**📄 4.3 Exploratory Data Analysis (EDA)**

**4.3.1 Objective of EDA**

The goal of Exploratory Data Analysis (EDA) is to thoroughly understand the structure, composition, and behavior of the collected datasets before feeding them into any machine learning or forecasting model. This process enables :

* Understanding the distribution and spread of individual features,
* Identifying trends, patterns, and early relationships between variables,
* Detecting missing values, inconsistencies, and possible outliers,
* Gaining insights to guide feature selection, transformation, and model strategy.

**4.3.2 Methodology Adopted**

The EDA process was conducted using **Jupyter Notebook**, within a **Conda-managed Python environment**. The following analysis was carried out on each dataset representing four different towns — **Bambili, Bamenda, Bafoussam, and Yaoundé** — using the CSV files provided.

The initial focus was on **Univariate Analysis**, i.e., analyzing each variable individually across all towns.

**4.3.3 Data Overview and Structure**

Each town’s dataset contains 27,214 entries and consists of the following columns:

* **Date**: Integer representing YYYYMMDD format.
* **Temperature** (°C)
* **Humidity** (%)
* **Irradiance** (W/m²)
* **Potential** (unitless value approximating solar energy potential)
* **Wind Speed** (m/s)

Each feature is of type float64 (except for Date, int64), with **no missing values detected**.

**4.3.4 Univariate Statistical Summary**

**a. Temperature**

* Ranges across towns:
  + Bambili: Mean ≈ 21.8°C, Max = 24.4°C
  + Bamenda: Mean ≈ 18.2°C, Max = 23.7°C
  + Bafoussam: Mean ≈ 20.6°C, Max = 26.5°C
  + Yaoundé : Mean ≈ 22.7°C, Max = 27.5°C
* Insight: **Yaoundé** is the warmest, **Bamenda** the coolest on average.

**b. Humidity**

* Mean values range from **91.9% (Bafoussam)** to **95.6% (Yaoundé)**.
* Insight: All towns have **very high humidity**, possibly influencing solar irradiance and energy production.

**c. Irradiance**

* Average irradiance :
  + Bambili ≈ 643.8 W/m²
  + Bamenda ≈ 631.1 W/m²
  + Bafoussam ≈ 657.3 W/m²
  + Yaoundé ≈ 558.9 W/m²
* Insight: **Bafoussam** shows the highest solar irradiance; **Yaoundé** the lowest despite being the warmest.

**d. Potential**

* Average potential :
  + Highest in **Bafoussam** (≈ 3.37), followed by Bamenda (3.23), Bambili (3.19), and Yaoundé (2.86).
* Insight: **Higher irradiance and moderate humidity in Bafoussam may contribute to its solar potential**.

**e. Wind Speed**

* Significantly varies across towns:
  + Bambili ≈ 0.52 m/s
  + Bamenda ≈ 0.51 m/s
  + Bafoussam ≈ 0.63 m/s
  + Yaoundé ≈ 1.43 m/s
* Insight: **Yaoundé has the highest wind activity**, potentially suitable for hybrid (solar + wind) energy strategies.

**4.3.5 Key Observations & Implications**

* The datasets are clean, consistent, and uniformly structured, with no nulls or corrupt entries.
* **Temperature and irradiance patterns** vary between towns, indicating **location-specific models may be needed**.
* High humidity across all towns could affect irradiance and thus model performance.
* **Yaoundé's high wind speed** could justify incorporating wind-based forecasting alongside solar models.
* **Bafoussam consistently scores highest on solar-related metrics**, making it an interesting benchmark town for evaluating models.

**4.3.6 Next Steps in EDA**

Further analysis will include:

* **Visualizations** (histograms, boxplots, line charts),
* **Correlation analysis** to understand feature interdependencies,
* **Multivariate analysis** (joint plots, pair plots),
* **Time-series decomposition** for seasonality and trend detection.