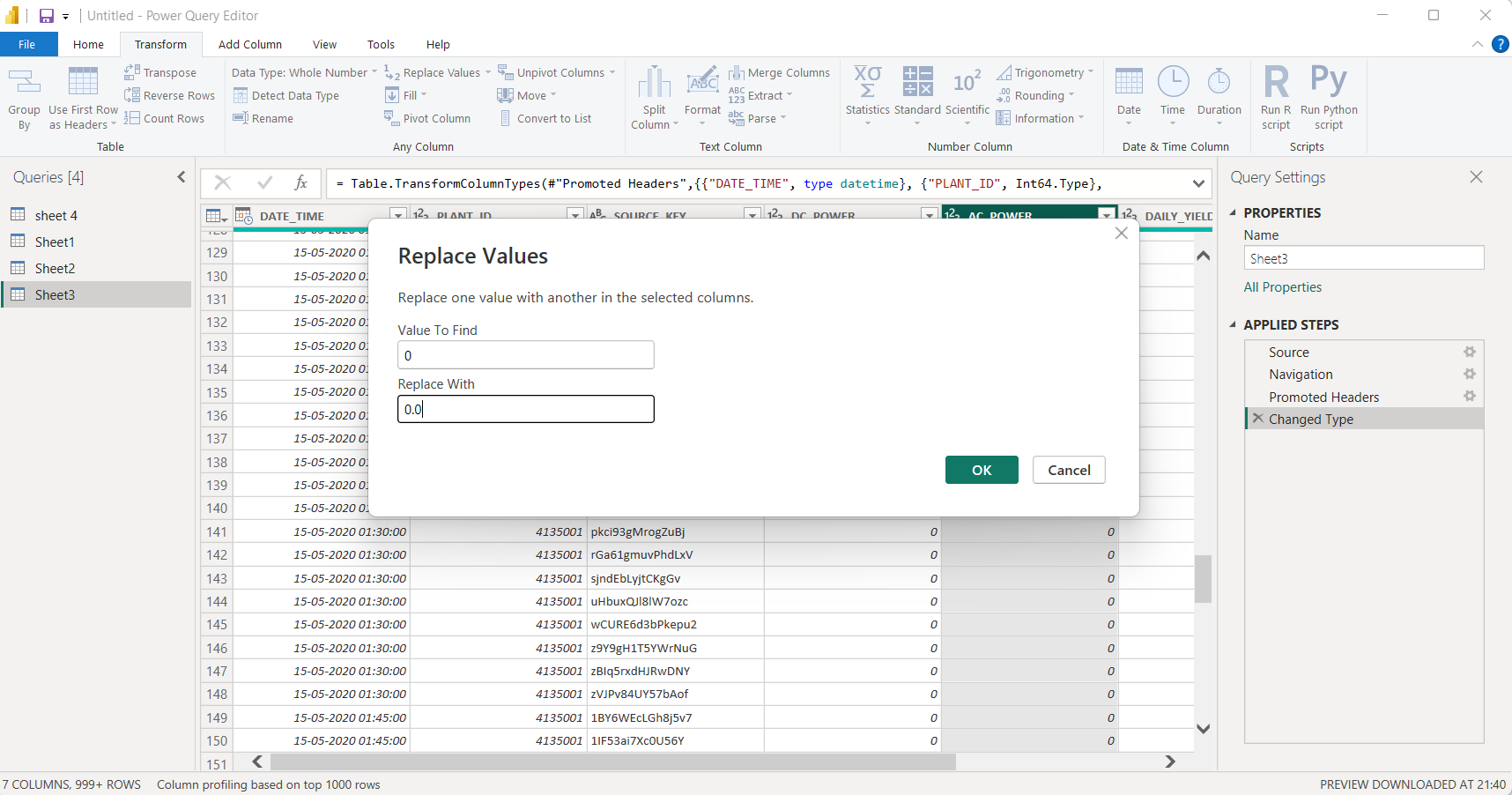


Figure 2.2.4 Replace Value

## **STEP 5**

1. Here apply change data type so click the column that needed to change the data type.

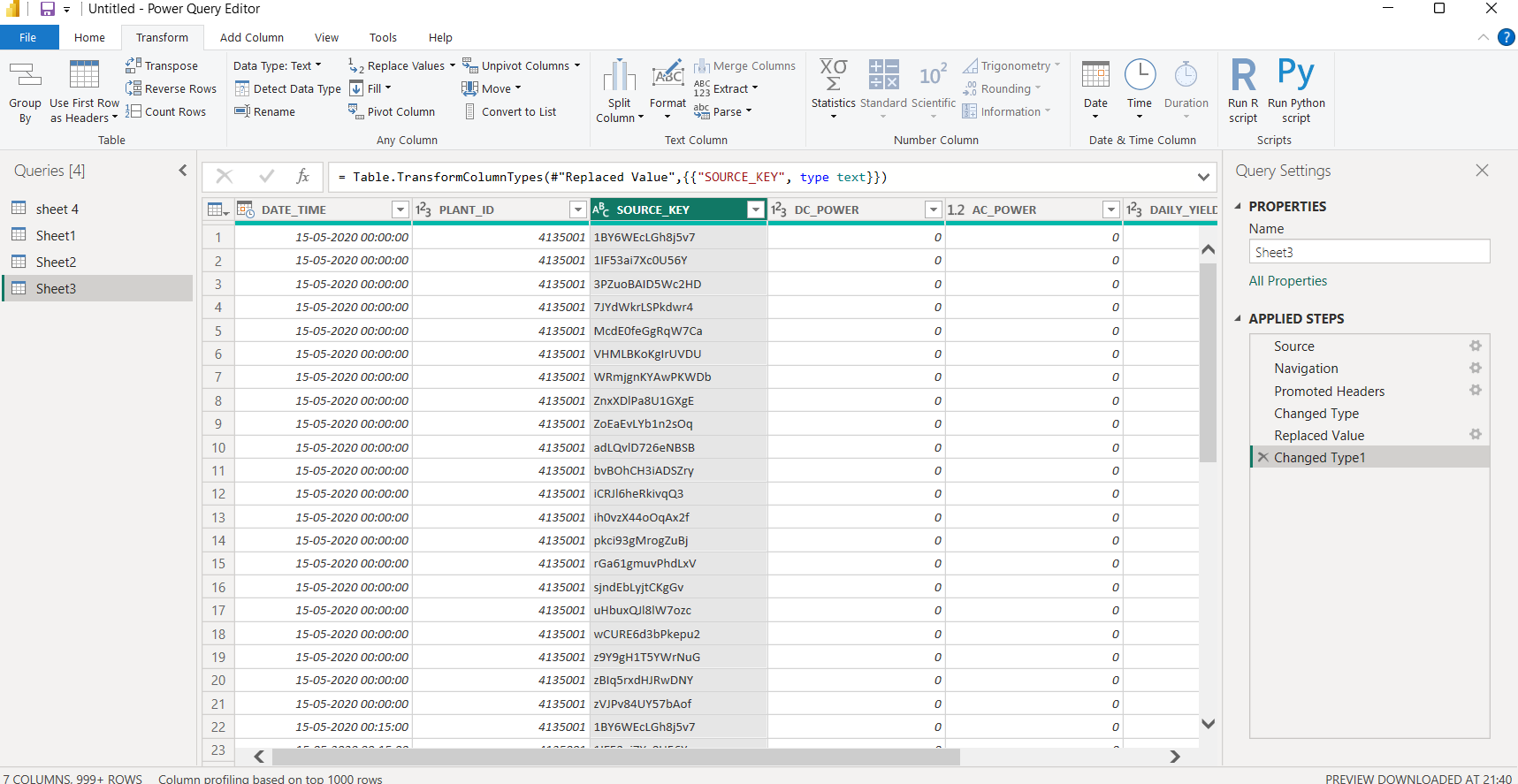
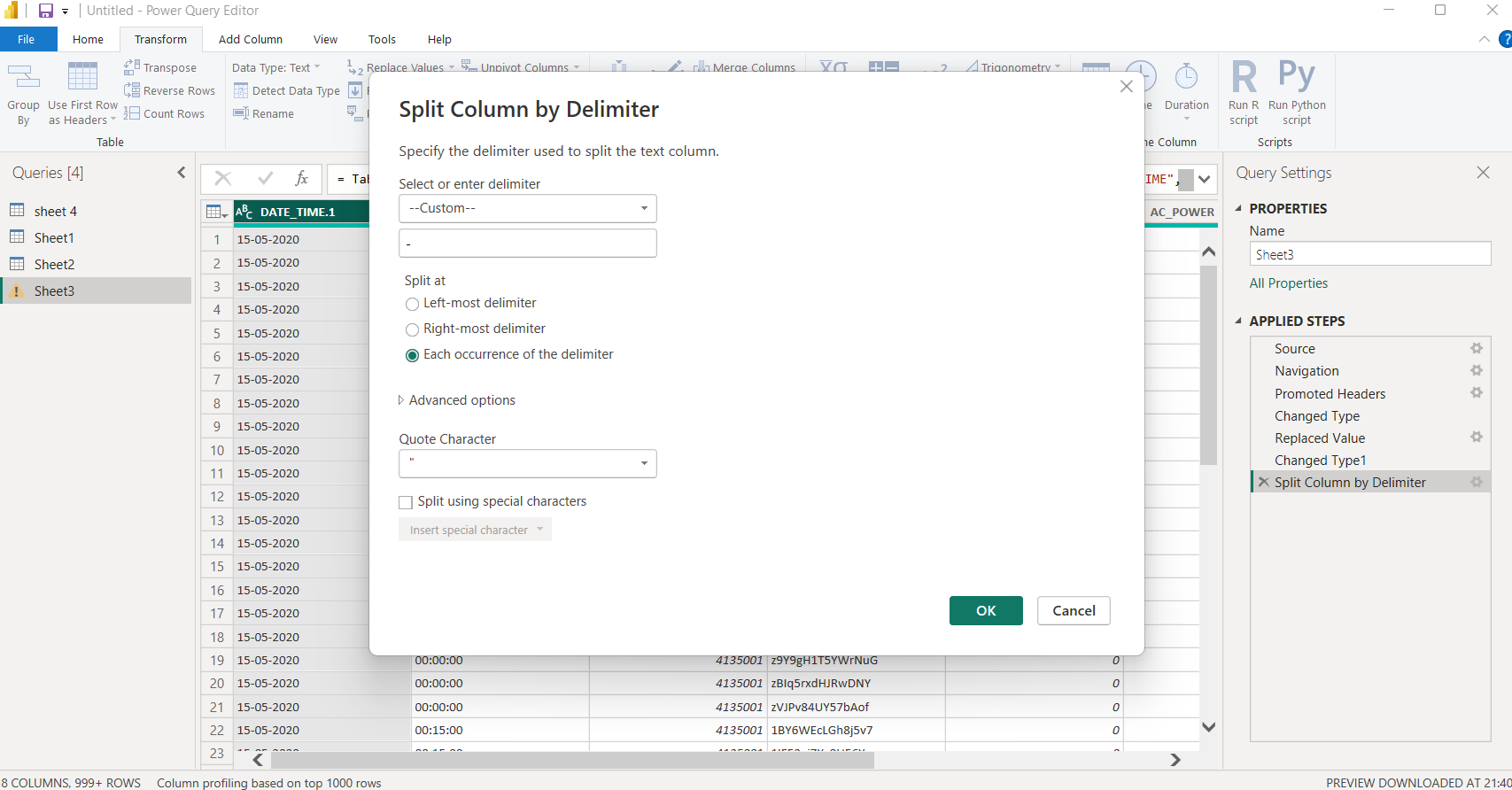
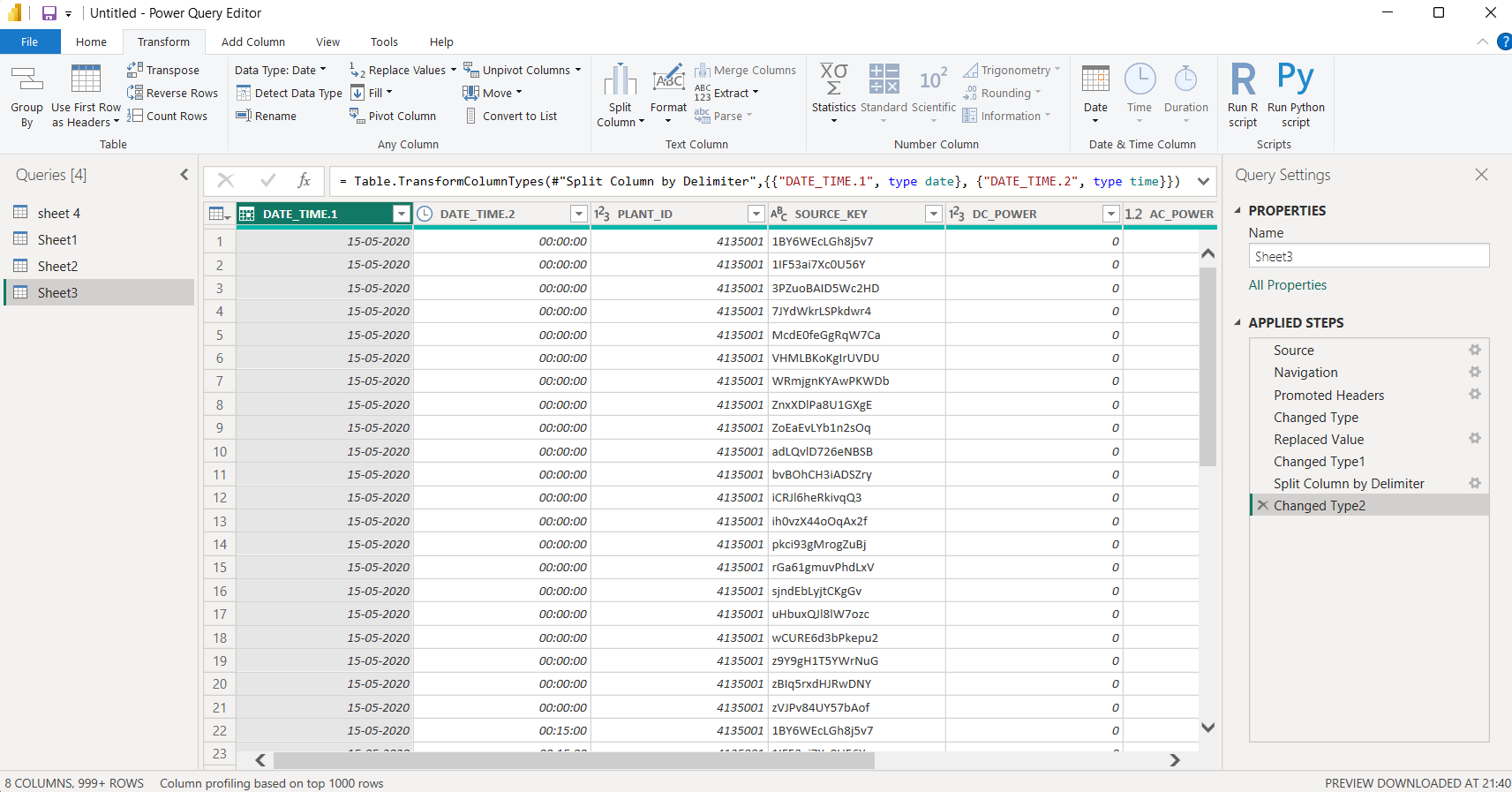


Figure 2.2.5 Change Column Type

## **STEP 6**

1. Now select the table in order to apply SPLIT COLUMN.
2. It is needed to split a particular column so that data can be accessed easily.
3. Select column to be split, RIGHT CLICK the column.
4. Drop down the list displays and select SPLIT COLUMN.
5. Then select split column by DELIMITER.
6. Similarly, we can also split column by using delimiters such as comma, colon, semi solon, hyphen, etc...
7. Now split the column “DISCRIPTION” from SUMMARY table into three different columns by using delimiter “space”.
8. Then rename the newly created columns as “YEAR”, “MONTH”, “DATE” and “TIME”.

Figure 2.2.6 Split Column Date\_Time

Figure 2.2.7 Split Column by Delimiter

## 

## **STEP 7**

1. Rename the split column.

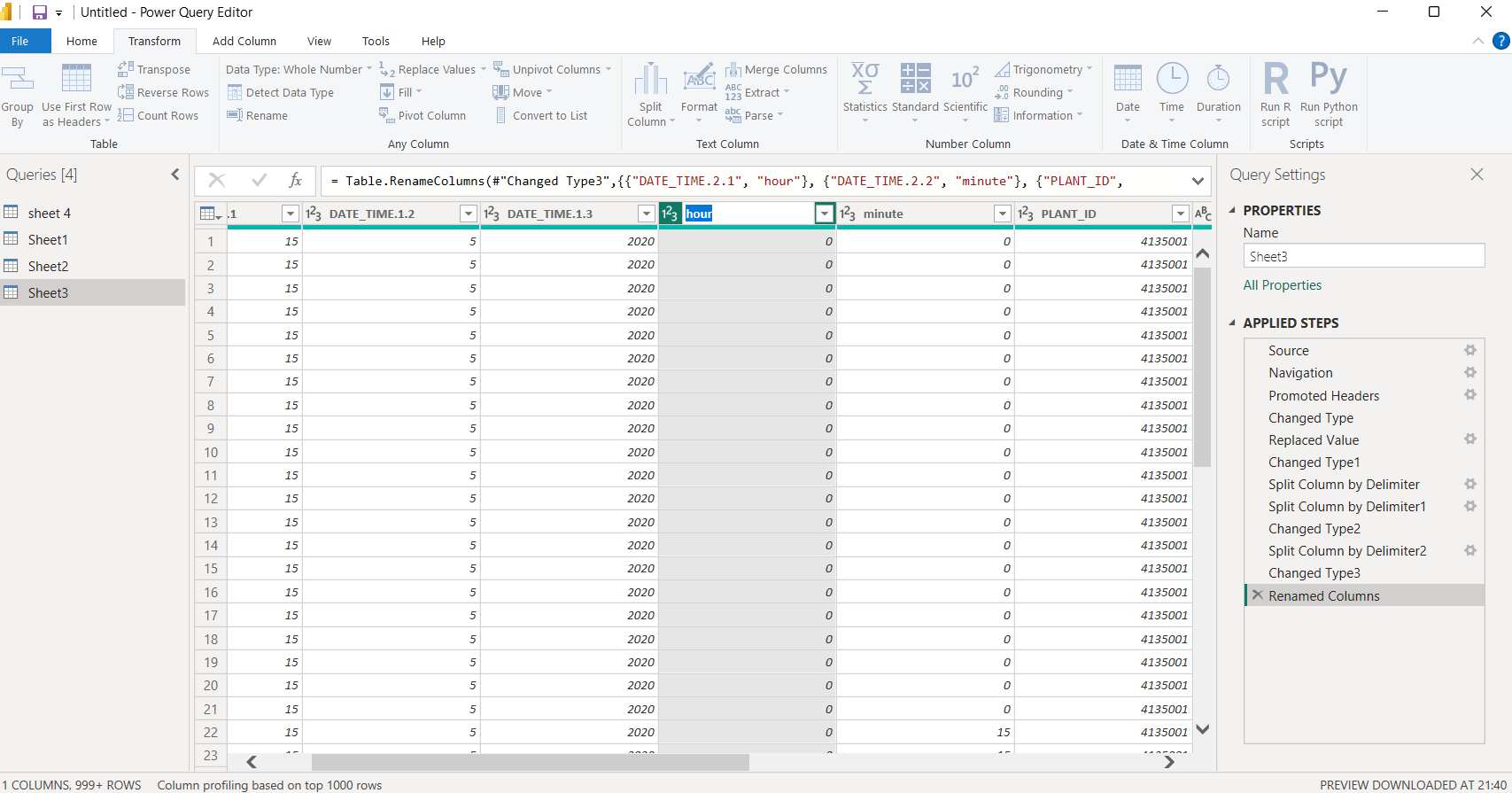


Figure 2.2.8 Rename Column

**2.3. DATA MODELING**

* One feature of BI tools that is used to create relationships between different data sources is data modeling. By identifying the relationships between multiple data sources, you may create captivating data visualizations.
* With the modeling feature, it may generate original calculations on the pre-existing tables, and these columns can be readily shown in Power BI visualizations. This makes it possible for businesses to develop new metrics and carry out special computations for them.
* In order to retrieve and visualize the data from several tables, data modeling is utilized to establish relationships between them. We can establish four different kinds of relationships, including:
  + One to One relationship
  + One to Many relationship
  + Many to One relationship
  + Many to Many relationship
* Comparative Modeling: Using parameters like temperature and radiation, models were developed to compare the two plants. The team was able to determine which environmental elements had the most effects on power generation by examining each plant both individually and collectively.
* Time-Series Models: These models looked at patterns and trends in climatic conditions and power generation over time. For example, daily maxima in temperature and irradiation patterns were examined using time-series analysis.
* Efficiency Predictive Modeling: Predictive models calculated the effects of different temperatures on solar power output efficiency. These models were essential for determining the circumstances that led to a decrease or increase in electricity generation**.**

**PROCEDURE**

**STEP 1:**

1. Here start merging of columns to create relationship.
2. Select the tables then merge them by using common attribute called as “Plant ID” of both tables which act as primary key.

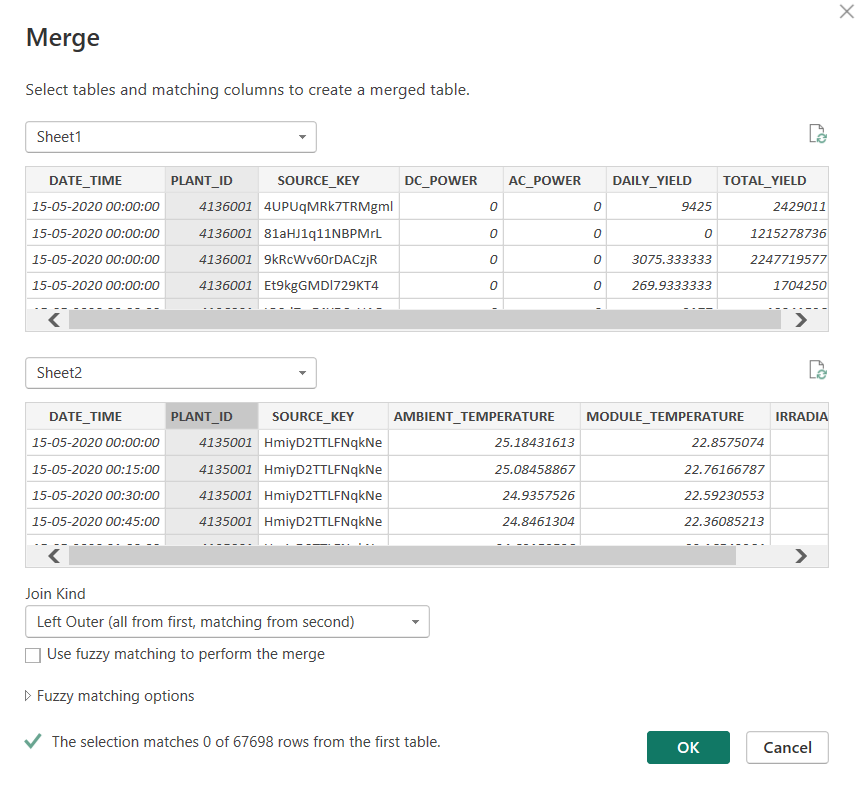


Figure 2.3.3 Merge tables “SHEET1” and “SHEET2”

## **STEP 2:**

* 1. After applying changes and merging tables select “close” and “apply” from ribbon of Power query Edito.
  2. At Power bi desktop it displays table as visualized below Figure 2.3.4.

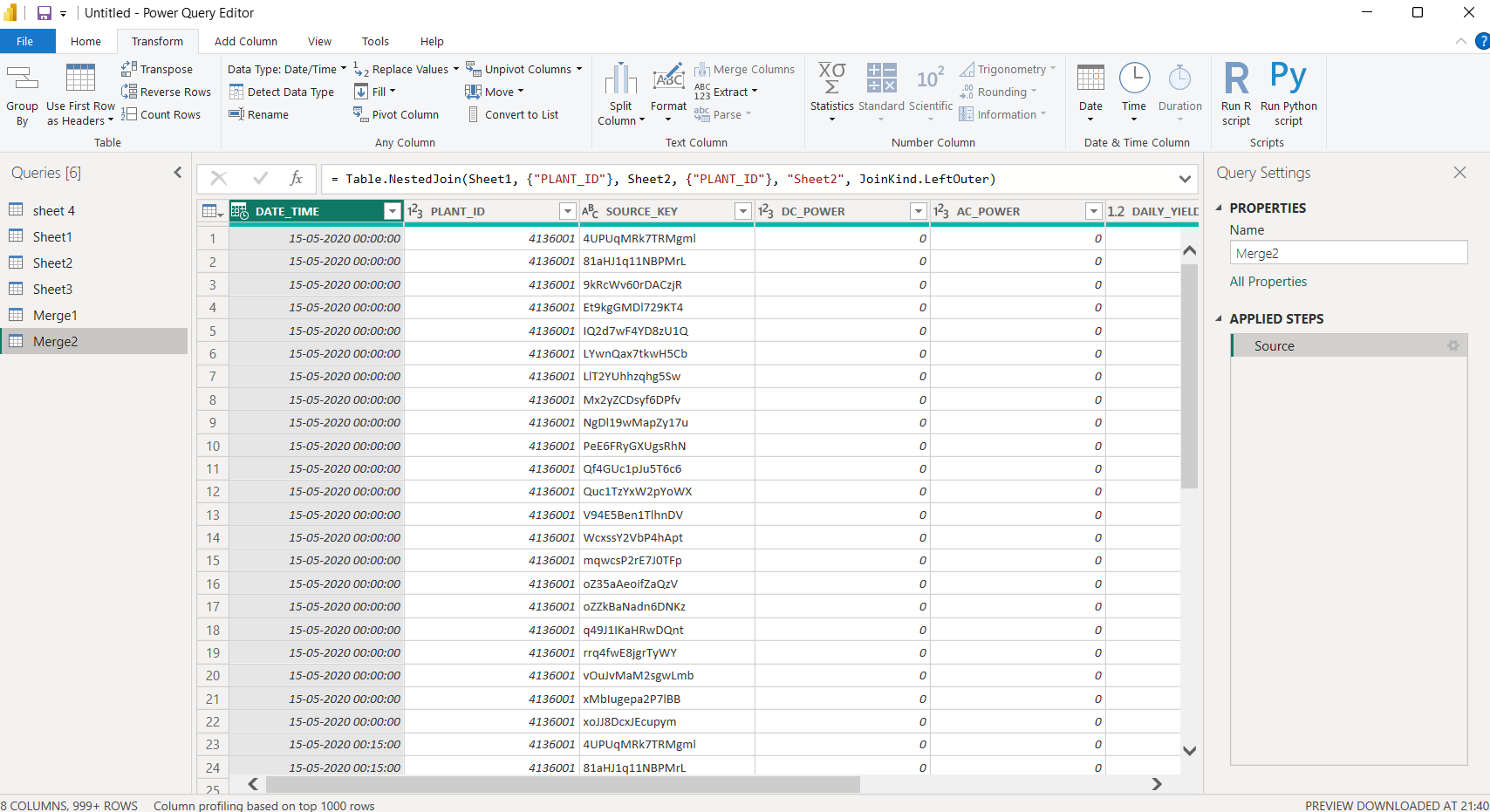
****

Figure 2.3.4 Power Bi desktop after transformation.

**2.4. DAX (DATA ANALYSIS EXPRESSION)**

* DAX in Power BI is a specialized function set with operators, formulas, and expressions that calculate and process data within existing tables, creating new insights from existing data.
* DAX enables users to define calculated columns, new measures, custom tables, quick measures, and implement time intelligence functions (for time-based columns).
* With DAX, users can quickly generate values like max, min, average, count, sum, and apply filters, as well as perform calculations like totals, variances, percentages, and basic arithmetic.
* In this project, the DAX queries that may be included are as follows:
  + **Aggregations**: DAX functions aggregate key metrics (like cumulative irradiation and temperature averages) over specific time frames, such as daily or weekly intervals.
  + **Conditional Calculations**: DAX enabled conditional calculations to identify and flag instances where certain conditions were met, such as exceptionally high or low power generation.
  + **Comparative Analysis**: Using DAX, the project team compared the two plants side-by-side, allowing for direct comparison of environmental factors like ambient temperature, module temperature, and irradiation levels to assess their impact on AC power generation.

## **PROCEDURE**

## **STEP 1:**

1. Creating Quick measure for table Summary.
2. Click Quick measure at ribbon and a menu pop up.
3. Select the calculation part.
4. Click the necessary parameter.

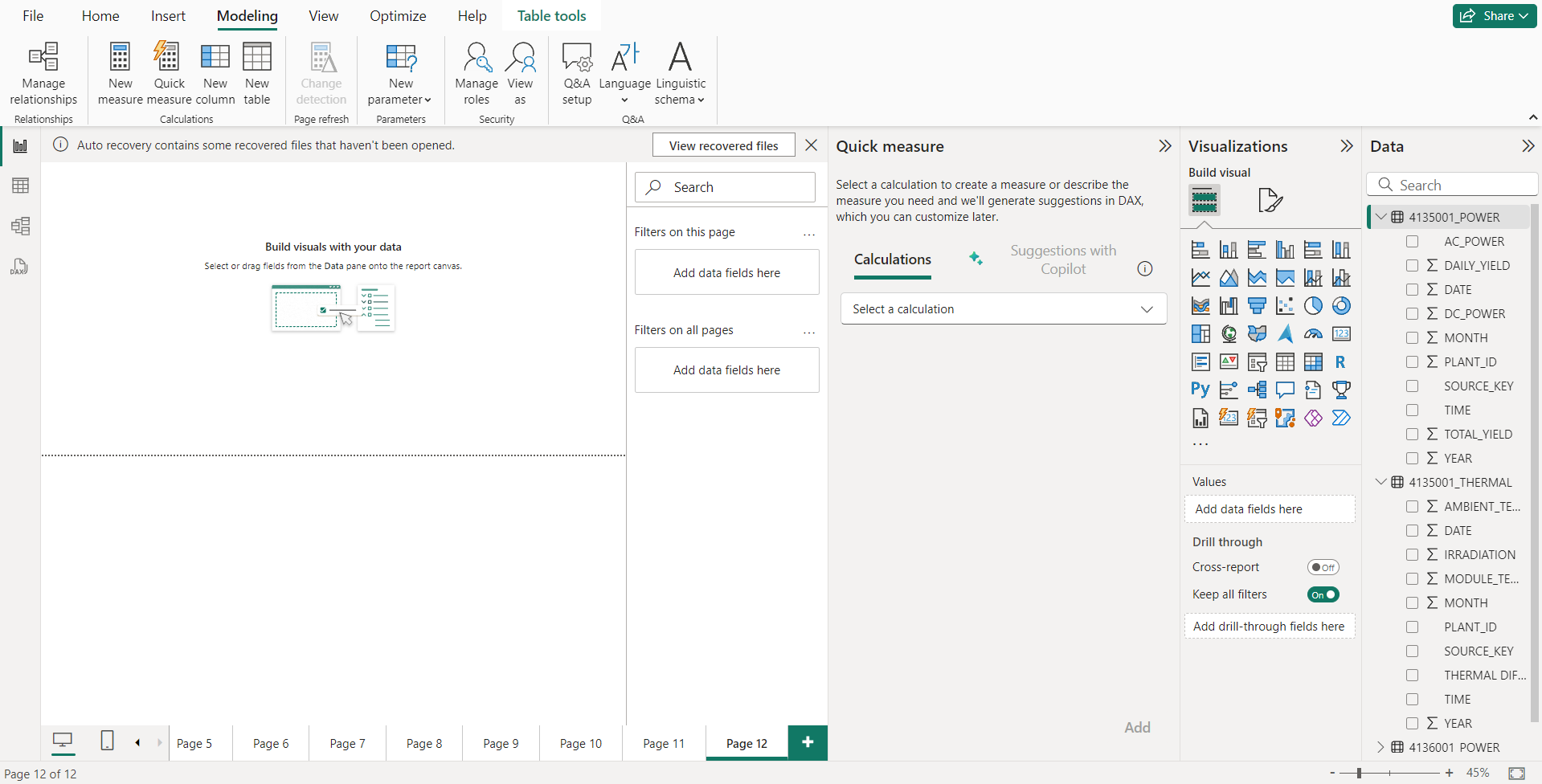


Figure 2.4.1 Quick Measure calculation

## **STEP 2**

1. Create a calculated column.
2. Select table,” Right click” it then selects “New column”.
3. Apply formulae for new table in the given box with new name click enter Formula for new column:

Average\_Daily\_Yield\_Plant\_1 =

AVERAGEX(

SUMMARIZE(

'Plant\_1\_Generation\_Data',

'Plant\_1\_Generation\_Data'[DATE\_TIME],

"Daily\_Yield", SUM('Plant\_1\_Generation\_Data'[DAILY\_YIELD])

),

[Daily\_Yield]

)

Average\_Daily\_Yield\_Plant\_2 =

AVERAGEX(

SUMMARIZE(

'Plant\_2\_Generation\_Data',

'Plant\_2\_Generation\_Data'[DATE\_TIME],

"Daily\_Yield", SUM('Plant\_2\_Generation\_Data'[DAILY\_YIELD])

),

[Daily\_Yield]

)

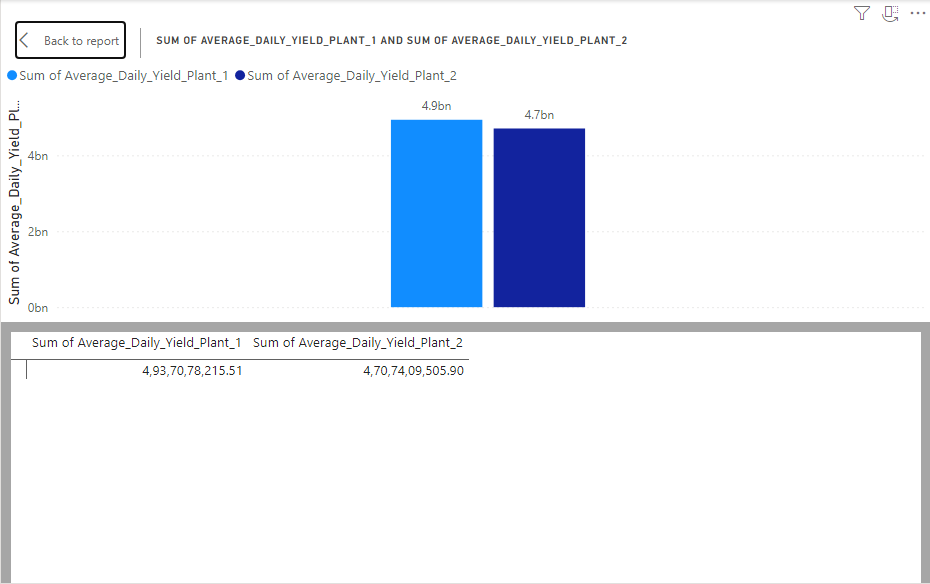


Figure 2.4.2 Average of daily yield for both the plants

## 

## **STEP 3**

1. Create new measure.
2. Select table,” Right click” it then selects “New column”.
3. Apply formulae for new table in the given box with new name click enter Formula for new column:

Total\_AC\_Power\_Plant\_1 = SUM('Plant\_1\_Generation\_Data'[AC\_POWER])

Total\_DC\_Power\_Plant\_1 = SUM('Plant\_1\_Generation\_Data'[DC\_POWER])

Power\_Efficiency\_Plant\_1 = DIVIDE([Total\_AC\_Power\_Plant\_1], [Total\_DC\_Power\_Plant\_1],0)

Total\_AC\_Power\_Plant\_2 = SUM('Plant\_2\_Generation\_Data'[AC\_POWER])

Total\_DC\_Power\_Plant\_2 = SUM('Plant\_2\_Generation\_Data'[DC\_POWER])

Power\_Efficiency\_Plant\_2 = DIVIDE([Total\_AC\_Power\_Plant\_2], [Total\_DC\_Power\_Plant\_2],0)

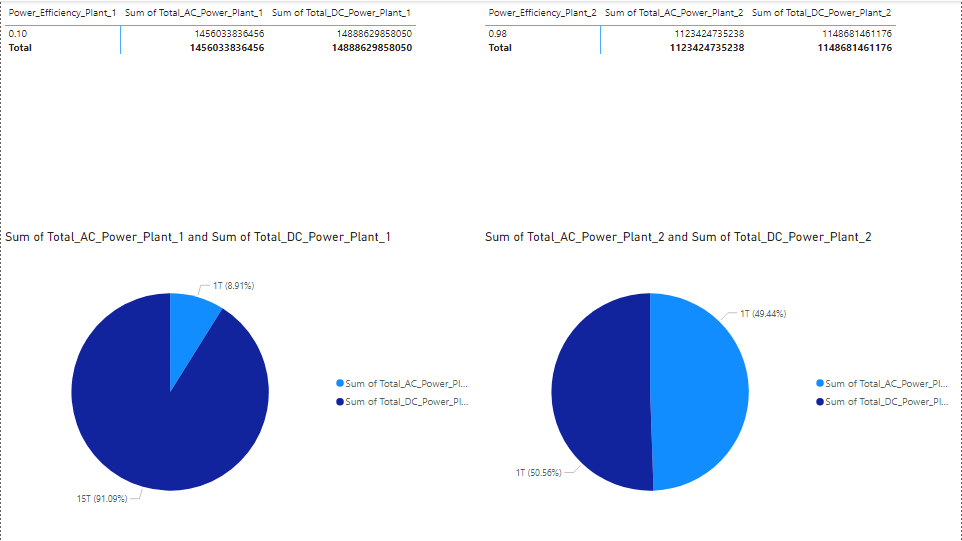


Figure 2.4.3 Average of daily yield for both the plants

## 

## **STEP 4:**

1. Create new measure.
2. Select table,” Right click” it then selects “New column”.
3. Apply formulae for new table in the given box with new name click enter Formula for new column:

Average\_Irradiation\_Plant1 = AVERAGE(‘Plant\_1\_Weather\_Sensor\_Data’[IRRADIATION])

Average\_Irradiation\_Plant\_2 =

AVERAGE(‘Plant\_2\_Weather\_Sensor\_Data’[IRRADIATION])

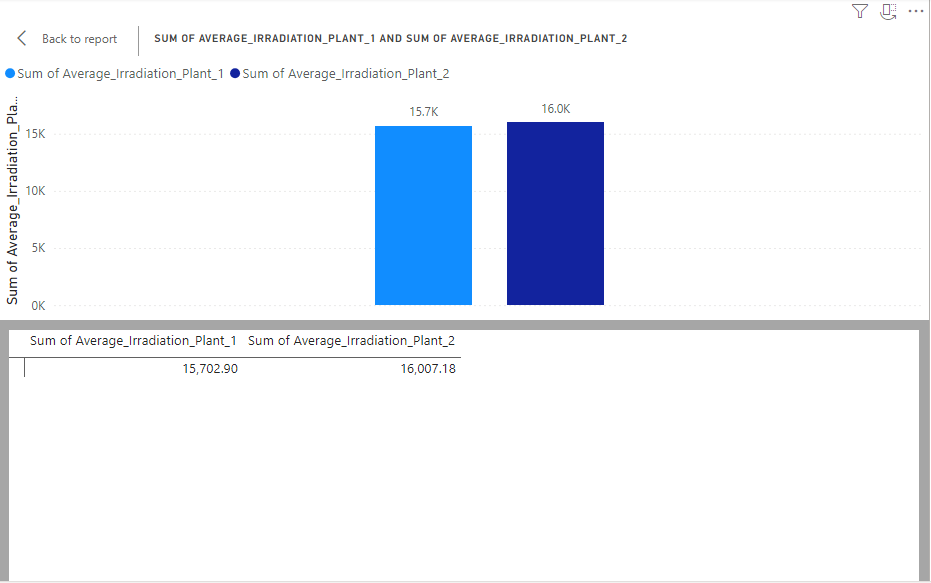


Figure 2.4.4 Average of Irradiation values of two plants

## 

## **STEP 5:**

1. Now create relationships for the various tables that we used.
2. From left ribbon select MODEL to create relationships.
3. All the tables of model will be displayed here.
4. For each table use Primary key or foreign key to create relationships either one to one, one to many, many to one, many to many respectively.

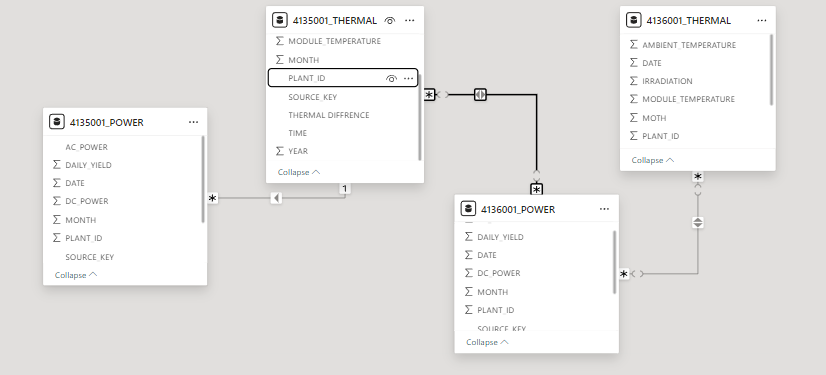
****

Figure 2.4.5 Relationship between the tables.

**CHAPTER 3**

**DATA ANALYSIS AND INTERPRETATION**

**3.1. DATA ANALYSIS**

* Data analysis is the process of examining, modifying, and tracking raw data in order to transform it into meaningful information.
* Utilizing data insights facilitates the process of making the decisions required for a business or company's expansion.
* To manage a data-driven company, in-depth data analysis is essential. Then, it is necessary to find studying various Power BI data analysis techniques both interesting and practical.
* The following outcomes are included in data analysis:
  + Utilized to generate a variety of charts using Power Bi images
  + Choose data from different tables, analyze it, and create graphics out of it.
  + Determine the outcome or ultimate answer based on the results of the analysis.
* In this project the team has analyzed how factors like ambient temperature influence the temperature of the solar panels and how both variables affect solar irradiation. The analysis identified peak times for solar energy generation and compared the efficiency of the two plants based on temperature and irradiation data.

**CHARTS:**

* Data analysis has been conducted on the solar plant power and thermal datasets.
* The analysis includes visualizations, DAX queries, and data modeling.
* A total of 19 questions were addressed, with corresponding charts and cards prepared.
* Each question is explored using specific visualization tools such as stacked column charts, pie chart, area chart, etc,.
* The results were then interpreted, and the final dashboard was published.

1. Split the data and time separately.

1. Select the column “DATE\_TIME”.
2. Select Split Column option.
3. Select the delimiter to split the column and click OK.

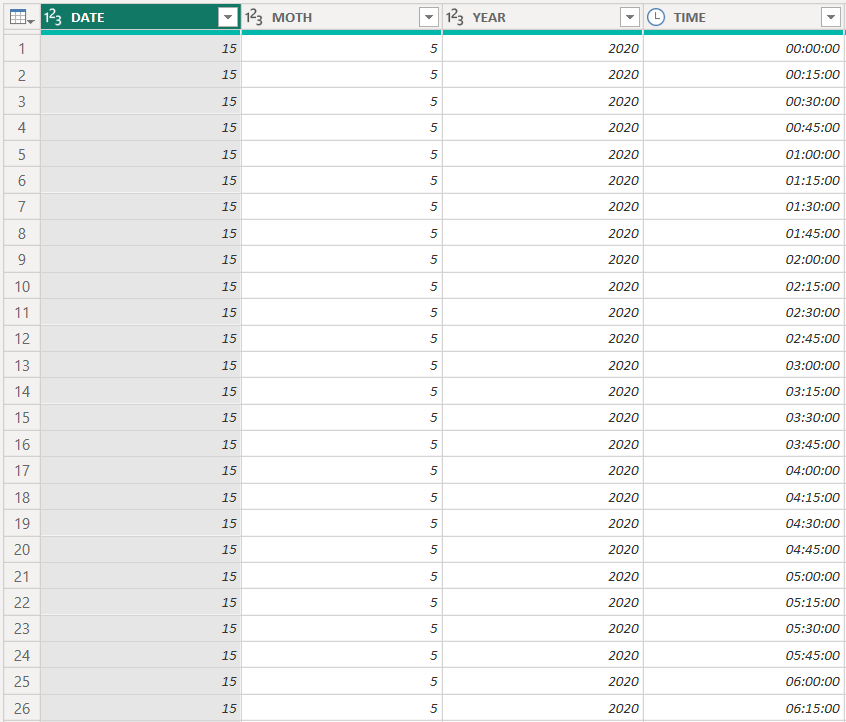


Figure 3.1.1 Split Column

2. Find the thermal differences.

1. Select the Add Column option and Select Custom Column.
2. Provide the formula for the column to fill the value.

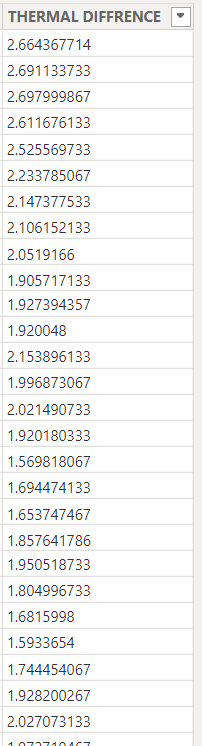


Figure 3.1.2 Boundaries by Home Captains

3. Analyze how ambient temperature affects module temperature.

1. Select the Stacked Area chart.
2. Provide “MODULE\_TEMPERATURE” for y-axis and “AMBIENT\_TEMPERATURE” for x-axis.

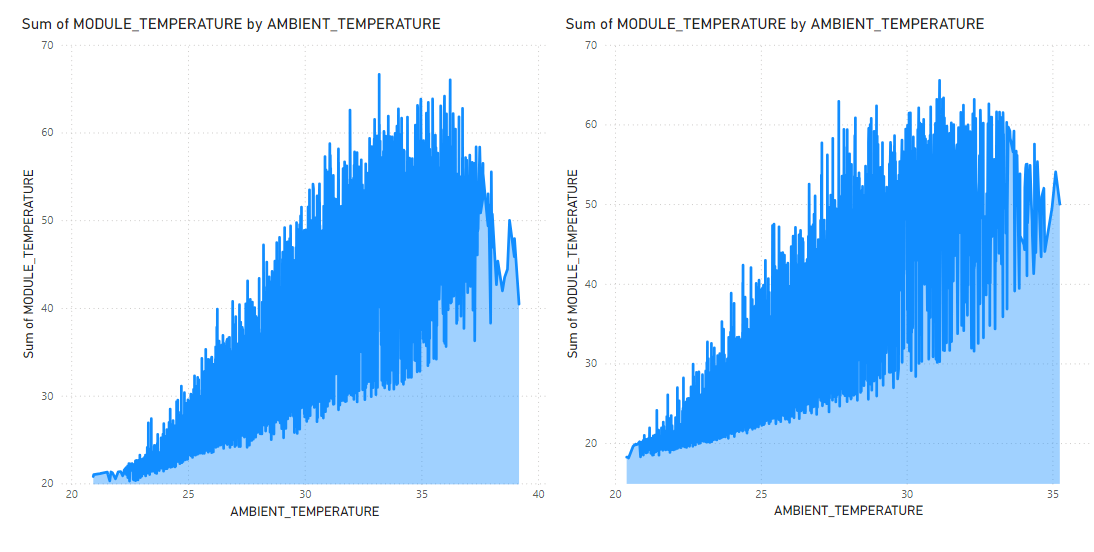


Figure 3.1.3 Module\_Temperature By Ambient\_Temperature

4. Find peak irradiation times.

1. Click the Line Chart symbol.
2. Provide the fields as “IRRADIATION” and “time”.
3. Plot the visualization.

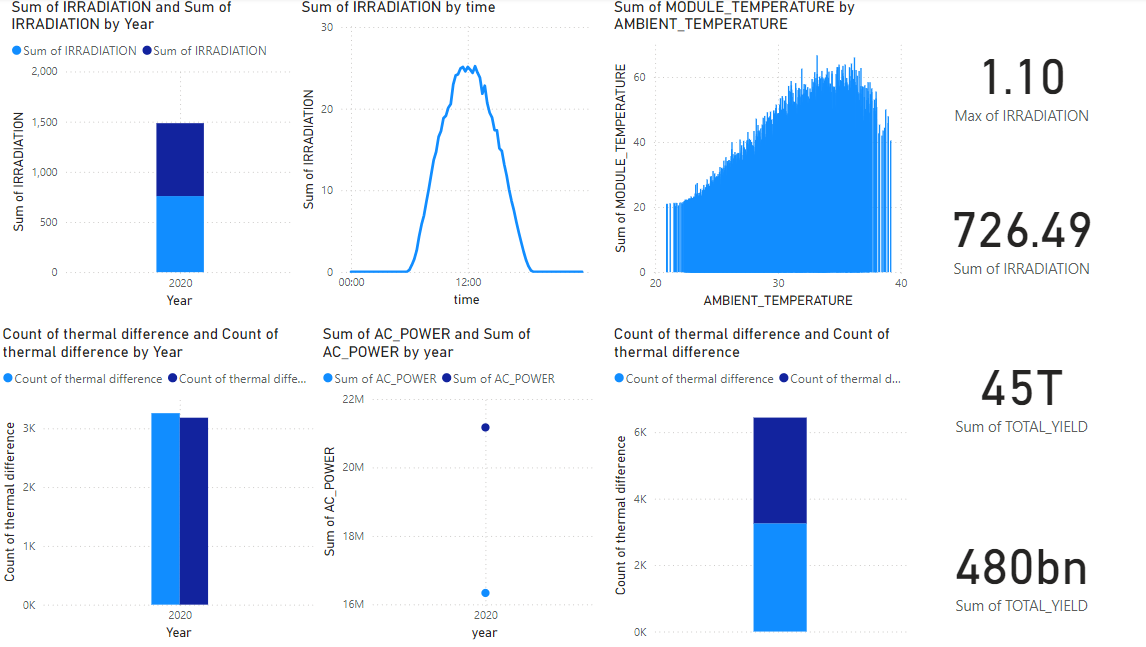
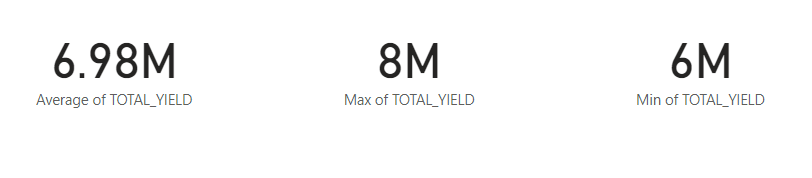


Figure 3.1.4 Peak of Irradiation by Time

5. Calculate average, max, and min temperatures.

1. Select the card.
2. Select the fields that you need.
3. Select the parameters such as average, max or min to get the visualization.

****Figure 3.1.5 Average, maximum and Minimum of Total\_yield

6. Calculate average, max, and min irradiation values.

1. Select card.
2. Provide the fields as per given.
3. Select the option for the final visualization.



Figure 3.1.6 Average, max and min values.

7. Compare different plants (using PLANT\_ID) to see which has higher ambient temperatures.

1. Select the donut chart.
2. Select the fields “AMBIENT\_TEMPERATURE” for both the plants.

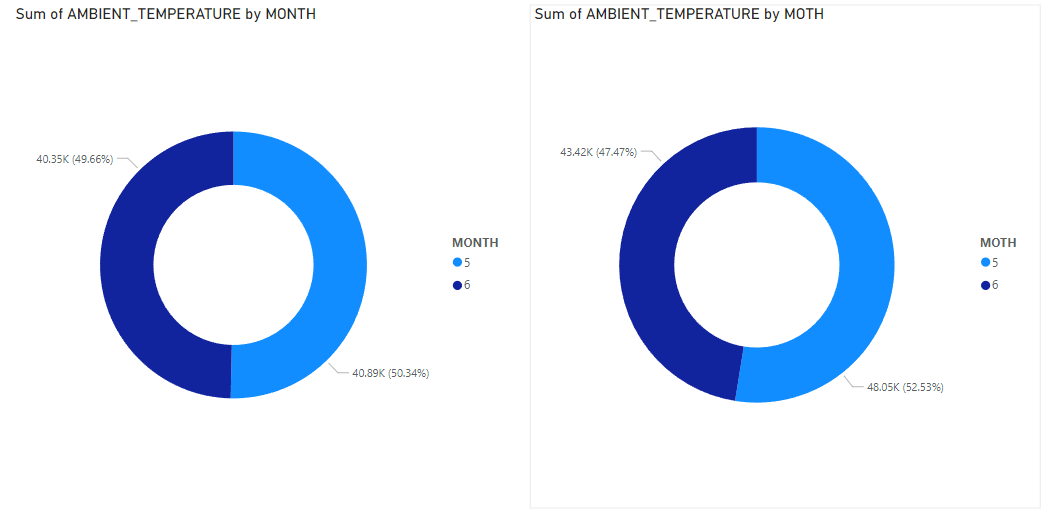
****

Figure 3.1.7 Ambient temperature of two different plants

8. Compare different plants (using PLANT\_ID) to see which has higher irradiation levels.

1. Select the stacked column chart.
2. Select the fields “IRRADIATION” and “MONTH” for visualization.

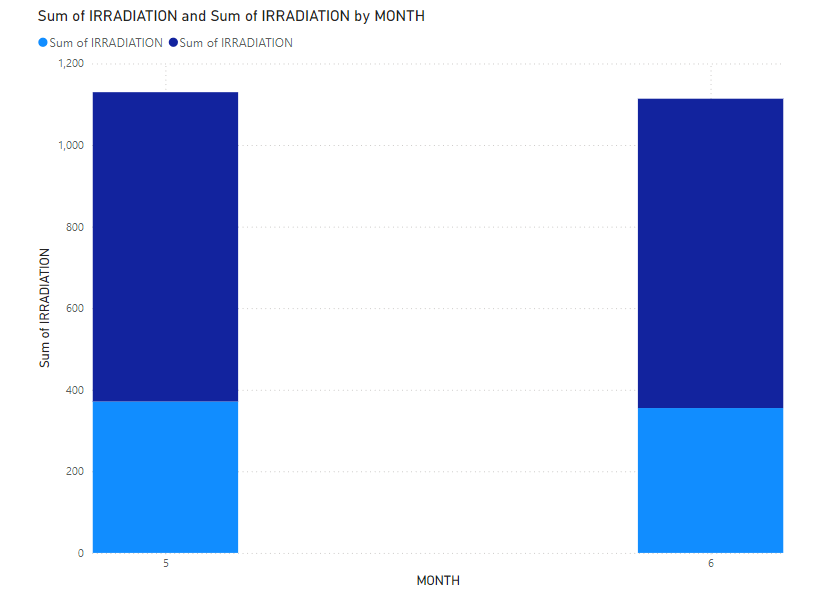
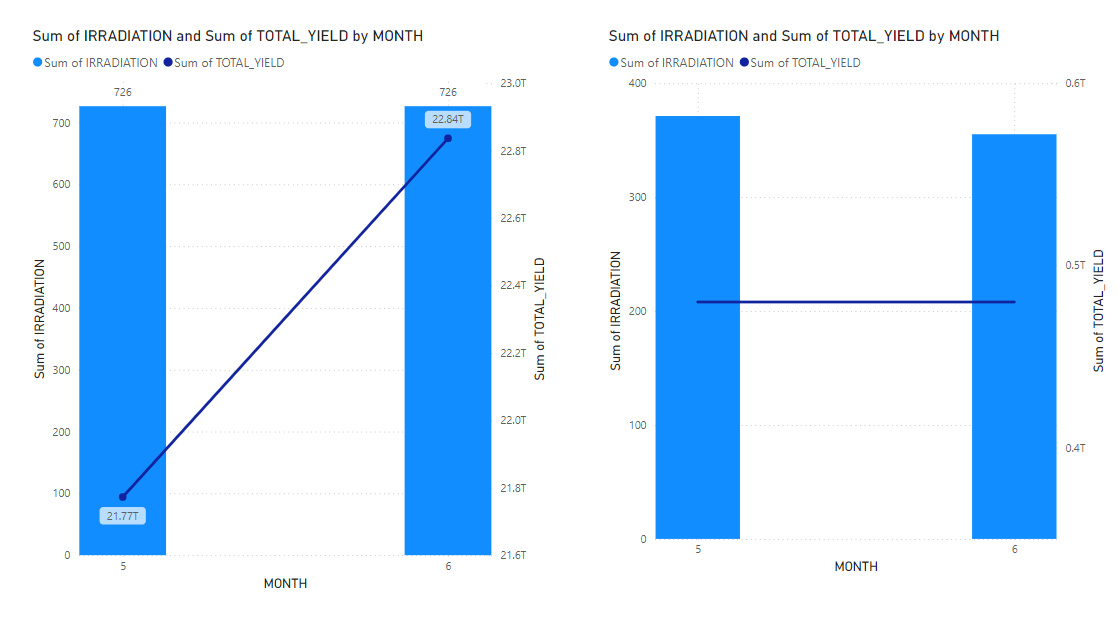
****

Figure 3.1.8 Irradiation levels of two plants by month

9. Calculate cumulative irradiation values over periods to assess overall energy potential.

1. Select the line and stacked column chart.
2. Provide the fields “IRRADIATION” and “MONTH” for visualization.

****Figure 3.1.9 Cumulative irradiation values by month

10. What is the total AC and DC Power Generated in both plants?

1. Select the stacked column chart.
2. Include the fields “AC\_POWER” and “DC\_POWER” to visualize.

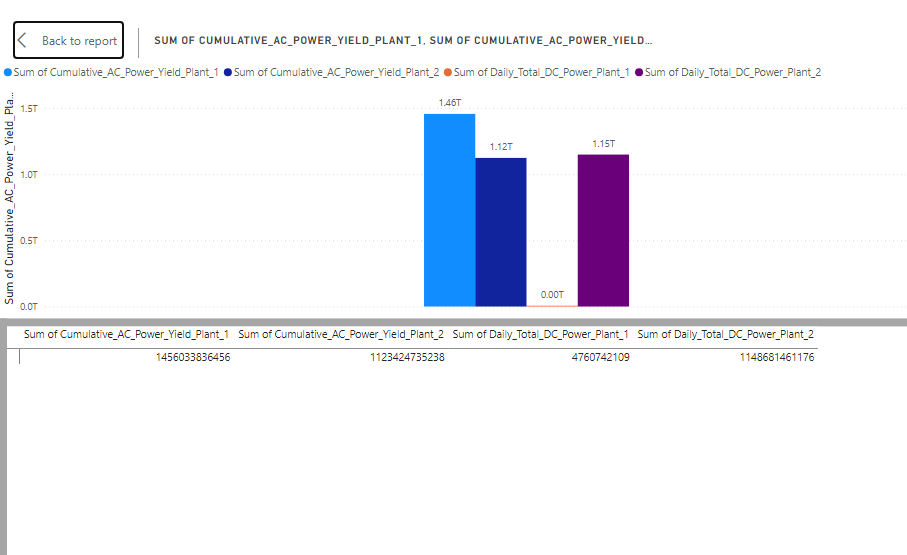


Figure 3.1.10. Total AC and DC power generated

11. Compare the total yield generated by two different plants.

1. Select the stacked area chart.
2. Include the field “TOTAL\_YEILD” and “MONTH” and visualize it.

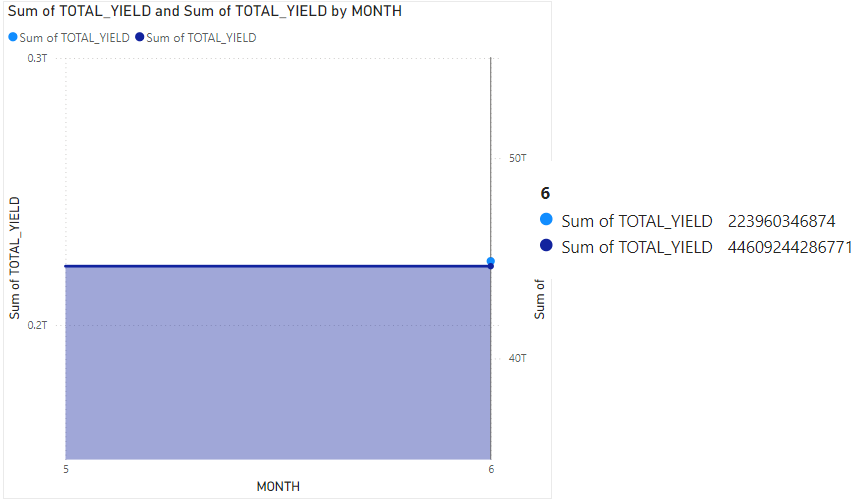
****

Figure 3.1.11. Total yield of two plants

12. Compare and contrast the thermal differences between the two solar plants.

1. Select the line chart.
2. Provide the fields “THERMAL DIFFERENCE” and “DATE” for visualization.

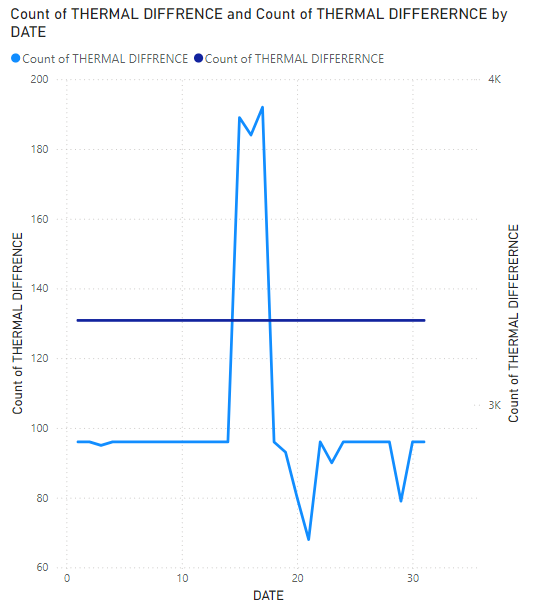
****

Figure 3.1.12 Comparison of Thermal Differences of two plants

13. Compare the AC\_POWER generated by two plants.

1. Select the area chart.
2. Include “AC\_POWER” of two plants and “DATE”.
3. Visualize the chart.

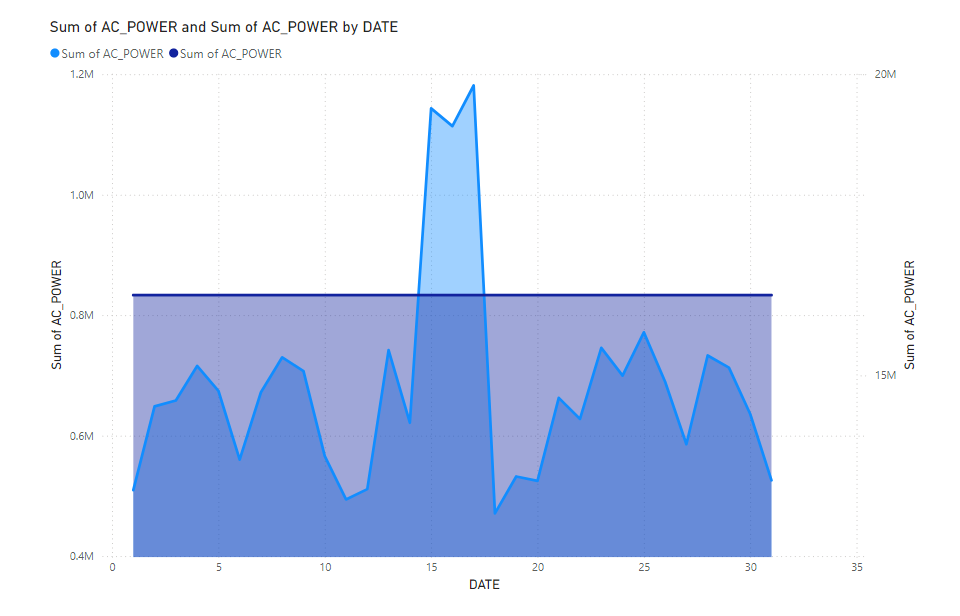
****

Figure 3.1.13 Comparison of Ac power of both the plants

14. Finding the adverse effect of the temperature difference to the power generation.

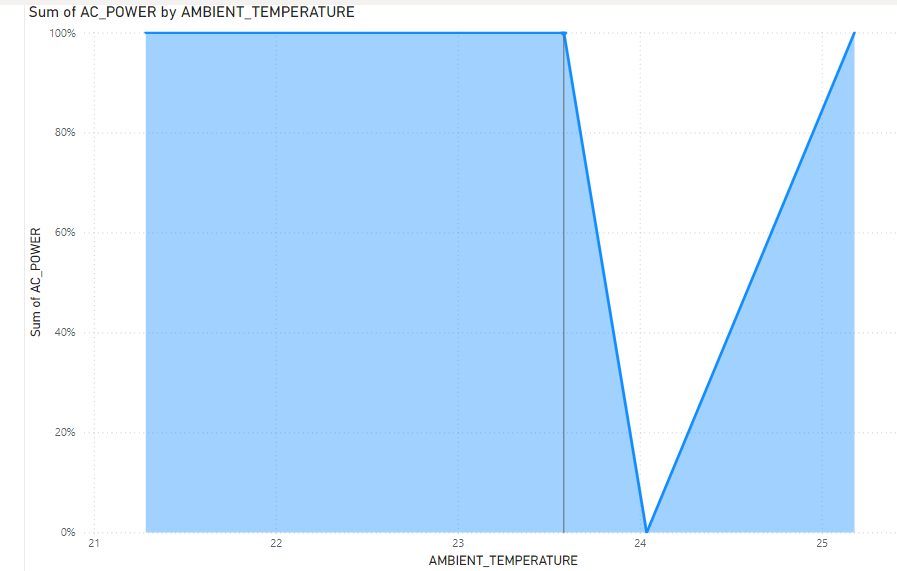
1. Select the stacked area chart.
2. Select the fields “AMBIENT\_TEMPERATURE” and “AC\_POWER” for visualization.****

Figure 3.1.14 Stacked area chart

15. How does the thermal difference affect the irradiation level?

1. Select the clustered column chart.
2. Provide the fields “IRRADIATION” and “THERMAL DIFFERENCE” to visualize.

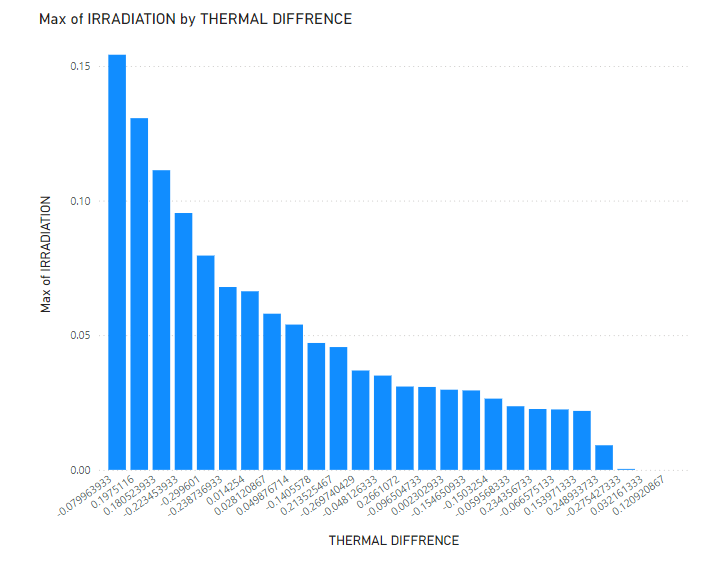
****

Figure 3.1.15 Irradiation by Thermal difference

**3.2. PUBLISHING DASHBOARD**

* A Power BI dashboard is a single page that tells a story using images. Since a well-designed dashboard is simply one page long, it only contains the essential elements of the story. Report creators place the images you see on the dashboard, called tiles.
* After selecting a tile, you frequently end yourself on the report page where the visualization was created. Reports provide the graphics for a dashboard, and each report is constructed from a single dataset. In actuality, a dashboard can be viewed as a gateway to the underlying statistics and reports.
* The report that was used to create the visualization may then be obtained by choosing Dashboards are a great way to monitor your business, look for answers, and see all of your most important metrics in one place.
* Visualizations on a dashboard may be derived from one or more underlying reports, as well as from a single underlying dataset.
* A dashboard may combine cloud and on-premises data to present a unified view, regardless of the data's storage location. The tiles on an interactive dashboard update in response to changes in the underlying data.
* **Dashboard Link:**

<https://app.powerbi.com/groups/me/reports/894b5a1c-2da0-48b2-b0bd-9db0aa3acd14/b87b955a4d281d1da807?experience=power-bi>

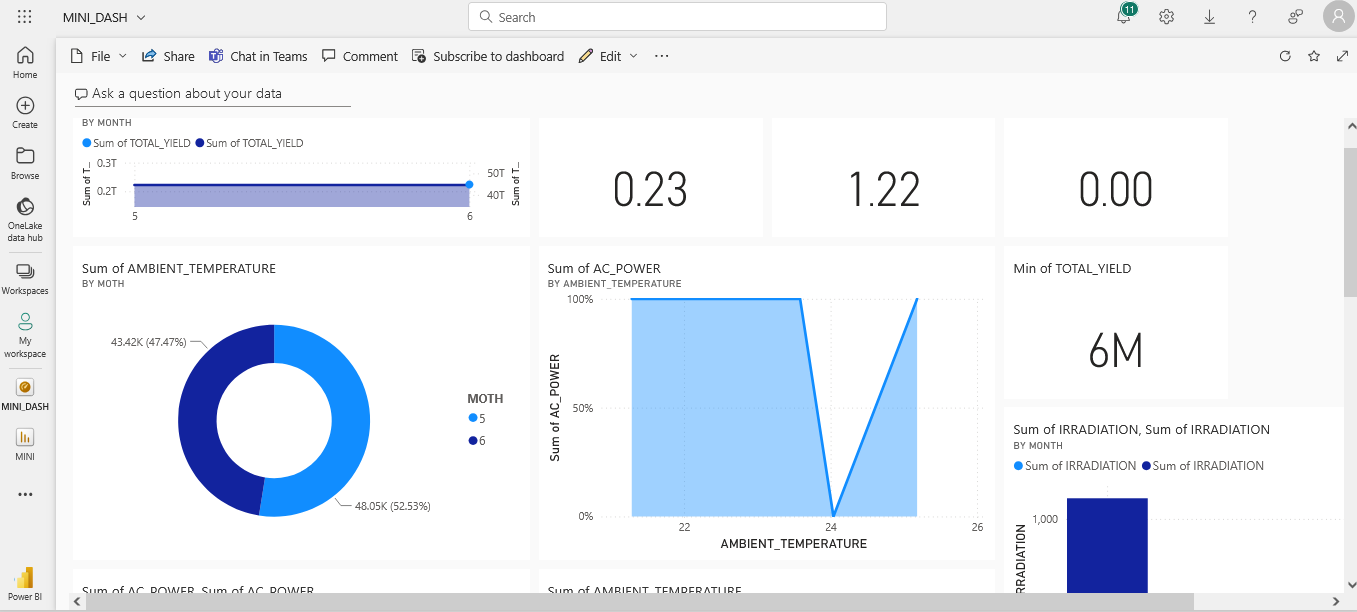
****

Figure 3.2.1 Published Dashboard

**3.3. INFERENCES**

1. Split the data and time separately.

INFERENCE: The date and time has been split successfully.

1. Find the thermal differences.

INFERENCE: Thermal differences have been found.

Thermal difference=(Ambient temp - Module temp)

1. Analyze how ambient temperature affects module temperature.

INFERENCE: Module temperature increases with respect to ambient temperature

1. Find peak irradiation times.

INFERENCE: The peak irradiation is at 12.00 pm.

1. Calculate average, max, and min temperatures.

INFERENCE: Average: 6.98 M, Maximum: 8 M, Minimum: 6 M

1. Calculate average, max, and min irradiation values.

INFERENCE: Average: 0.23, Maximum: 1.22, Minimum: 0

1. Compare different plants (using PLANT\_ID) to see which has higher ambient temperatures.

INFERENCE: Plant ID 4136001 has acquired highest ambient temperature

1. Compare different plants (using PLANT\_ID) to see which has higher irradiation levels.

INFERENCE: The PLANT ID 4135001 has higher irradiation level

1. Calculate cumulative irradiation values over periods to assess overall energy potential.

INFERENCE:

* + 1. PLANT ID 4135001: The total yield increases over time
    2. PLANT ID 4136001: The total yield remains same over the

1. Find the relationship between the two dataset.

INFERENCE: The AC power attribute has many to one relationship to the thermal difference attribute.

1. Compare the total yield generated by two different plants.

INFERENCE: The total yield generated by PLANT 4135001 is much lower than that has been generated by the PLANT 4136001

1. Compare and contrast the thermal differences between the two solar plants.

INFERENCE: The thermal difference of PLANT ID 4135001 varies while of the other plant remains same.

1. Compare the AC\_POWER generated by two plants.

INFERENCE: The PLANT ID 4135001 generates varying AC power while the other generates 16334048 sum of AC power all the days.

1. Finding the adverse effect of the temperature difference to the power generation.

INFERENCE: The power generated remains same until the threshold value of ambient temperature with value 23.4 and varies as the temperature increases.

1. How does the thermal difference affect the irradiation level?

INFERENCE: The irradiation level decreases with increase in thermal difference.

**CHAPTER 4**

**CONCLUSION**

**4.1. RECOMMENDATIONS**

Based on the analysis, it is recommended to monitor ambient and module temperatures more closely and to implement cooling strategies where necessary to prevent overheating, especially in high-temperature regions. Additionally, adjusting the positioning of panels based on irradiation data could help increase energy efficiency.

**CHAPTER 5**

**REFERENCE**

1. <https://data.world/gymprathap/india-solar-power-generation-dataset/workspace/file?filename=India-Solar-Power-Generation-Dataset.zip>