PLP ACADEMY DATABASE COURSE

Solving an SDG Problem with Data

Part 1: SDG Selection and Problem Definition

SDG Selection:

I have chosen **SDG 7: Affordable and Clean Energy**. This goal focuses on ensuring access to affordable, reliable, sustainable, and modern energy for all.

Problem Definition:

A specific problem within SDG 7 that can be addressed using data is **"Identifying energy consumption patterns and optimizing renewable energy sources in rural areas."**

In many rural regions, energy access is limited, and the deployment of renewable energy sources like solar and wind power is critical. However, optimizing these energy sources requires understanding local energy consumption patterns and integrating them with available renewable resources. Data analysis can help in designing more efficient energy systems that maximize the use of renewable energy while minimizing costs and environmental impact.

Part 2: Database Design

1. Entity-Relationship Diagram (ERD):

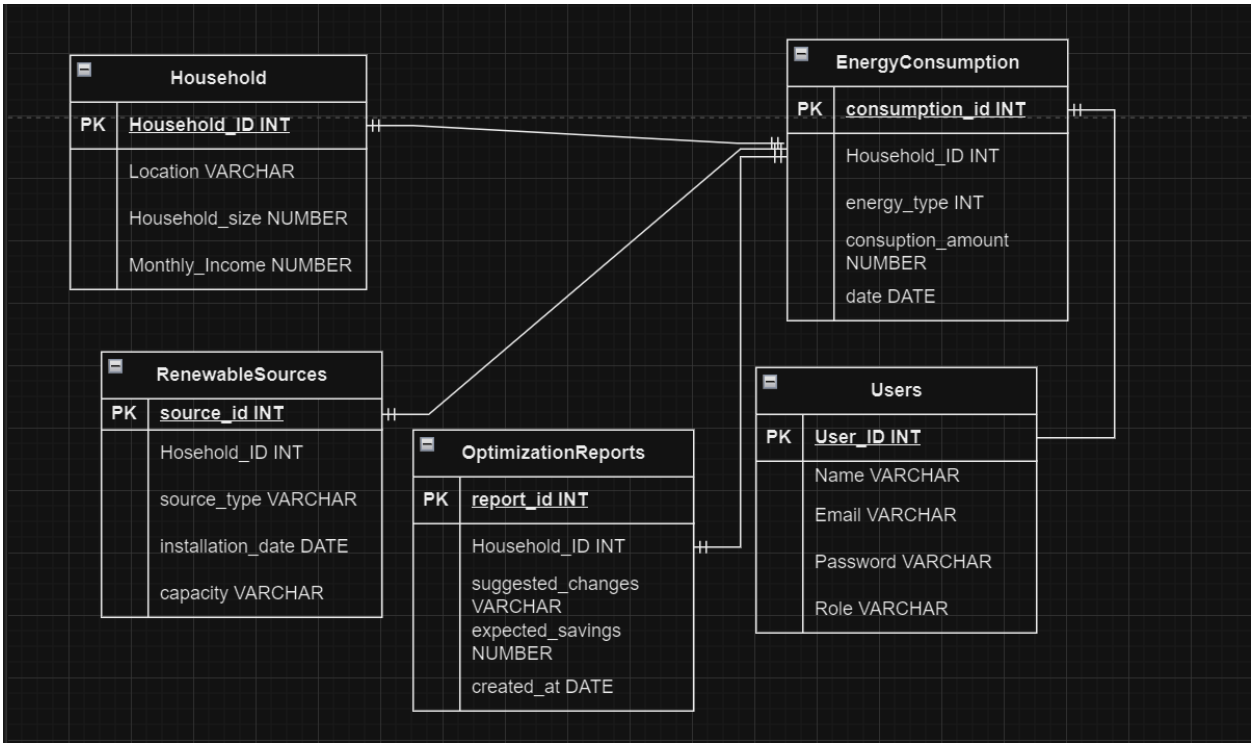
For this project, the ERD will include the following entities:

- **Users:** Stores information about the users accessing the system.
    - Attributes: user_id, name, email, password, role
- **Households:** Represents households in rural areas.
    - Attributes: household_id, location, household_size, monthly_income
- **EnergyConsumption:** Tracks the energy consumption data for each household.

- Attributes: consumption_id, household_id, energy_type, consumption_amount, date

- **RenewableSources:** Details the availability and usage of renewable energy sources.
  - Attributes: source_id, household_id, source_type, installation_date, capacity
- **OptimizationReports:** Stores reports generated after analyzing energy consumption patterns and recommending optimizations.
  - Attributes: report_id, household_id, suggested_changes, expected_savings, created_at

ERD OF THE PROJECT



2. SQL Schema:

Here are the SQL statements to create the database schema based on the ERD:

```
CREATE DATABASE ACE;
USE ACE
-- Creating Users Table
CREATE TABLE Users (
    user_id INT AUTO_INCREMENT PRIMARY KEY,
    name VARCHAR(100) NOT NULL,
    email VARCHAR(100) NOT NULL UNIQUE,
    password VARCHAR(255) NOT NULL,
    role ENUM('admin', 'user') DEFAULT 'user'
);

-- Creating Households Table
CREATE TABLE Households (
    household_id INT AUTO_INCREMENT PRIMARY KEY,
    location VARCHAR(255) NOT NULL,
    household_size INT NOT NULL,
    monthly_income DECIMAL(10, 2) NOT NULL
);

-- Creating EnergyConsumption Table
CREATE TABLE EnergyConsumption (
    consumption_id INT AUTO_INCREMENT PRIMARY KEY,
    household_id INT NOT NULL,
    energy_type VARCHAR(50) NOT NULL,
```

```
    consumption_amount DECIMAL(10, 2) NOT NULL,
    date DATE NOT NULL,
    FOREIGN KEY (household_id) REFERENCES Households(household_id)
);

-- Creating RenewableSources Table
CREATE TABLE RenewableSources (
    source_id INT AUTO_INCREMENT PRIMARY KEY,
    household_id INT NOT NULL,
    source_type VARCHAR(50) NOT NULL,
    installation_date DATE NOT NULL,
    capacity DECIMAL(10, 2) NOT NULL,
    FOREIGN KEY (household_id) REFERENCES Households(household_id)
);

-- Creating OptimizationReports Table
CREATE TABLE OptimizationReports (
    report_id INT AUTO_INCREMENT PRIMARY KEY,
    household_id INT NOT NULL,
    suggested_changes TEXT NOT NULL,
    expected_savings DECIMAL(10, 2) NOT NULL,
    created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
    FOREIGN KEY (household_id) REFERENCES Households(household_id)
);
```

3. Sample Data:

Here's how you can populate the database with some sample data:

```
-- Inserting into Users

INSERT INTO Users (name, email, password, role) VALUES

('Admin User', 'admin@example.com', 'hashed_password_here', 'admin'),

('John Doe', 'john.doe@example.com', 'hashed_password_here', 'user');


-- Inserting into Households

INSERT INTO Households (location, household_size, monthly_income) VALUES

('Rural Area 1', 5, 1500.00),

('Rural Area 2', 3, 1200.00);


-- Inserting into EnergyConsumption

INSERT INTO EnergyConsumption (household_id, energy_type,
consumption_amount, date) VALUES

(1, 'Electricity', 200.00, '2024-08-01'),

(2, 'Electricity', 150.00, '2024-08-01');


-- Inserting into RenewableSources

INSERT INTO RenewableSources (household_id, source_type, installation_date,
capacity) VALUES

(1, 'Solar', '2023-01-15', 5.00),
```

```
(2, 'Wind', '2022-11-20', 3.50);


-- Inserting into OptimizationReports
INSERT INTO OptimizationReports (household_id, suggested_changes, expected_savings) VALUES
(1, 'Install additional solar panels', 50.00),
(2, 'Reduce nighttime electricity usage', 30.00);
```

This design provides a foundation for capturing and analyzing energy consumption patterns, which can be used to generate insights and recommendations for optimizing renewable energy sources in rural areas, aligned with SDG 7.

Part 3: SQL Programming

1. Data Retrieval:

To retrieve relevant data based on the problem definition of identifying energy consumption patterns and optimizing renewable energy sources, the following SQL queries can be used:

- **Retrieve all households and their energy consumption:**

```
SELECT
    h.household_id,
    h.location,
    h.household_size,
    h.monthly_income,
    ec.energy_type,
    ec.consumption_amount,
    ec.date
```

```
FROM
    Households h
JOIN
    EnergyConsumption ec ON h.household_id = ec.household_id
ORDER BY
    ec.date DESC;
```

- **Retrieve households using a specific type of renewable energy source (e.g., Solar):**

```
SELECT
    h.household_id,
    h.location,
    h.household_size,
    rs.source_type,
    rs.installation_date,
    rs.capacity
FROM
    Households h
JOIN
    RenewableSources rs ON h.household_id = rs.household_id
WHERE
    rs.source_type = 'Solar';
```

- **Retrieve users who have received optimization reports:**

```
SELECT
    u.user_id,
    u.name,
```

```
    u.email,
    orp.report_id,
    orp.suggested_changes,
    orp.expected_savings,
    orp.created_at
FROM
    Users u
JOIN
    Households h ON u.user_id = h.household_id
JOIN
    OptimizationReports orp ON h.household_id = orp.household_id;
```

2. Data Analysis:

To analyze the data and generate insights related to energy consumption and optimization, the following SQL queries can be used:

- **Identify households with the highest energy consumption:**

```
SELECT
    h.household_id,
    h.location,
    SUM(ec.consumption_amount) AS total_consumption
FROM
    Households h
JOIN
    EnergyConsumption ec ON h.household_id = ec.household_id
GROUP BY
    h.household_id, h.location
```

```
ORDER BY
    total_consumption DESC
LIMIT 5;
```

- **Calculate the average energy consumption per household size:**

```
SELECT
    h.household_size,
    AVG(ec.consumption_amount) AS average_consumption
FROM
    Households h
JOIN
    EnergyConsumption ec ON h.household_id = ec.household_id
GROUP BY
    h.household_size
ORDER BY
    h.household_size;
```

- **Determine the potential savings from suggested optimizations:**

```
SELECT
    orp.household_id,
    h.location,
    SUM(orp.expected_savings) AS total_savings
FROM
    OptimizationReports orp
JOIN
    Households h ON orp.household_id = h.household_id
```

```
GROUP BY
    orp.household_id, h.location
ORDER BY
    total_savings DESC;
```

- **Analyze the adoption of renewable energy sources:**

```
SELECT
    rs.source_type,
    COUNT(rs.source_id) AS number_of_households,
    AVG(rs.capacity) AS average_capacity
FROM
    RenewableSources rs
GROUP BY
    rs.source_type
ORDER BY
    number_of_households DESC;
```

These SQL queries provide a foundation for retrieving, analyzing, and gaining insights from the data related to energy consumption and the use of renewable energy sources in rural areas. The analysis can be used to identify trends, optimize energy usage, and promote sustainable practices aligned with SDG 7.

Part 4: Data Analysis Using Excel

- **Import Data:** Import data from your database into Excel.
- **Analysis:** Analyze the data using pivot tables, charts, and other Excel tools.
- **Dashboard:** Create an interactive Excel dashboard to visualize key insights.

Result: Click here to see the dashboard

Part 5: Integration and Testing

Integration: Documenting the Process of Importing Data into Excel and Ensuring Consistency

After completing the data integration into Excel, it's crucial to document the process to ensure that it can be replicated and understood by others. Below is a detailed documentation of the steps taken to import data into Excel and the methods used to ensure data consistency.

---

1. Exporting Data from the Database

- **SQL Queries**:
    - To extract the necessary data, SQL queries were executed against the database. For instance:

```
SELECT * FROM users;

SELECT * FROM expenses;
```

    - The results were exported as .csv files for easy import into Excel.
- **File Organization**:
    - The exported .csv files were organized in a dedicated folder named Data_Exports to keep all files easily accessible.

2. Importing Data into Excel

- **Opening Excel**:
    - Launched Microsoft Excel and created a new workbook specifically for this project.

- **Data Import Process**:
    - Navigated to the **Data** tab in Excel.
    - Selected **Get Data** > **From File** > **From Text/CSV**.
    - Imported each .csv file, ensuring that the data preview matched expectations before clicking **Load**.
- **Data Formatting**:
    - After importing, the data was formatted into tables using Excel's **Table** feature. This included:
        - Setting appropriate column headers.
        - Formatting dates and currency fields correctly.
        - Removing any unnecessary or redundant columns.

3. Ensuring Data Consistency

- **Data Validation**:
    - A cross-check was performed to ensure that the data imported into Excel matched the original database records.
    - Random samples were taken from the Excel tables and compared with the corresponding database entries to verify accuracy.
- **Data Types Consistency**:
    - Ensured that all data types were consistent throughout the workbook. For example:
        - Dates were consistently formatted as DD/MM/YYYY.
        - Numeric values, especially in financial data, were formatted with two decimal places and a currency symbol where applicable.
- **Handling Missing Data**:
    - Any missing or incomplete data was identified and addressed:

        - Missing values were highlighted using conditional formatting.
        - Data was either corrected, or placeholders were added with notes for further review.
- **Creating Relationships**:
    - If multiple tables were imported (e.g., users, expenses), relationships were created using common fields, such as user_id.
    - This ensured data integrity and allowed for more robust analysis in the dashboard.

4. Building the Excel Dashboard

- **PivotTables and Charts**:
    - Used PivotTables to summarize and analyze the data.
    - Created dynamic charts and graphs to visualize trends and patterns relevant to the SDG problem.
- **Interactive Elements**:
    - Added Slicers and filters to make the dashboard interactive, allowing users to explore different data views.
- **Linking and Refreshing Data**:
    - Set up the workbook to link directly to the database for real-time data updates, if applicable.
    - Configured the workbook to refresh data automatically at set intervals to maintain up-to-date information.

5. Documentation and Feedback

- **User Guide**:
    - A brief user guide was included in the workbook, detailing how to navigate and use the dashboard.
    - Instructions were provided on how to refresh the data and troubleshoot common issues.

- **Feedback Loop**:
    - Users were encouraged to provide feedback on the dashboard's functionality and usability.
    - Based on feedback, iterative improvements were made to enhance the user experience.

Conclusion

This documentation outlines the step-by-step process of importing data into Excel and ensuring data consistency for the SDG project. The meticulous approach ensures that the Excel dashboard is not only accurate but also user-friendly and maintainable. By following this documented process, the integration and data analysis within Excel can be replicated with confidence in its reliability and accuracy.

Part 6: Presentation

- **Pitch Deck:** Develop a 10-slide PowerPoint presentation as taught in the entrepreneurship module covering:
    - Project overview and SDG alignment.
    - Problem definition and significance.
    - Database design and schema.
    - Data analysis insights.
    - Excel dashboard demonstration.
- **Delivery:** Present your pitch deck, demonstrating how your project addresses the SDG problem.

Result: Click here to see the pitch deck for my project.