TUNIS BUSINESS SCHOOL
UNIVERSITY OF TUNIS

Report

IT 300

Business Intelligence and Database Management Systems

Business Intelligence Mini-Project
SunWise Analytics

Authors:
Hiba Chaabouni
Ghada Ben Amira
Alaa Dbira
Yomna Brahmi
Ali Ayari

Submitted to:
Professor Amani Azzouz

Contents

1. Introduction

Solar energy is attracting growing interest in Tunisia and the MENA region due to its favorable natural conditions, drawing attention from governments and investors alike. However, solar performance remains highly dependent on weather variability, making reliable analysis essential for informed investment and planning decisions.

This project analyzes solar and weather data to understand how environmental factors affect energy production. Based on these observations, the analysis is applied to the MENA region to support investors and decision-makers in selecting suitable locations, improving forecasts, and planning efficient and sustainable solar projects.

Due to limited MENA solar data, Tunisia's real measurements were used as a baseline, with regional indices for solar intensity, grid availability, and population proximity to estimate conditions across the region.

2. Implementation

2.1. Data Gathering

Type: CSV
Source: Kaggle
[Renewable Energy and Weather Conditions]

2.2 Data Preparation

///Using Python

Loaded Data

- CSV file loaded using Python's pandas library.
- Initial inspection performed using .info() and .describe().

Standardized Column Structure

- Converted column names to lowercase.

- Removed spaces and special characters ([ ] /).
- Ensured consistent and readable column naming.

Cleaned Data

- Converted the time column to datetime format.
- Extracted year and day from the time column.
- Verified that the dataset contains no missing values.
- Removed columns containing only zero values.
- Removed outliers from the energy_delta_wh column using the IQR method.
- Categorized cloud coverage into Low, Medium, and High using bins.
- Normalized numerical features(ghi, temp, wind_speed)using min-max scaling.
- Ensured data consistency and readiness for analysis.

Feature Engineering

- Created time-based features for trend and seasonal analysis.
- Generated categorical weather indicators.
- Prepared normalized numerical features for comparability.

2.3 Data Modeling

2.3.1. Storage

Database Tool: MySQL

MySQL Code:

```sql
USE Solar_weather;

-- 1. Create Date Dimension Table
CREATE TABLE dim_date (
    time_id INT AUTO_INCREMENT PRIMARY KEY,
    date DATE,
    hour INT,
    day INT,
    month INT,
    quarter INT,
    year INT,
    season VARCHAR(20)
);

-- 2. Create Weather Dimension Table
```

```sql
CREATE TABLE dim_weather (
    weather_id INT AUTO_INCREMENT PRIMARY KEY,
    ghi FLOAT,
    temperature FLOAT,
    humidity INT,
    wind_speed FLOAT,
    isSun TINYINT
);

-- 3. Create Location Dimension Table
CREATE TABLE dim_location (
    location_id INT AUTO_INCREMENT PRIMARY KEY,
    country VARCHAR(100),
    region VARCHAR(100),
    latitude FLOAT,
    longitude FLOAT,
    mena_region VARCHAR(100)
);

-- 4. Create Solar Energy Fact Table
CREATE TABLE fact_solar_energy (
    fact_id INT AUTO_INCREMENT PRIMARY KEY,
    time_id INT,
    weather_id INT,
    location_id INT,
    energy_delta_wh FLOAT,
    FOREIGN KEY (time_id) REFERENCES dim_date(time_id),
    FOREIGN KEY (weather_id) REFERENCES dim_weather(weather_id),
    FOREIGN KEY (location_id) REFERENCES dim_location(location_id)
);
```

2.3.2. Fact: `fact_solar_energy`

The **Solar Energy Fact Table** stores the core quantitative measure of the data warehouse, representing changes in solar energy production under specific temporal, weather, and geographical conditions.

Attributes:

- `fact_id`: **PK**
- `time_id`: **FK** → `dim_date`
- `weather_id`: **FK** → `dim_weather`
- `location_id`: **FK** → `dim_location`
- `energy_delta_wh`: Measure (energy variation in watt-hours)

2.3.3. Dimensions

2.3.3.1. Date Dimension: `dim_date`

Stores temporal information to support time-based analysis.

- `time_id`: **PK**
- `date`
- `hour`
- `day`
- `month`
- `quarter`
- `year`
- `season`

2.3.3.2. Weather Dimension: `dim_weather`

Contains meteorological factors influencing solar energy production.

- `weather_id`: **PK**
- `ghi` (Global Horizontal Irradiance)
- `temperature`
- `humidity`
- `wind_speed`
- `isSun` (binary indicator for sunny conditions)

2.3.3.3. Location Dimension: `dim_location`

Represents geographical and regional attributes, with a focus on MENA applicability.

- `location_id`: **PK**
- `country`

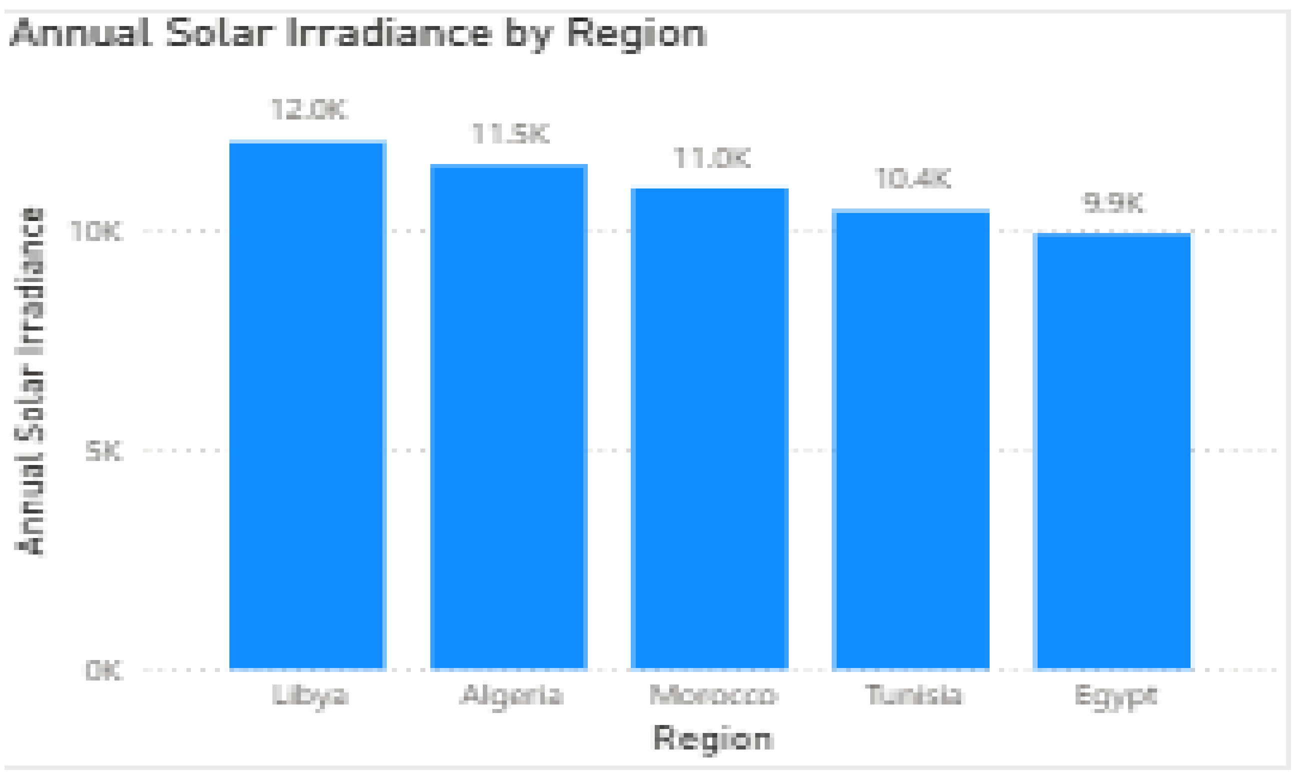
- `region`
- `latitude`
- `longitude`
- `mena_region`

2.3.4. Validation & Consistency Checks

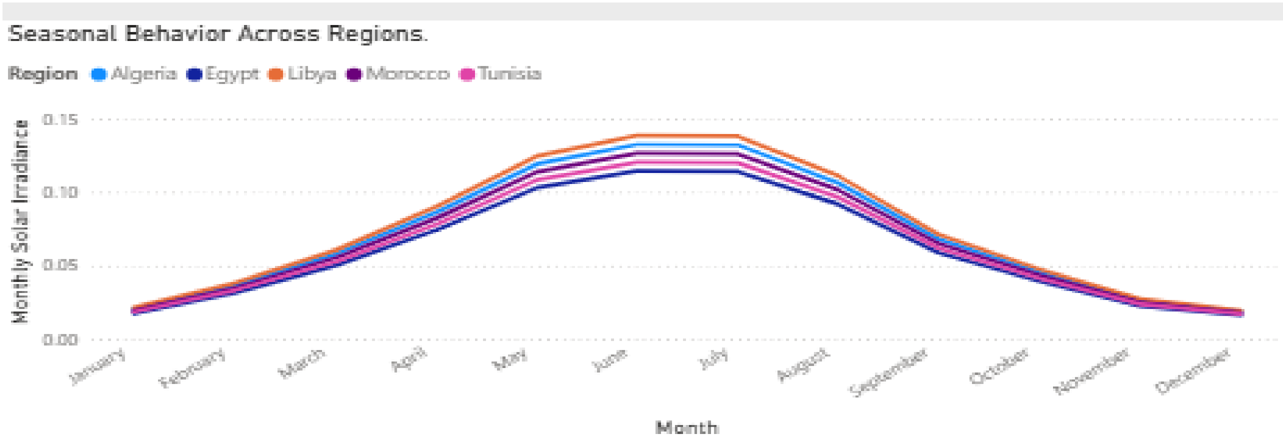
```
-- Check fact table row count
SELECT COUNT(*) AS total_fact_rows
FROM fact_solar_energy;>

-- Check missing foreign keys
SELECT *
FROM fact_solar_energy
WHERE time_id IS NULL
   OR weather_id IS NULL
   OR location_id IS NULL;

-- Verify joins work correctly
SELECT
    f.energy_delta_wh,
    d.date,
    l.region,
    w.ghi
FROM fact_solar_energy f
JOIN dim_date d ON f.time_id = d.time_id
JOIN dim_location l ON f.location_id = l.location_id
JOIN dim_weather w ON f.weather_id = w.weather_id
LIMIT 10;
```

3. Data Visualization:

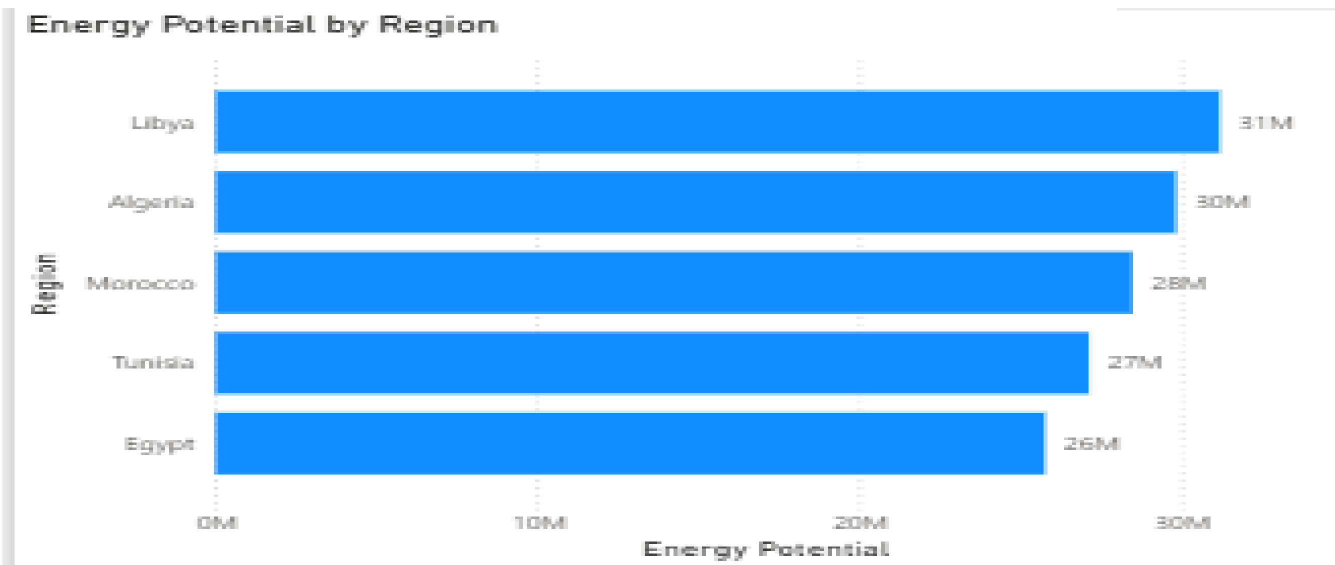
- **Which regions in the MENA region have the highest annual solar irradiance?**



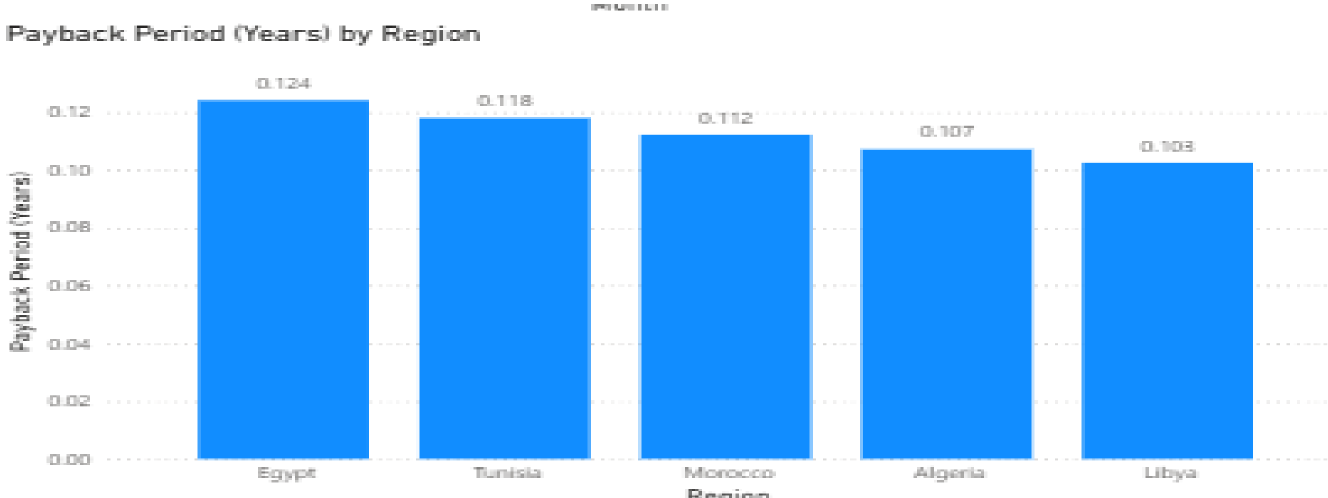
- **How does solar irradiance vary seasonally in Tunisia and neighboring countries?**



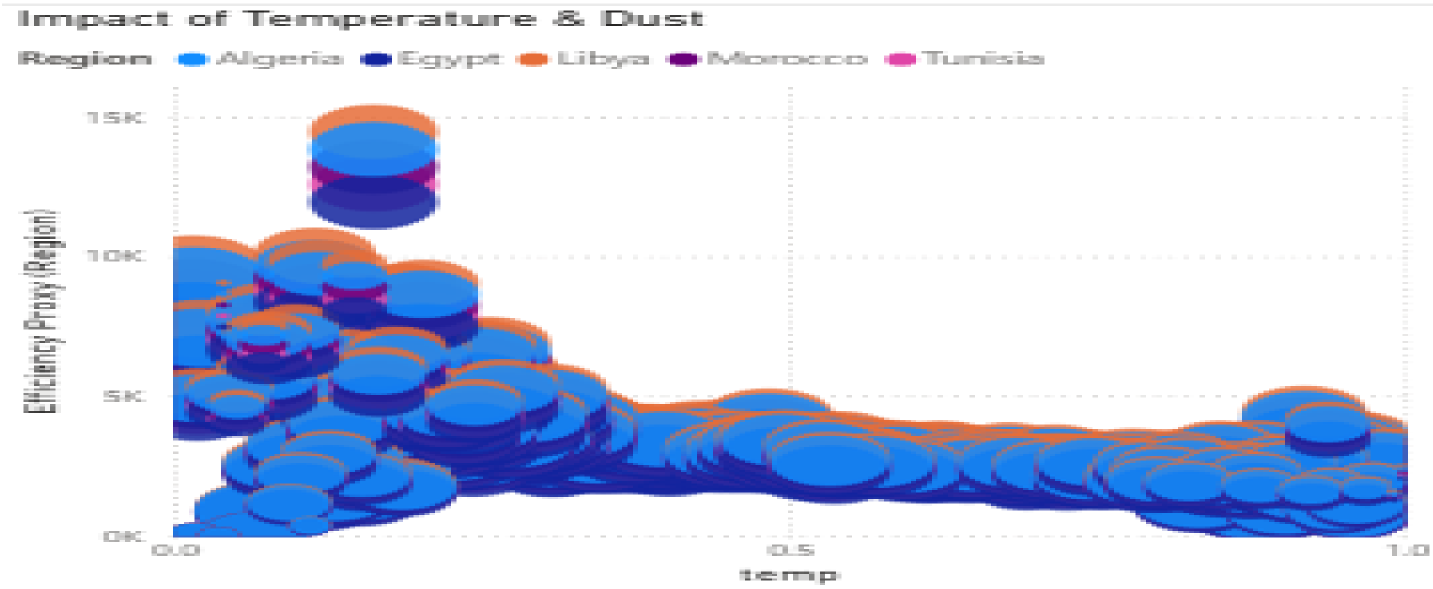
- **Which potential sites have the highest potential energy output per square meter?**



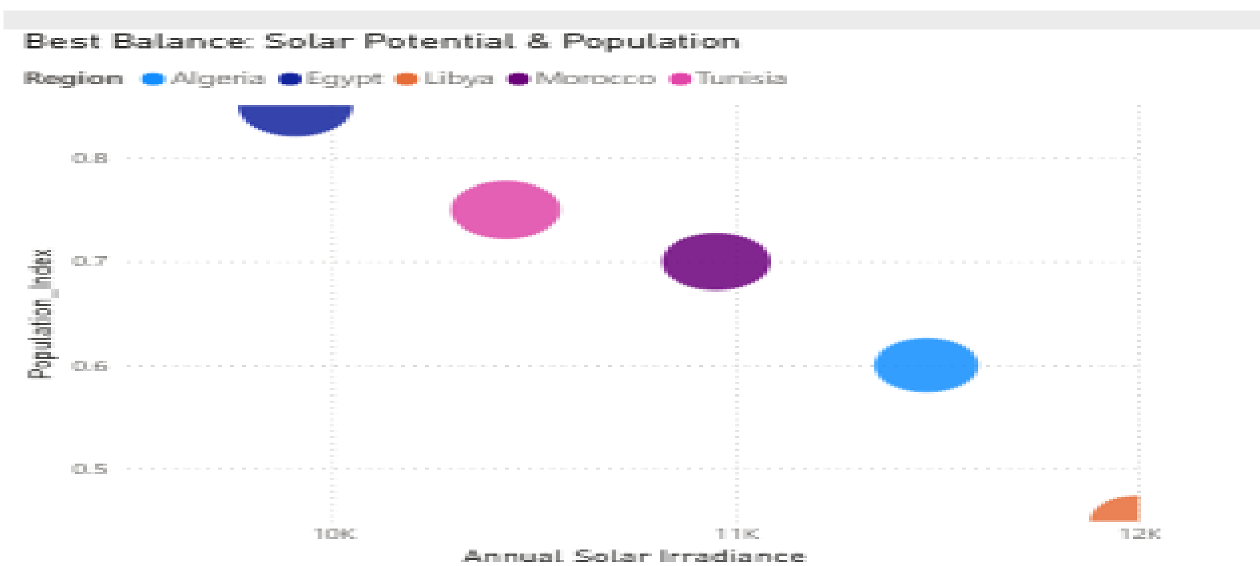
- **What is the estimated payback period for solar installations in different regions?**



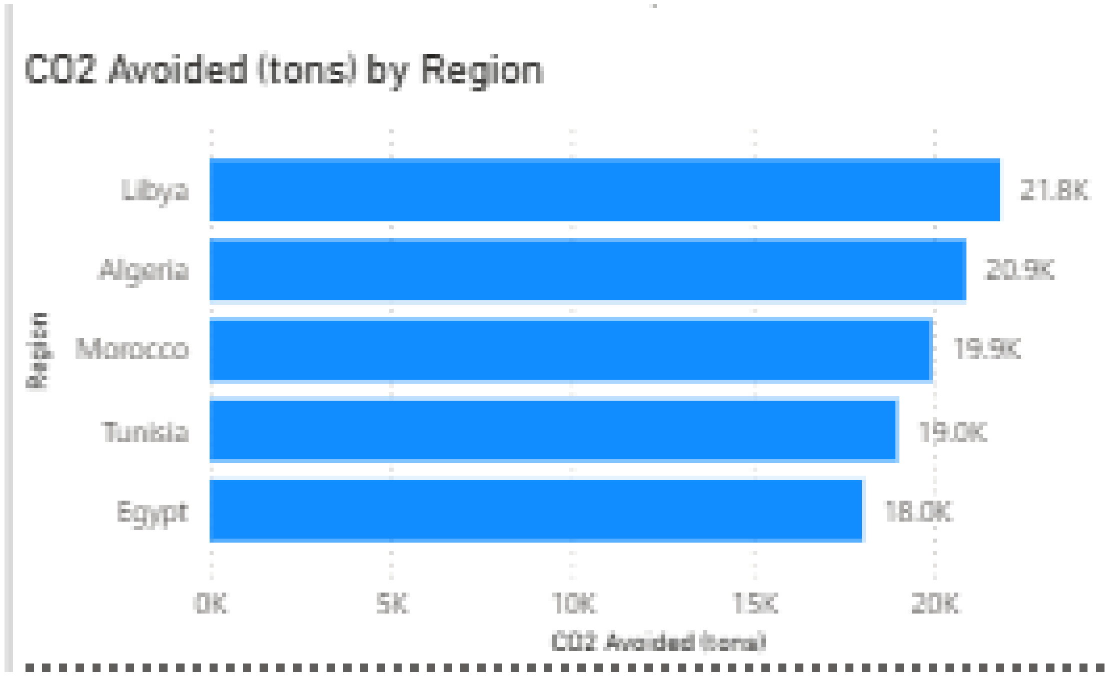
- **How do temperature and dust storms in North Africa affect solar panel efficiency?**



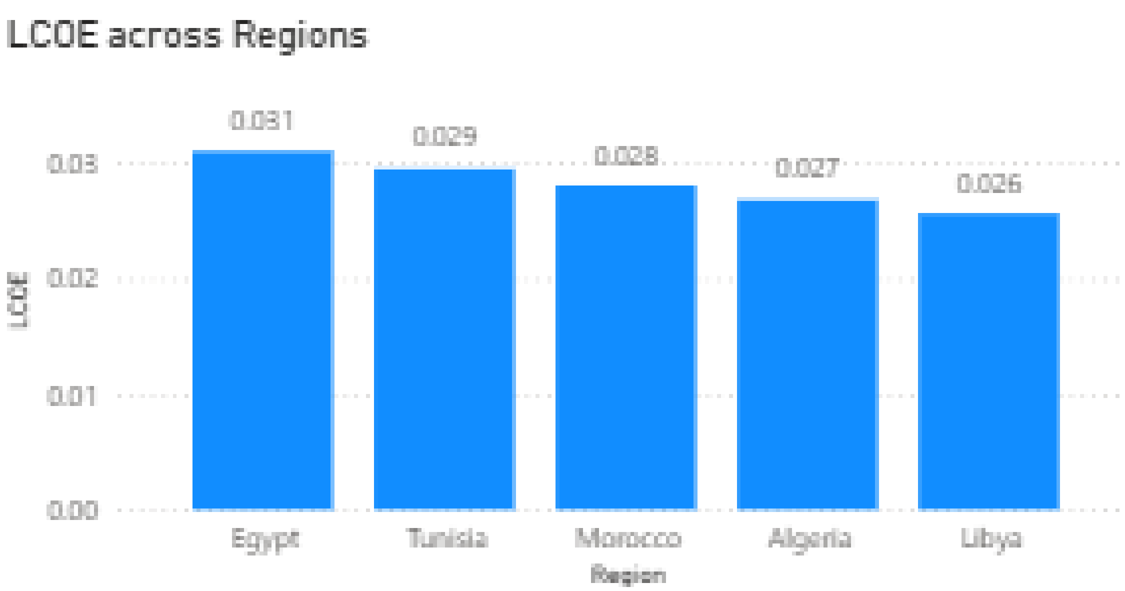
- **Which sites provide the best balance between proximity to population centers and solar potential?**



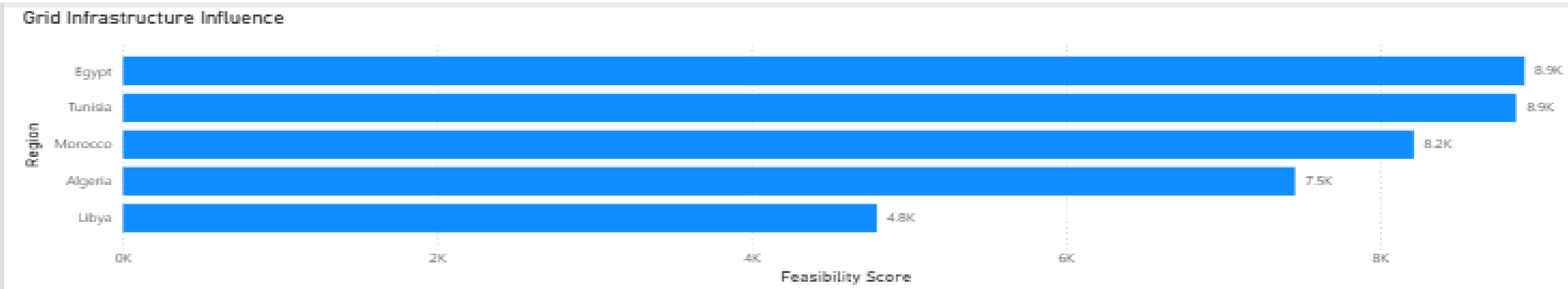
- **How much CO_2 emissions could be avoided if a solar project is installed in each region?**



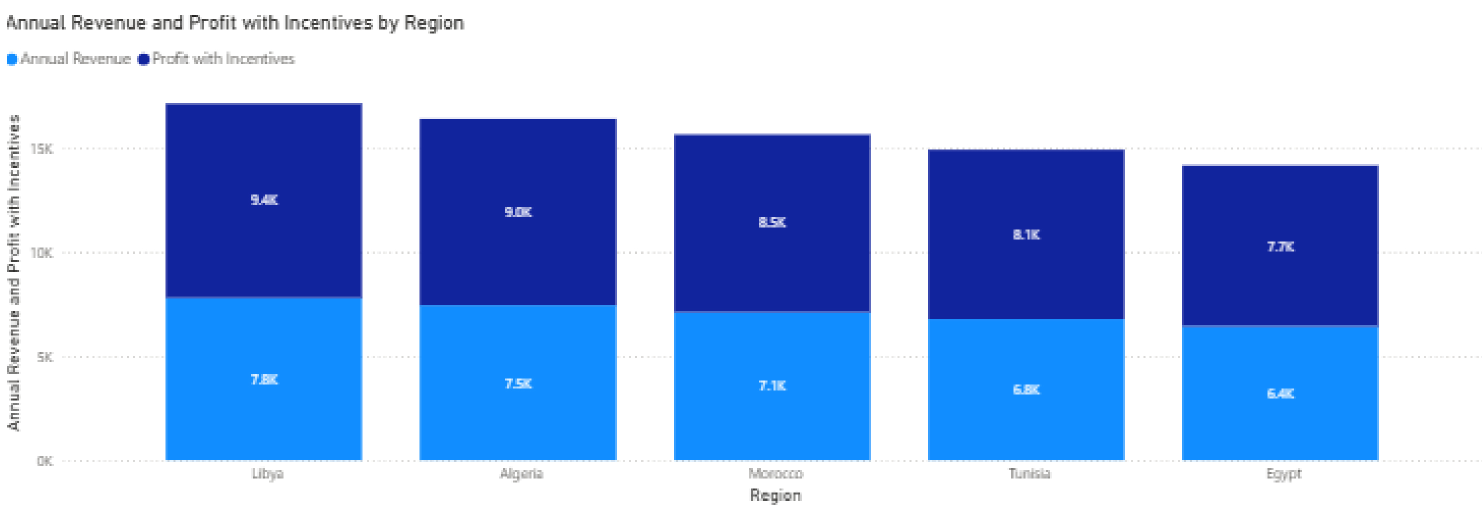
- **What is the levelized cost of electricity (LCOE) for potential solar installations across MENA countries?**



- **How does the grid infrastructure availability influence site feasibility?**



- **What is the profitability under different incentive schemes (e.g., feed-in tariffs, tax incentives) in Tunisia and MENA?**



- **What is the risk of project downtime due to local weather patterns (sandstorms, high temperatures)?**

Month	Algeria	Egypt	Libya	Morocco	Tunisia
January	80.71	80.71	80.71	80.71	80.71
February	74.06	74.06	74.06	74.06	74.06
March	65.56	65.56	65.56	65.56	65.56
April	56.09	56.09	56.09	56.09	56.09
May	61.28	61.28	61.28	61.28	61.28
June	59.45	59.45	59.45	59.45	59.45
July	62.85	62.85	62.85	62.85	62.85
August	63.19	63.19	63.19	63.19	63.19
September	67.53	67.53	67.53	67.53	67.53
October	71.12	71.12	71.12	71.12	71.12
November	81.79	81.79	81.79	81.79	81.79
December	84.31	84.31	84.31	84.31	84.31

4. Annex:

- Created a MENA Region table:

```
MENA_Region =
DATATABLE(
    "Region", STRING,
    "Country", STRING,
    "Solar_Index", DOUBLE,
    "Grid_Index", DOUBLE,
    "Population_Index", DOUBLE,
    {
        {"Tunisia", "Tunisia", 1.00, 0.85, 0.75},
        {"Morocco", "Morocco", 1.05, 0.75, 0.70},
        {"Algeria", "Algeria", 1.10, 0.65, 0.60},
        {"Egypt", "Egypt", 0.95, 0.90, 0.85},
        {"Libya", "Libya", 1.15, 0.40, 0.45}
    }
)
```

- Created a new measure called **Adjusted GHI => a key trick that will unlock everything:**

```
Adjusted GHI = SUM('solar_weather_clean 3'[ghi]) *
AVERAGE('MENA_Region'[Solar_Index])
```

- Created a new measure named **Annual Solar Irradiance:**

```
Annual Solar Irradiance = SUM('solar_weather_clean 3'[ghi])
* AVERAGE('MENA_Region'[Solar_Index])
```

- Another measure for the seasonal variation named **Seasonal GHI**:

```
Seasonal GHI = AVERAGE('solar_weather_clean 3'[ghi]) *
AVERAGE('MENA_Region'[Solar_Index])
```

- A new measure for the highest potential energy output per m²: Energy Potential:

```
Energy Potential = SUM('solar_weather_clean
3'[energy_deltawh]) * AVERAGE('MENA_Region'[Solar_Index])
```

- An **Annual Revenue** measure with the following assumptions:
  cost= 800TND/m^2
  Electricity price= .25TND/kWh (Lower bar = better investment):

```
Annual Revenue = (SUM('solar_weather_clean
3'[energy_deltawh]) / 1000) * 0.25 *
AVERAGE('MENA_Region'[Solar_Index])

Payback Period (Years) = 800 / [Annual Revenue]
```

- Impact of temp & dust : Tunisia uses real weather / MENA reflects scaled conditions:

```
Efficiency Proxy = AVERAGE('solar_weather_clean
3'[energy_deltawh]) / AVERAGE('solar_weather_clean 3'[ghi])
```

- Best balance: solar + population : best regions = top-right

```
Balanced Site Score = [Annual Solar Irradiance] *
AVERAGE('MENA_Region'[Population_Index])
```

- Another measure for the **CO2 emissions avoided:**

```
CO2 Avoided (tons) =
(SUM('solar_weather_clean 3'[energy_deltawh]) / 1000)
* 0.7 * AVERAGE('MENA_Region'[Solar_Index])
```

- A new measure for **LCOE across Tunisia & MENA => lower = more competitive:**

```
LCOE = 800 / ((SUM('solar_weather_clean 3'[energy_deltawh])
/ 1000)
* AVERAGE('MENA_Region'[Solar_Index]))
```

- A new measure named **Feasibility Score:**

```
Feasibility Score = [Annual Solar Irradiance] *
AVERAGE('MENA_Region'[Grid_Index])
```

- Another new measure for the **Profitability with incentives:**

```
Profit with Incentives = [Annual Revenue] * 1.2
```

- A final measure for downtime risk: Cloud cover and wind speed were used as proxies for sandstorm risk, which is commonly applied in early-stage solar feasibility analyses when direct dust data are unavailable.

```
Downtime Risk Index = AVERAGE('solar_weather_clean 3'[temp])
+ AVERAGE('solar_weather_clean3'[clouds_all]) +
AVERAGE('solar_weather_clean 3'[wind_speed])
```

5. Conclusion:

This project shows where and when solar energy performs best in Tunisia and the MENA region, helping identify strong locations for investment.

Using business intelligence indicators such as energy potential, payback period, LCOE, and CO_2 emissions avoided, this project shows: where solar projects make the most sense and why. The analysis helps compare regions, highlight the most cost-effective locations, and estimate how quickly investments can be recovered. These results give investors and decision-makers a practical basis for choosing suitable sites, prioritizing regions, and evaluating the financial and environmental benefits of solar projects.

*However, it is important to note that the dataset used is not fully representative of all MENA countries. While regional scaling factors were applied, real-world implementation would require validation using localized and official data sources to ensure accuracy. Overall, this project demonstrates the value of business intelligence tools in supporting strategic renewable energy decisions in the region.