EDA / Descriptive Statistics

## Introduction:

Solar energy has become a crucial component of sustainable power solutions worldwide, with solar farms playing a key role in harnessing this renewable resource. The client's solar farm, however, has been experiencing a power drop per string, which has yet to be thoroughly investigated. To maximize energy efficiency, it is essential to identify and address the underlying causes of this underperformance. Our project aims to enhance the output of the solar farm while minimizing operational costs. By ensuring that 95% maximum output is achieved across all strings, even with weather variations, and by increasing annual revenue by $1.75 million, we can significantly improve both the efficiency and profitability of the client's solar operations.

## Overall design strategy

### Overall Design Strategy

The solar dataset comprises 366 records and includes 23 essential columns, capturing various aspects of solar power generation and performance metrics. Our design strategy focuses on extracting meaningful insights from this dataset to identify and address the causes of power drops per string and improve overall energy efficiency.

### Data Preparation

The dataset includes critical columns such as:

* **Date**: Recording date for each entry.
* **Inverter Measurements**: 'ICR-3 - INV1', 'ICR-3 - INV2', 'ICR-3 - INV3', 'ICR-3 - INV4', 'ICR-4 - INV1', 'ICR-4 - INV2', 'ICR-4 - INV3', 'ICR-4 - INV4'.
* **Generation Metrics**: 'Total Daily Integrated Generation MWh', 'PSS Main Meter (Export/Import) MWh', 'PSS Check Meter (Export/Import) MWh', 'Daily Generation Plant end meter (net) MWh'.
* **Irradiance Measurements**: 'Global Tilted Irradiation/Irradiance (GTI) kWh/m2', 'Global horizontal irradiance (GHI) kWh/m2'.
* **Performance and Downtime**: 'Performance Ratio (PR) %', 'Grid Downtime HH’,'Plant Downtime HH’
* **Remarks**: Any additional notes.

## Data Overview

The solar dataset comprises 366 records, extracted to gain insights into the performance and efficiency of the client's solar farm. This dataset includes crucial measurements and metrics necessary for comprehensive analysis. Below is an overview of the key data components and their sources:

### Data Extraction

The data was collected from the client's solar farm monitoring systems, encompassing various aspects of power generation and performance metrics. This data includes:

* **Date**: The recording date for each entry.
* **Inverter Measurements**: Data from eight inverters including 'ICR-3 - INV1', 'ICR-3 - INV2', 'ICR-3 - INV3', 'ICR-3 - INV4', 'ICR-4 - INV1', 'ICR-4 - INV2', 'ICR-4 - INV3', 'ICR-4 - INV4'.
* **Generation Metrics**: Total Daily Integrated Generation (MWh), PSS Main Meter and Check Meter readings for Export and Import (MWh), and net daily generation metrics.
* **Irradiance Measurements**: Global Tilted Irradiation/Irradiance (GTI) in kWh/m² and Global Horizontal Irradiance (GHI) in kWh/m².
* **Performance and Downtime**: Performance Ratio (PR) percentage, Grid Downtime, and Plant Downtime in HH

format.

* **Remarks**: Any additional notes or observations.

### Data Integration and Tools

The extracted data was initially stored in text files and subsequently uploaded into our data visualization and analysis tools, including Python, Power BI, Excel, and Google Looker Studio. These files were joined based on relevant keys to ensure a cohesive dataset for analysis. Following the initial data load, an additional extract summarizing team-level data was created using custom SQL queries. This step was essential to aggregate the data for more detailed and insightful visualizations.

* **Python**: Used for initial data processing, cleaning, and creating custom SQL outputs.
* **Power BI**: Employed to create dynamic and interactive visualizations, offering real-time insights into the data.
* **Excel**: Utilized for data validation, preliminary analysis, and creating supplementary data tables.
* **Google Looker Studio**: Used to build comprehensive dashboards and share interactive reports with stakeholders.

### Summary

The comprehensive dataset and its integration into these powerful tools enable a thorough analysis of the solar farm's performance. By examining inverter performance, generation metrics, irradiance impact, and downtime data, we aim to identify the causes of power drops per string and develop strategies to maximize energy efficiency and operational profitability. This multi-tool approach ensures robust data handling, insightful visualization, and effective communication of findings to drive decision-making and improvements.

## Users

* **Plant Managers**: Use the dashboard to see overall solar plant performance and plan improvements.
* **Technical Staff**: Monitor inverter performance, track energy trends, and fix technical problems quickly.
* **Investors**: Get quick insights into plant performance and profitability.
* **Regulatory Bodies**: Ensure the plant meets energy standards and operates efficiently.

### Questions Answered by the Dashboard

**Plant Managers:**

* **Performance Trend**: How is the plant's energy generation changing over time?
* **Best and Worst Months**: Which months produce the most and least energy?
* **Weather Impact**: How does weather affect energy production?
* **Inverter Performance**: Which inverters are working well and which aren't?

**Technical Staff:**

* **Inverter Trends**: How does each inverter's performance change over time?
* **Daily Efficiency**: How efficient is the plant each day?
* **Anomalies**: Are there any unusual drops in daily energy production?
* **Weather Effects**: How do different weather conditions affect daily and cumulative energy generation?

**Investors:**

* **Average Output**: What is the plant's average energy output over time?
* **Reliability and Profitability**: Are the energy generation trends reliable and profitable?
* **Comparisons**: How does the plant's performance match initial projections and industry standards?

**Regulatory Bodies:**

* **Regulatory Targets**: Is the plant meeting energy generation targets?
* **Consistency**: Is the plant's performance consistent throughout the year?
* **Downtimes**: Are there any significant downtimes or inefficiencies?

### Visualizations and Answers

**Plant Managers:**

* **Performance Trend**: Line chart of daily and monthly energy output.
* **Best and Worst Months**: Bar charts showing average monthly energy output.
* **Weather Impact**: Scatter plots comparing energy generation on sunny, cloudy, and windy days.
* **Inverter Performance**: Charts showing individual inverter performance.

**Technical Staff:**

* **Inverter Trends**: Line charts of each inverter's daily and monthly output.
* **Daily Efficiency**: Scatter plots of daily energy generation vs. solar irradiance.
* **Anomalies**: Detailed trend lines for daily energy generation.
* **Weather Effects**: Line charts of energy generation under various weather conditions.

**Investors:**

* **Average Output**: Summary statistics and line charts of average output.
* **Reliability and Profitability**: Yearly trend lines and performance summary charts.
* **Comparisons**: Benchmark comparison charts.

**Regulatory Bodies:**

* **Regulatory Targets**: Compliance charts and performance dashboards.
* **Consistency**: Yearly and monthly performance trend lines.
* **Downtimes**: Anomaly detection charts and downtime logs.

## Conclusion

Using the dashboard helps various stakeholders quickly understand the solar plant's performance, spot issues, and plan for improvements. Visualizations make complex data easy to grasp, aiding in decision-making and optimizing the plant's operation. With ongoing tech advancements, the dashboard's insights can become even more precise and useful.