Strategy Report: Data Analysis for Solar Installation Strategy

Introduction

MoonLight Energy Solutions is committed to enhancing operational efficiency and sustainability through targeted solar investments. As part of the company's ongoing efforts, a key task is to analyze environmental data to identify potential regions for solar installation. This report presents a statistical analysis and exploratory data analysis (EDA) of environmental measurements provided by the engineering team, with a specific focus on identifying regions with the highest potential for solar installation.

Data Overview

The data provided consists of environmental measurements recorded over a set period. Key variables include solar irradiance (Global Horizontal Irradiance - GHI, Direct Normal Irradiance - DNI, and Diffuse Horizontal Irradiance - DHI), temperature (Tamb), wind speed (WS), and other environmental factors such as precipitation, atmospheric pressure (BP), and cleaning status.

Data Cleaning and Preprocessing

During the initial data inspection, a few observations were made:

- Negative Values in Irradiance Columns: The columns GHI, DNI, and DHI contain negative values in several records ( -1.2 for GHI, -0.2 for DNI, -1.1 for DHI). In the context of solar irradiance data, negative values are unexpected since irradiance should always be non-negative. These negative values may indicate:

- Data Errors: Incorrect measurements or malfunctions in the data collection process.

- Missing or Invalid Data: Negative values might be used as placeholders to signify missing or invalid measurements.

- Action Taken: The negative values in the irradiance columns need to be treated. It is recommended to replace these values with `NaN` (Not a Number) for missing data or to apply a data imputation method, such as replacing negative values with the mean or median of the valid values. Further investigation into the data collection process is also recommended to prevent these errors in future datasets.

Statistical Analysis and Exploratory Data Analysis (EDA)

1. Summary Statistics:

- The mean values for GHI, DNI, and DHI suggest average irradiance levels over the measured period. However, the standard deviation is quite large, indicating high variability in these values, which is typical of environmental measurements.

- ModA and ModB values represent solar panel model efficiency, with average values of 236.59 and 228.88, respectively, suggesting strong performance during the measurement period.

- Other environmental variables, such as Tamb (temperature), RH (relative humidity), and WS (wind speed), show consistent patterns but may require further analysis to understand their relationship with solar energy generation.

2. Correlations

- The analysis of correlation between variables, particularly irradiance (GHI, DNI, DHI) and temperature (Tamb), will help determine how environmental conditions impact solar panel performance.

- Wind speed (WS) and wind gusts (WSgust) should also be analyzed to assess whether high winds negatively affect solar panel efficiency or contribute to increased dust and dirt accumulation, impacting energy production.

3. Outliers

- Outliers in wind speed (WS) and gusts (WSgust) could indicate extreme weather conditions that affect solar energy production.

- It's important to investigate these outliers, as they may provide insights into environmental factors that significantly alter solar energy output.