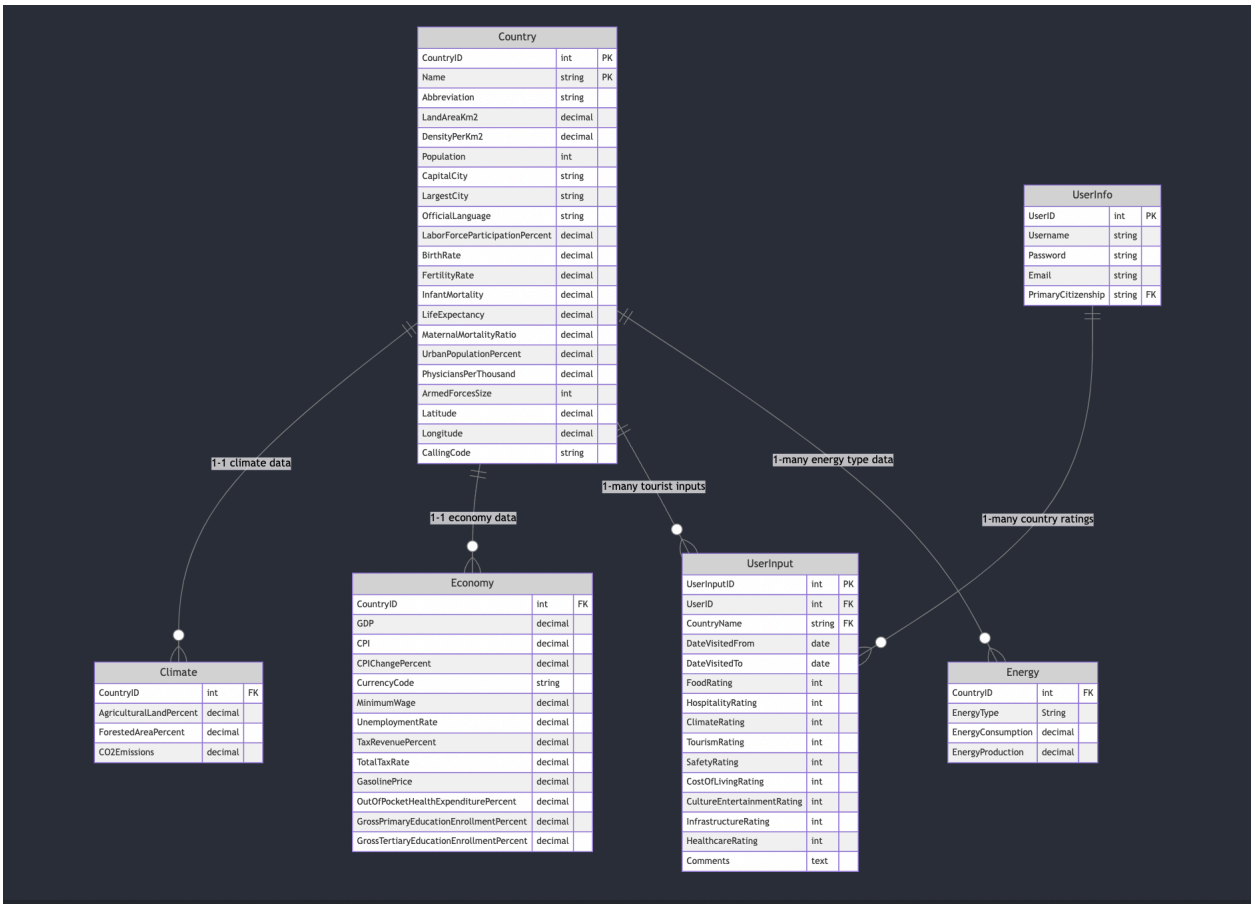
# Country Pollution App Conceptual and Logical Database Design

## Entity-Relationship Diagram (ERD)




### ER Diagram Description

The Entity-Relationship Diagram (ERD) for the Country Pollution App consists of six entities:

1. **Country**
2. **Climate**
3. **Economy**
4. **Energy**
5. **UserInfo**

6. **UserInput**

There are 5 total relationships between these entities (between 2 types):

- **Country** has a **1-to-1** relationship with **Climate**.
- **Country** has a **1-to-1** relationship with **Economy**.
- **Country** has a **1-to-many** relationship with **Energy**.
- **UserInfo** has a **1-to-many** relationship with **UserInput**.
- **Country** has a **1-to-many** relationship with **UserInput**.

**Assumptions and Explanations**

**1. Country**

- **Assumptions:**
  - Each country is uniquely identified by a CountryID.
  - A country's Name is also unique and can serve as another primary key.
  - Attributes like Population, CapitalCity, and OfficialLanguage are properties that describe the demographics and geographic information about the country.
- **Explanation:**
  - The Country entity is central to the application as it represents the primary subject of analysis.
  - We modeled Country as an entity to encapsulate all country-specific data in one place.
  - Attributes were chosen for users to learn facts about the country they're analyzing

**2. Climate**

- **Assumptions:**
  - Each country has one set of climate data.
  - Climate data includes metrics like AgriculturalLandPercent, ForestedAreaPercent, and CO2Emissions.
- **Explanation:**
  - Climate is modeled as a separate entity to manage the most up-to-date climate-related attributes, which are substantial and may change over time.
  - The 1-to-1 relationship with Country reflects that each country has a unique climate profile.

**3. Economy**

- **Assumptions:**
    - Each country has one set of economic data.
    - Economic data includes indicators like GDP, CPI, and UnemploymentRate.
- **Explanation:**
    - Similar to Climate, Economy is a separate entity to handle economic attributes efficiently.
    - The 1-to-1 relationship ensures that each country's economic data is directly linked to it.

**4. Energy**

- **Assumptions:**
    - A country can have multiple energy sources (natural gas, coal, solar, hydro, etc.).
    - Each energy source has specific data like EnergyType, EnergyConsumption, and EnergyProduction.
- **Explanation:**
    - Energy is modeled as an entity to accommodate multiple energy records per country.
    - This design avoids data redundancy and allows for detailed energy data management.
    - The 1-to-many relationship with Country shows that a country can have multiple energy sources.

**5. UserInfo**

- **Assumptions:**
    - Each user is uniquely identified by a UserID.
    - Users have attributes like Username, Email, and PrimaryCitizenship.
- **Explanation:**
    - UserInfo is the sole entity representing user accounts, adhering to the project requirement of having at most one user entity.
    - It stores essential user information for authentication and personalization.

**6. UserInput**

- **Assumptions:**
    - Users can provide multiple inputs or ratings for different countries.
    - Each input includes ratings on various aspects like FoodRating, SafetyRating, and Comments.
- **Explanation:**
    - UserInput captures user-generated content, crucial for the application's interactive features.

    - The 1-to-many relationship with UserInfo allows users to submit multiple inputs.
    - The 1-to-many relationship with Country enables aggregation of inputs for each country.

**Relationships and Cardinality**

**Country to Climate (1-to-1)**

- **Assumptions:**
    - Each country has one unique set of climate data.
- **Explanation:**
    - This relationship ensures that climate data is directly associated with its respective country without duplication.

**Country to Economy (1-to-1)**

- **Assumptions:**
    - Each country has one unique set of economic data.
- **Explanation:**
    - Economic data is specific to a country and doesn't vary per user, justifying the 1-to-1 relationship.

**Country to Energy (1-to-many)**

- **Assumptions:**
    - A country can have multiple energy sources.
- **Explanation:**
    - The 1-to-many relationship allows for multiple energy records (different energy sources) linked to a single country.

**UserInfo to UserInput (1-to-many)**

- **Assumptions:**
    - A user can submit multiple inputs for the same or different countries.
- **Explanation:**
    - This relationship enables users to provide numerous ratings and feedback entries over time.

**Country to UserInput (1-to-many)**

- **Assumptions:**
    - A country can have multiple inputs from different users.
- **Explanation:**

    - Aggregates user inputs for a country, facilitating collective analysis and integration of user feedback.

**Database Schema Normalization Process**

To ensure data integrity and eliminate redundancy, we have normalized our database schema. Below, we provide a detailed normalization process for each entity, including the identification of functional dependencies and demonstration of how each table satisfies the requirements for Boyce-Codd Normal Form (BCNF).

**1. First Normal Form (1NF)**

**Definition**: A table is in 1NF if all its attributes are atomic, meaning each attribute contains only indivisible values, and there are no repeating groups or arrays.

**Application to Our Schema**:

- All tables have atomic attributes with single values.
- There are no repeating groups or arrays in any of the entities.

**Conclusion**: All entities in our schema meet the requirements for 1NF.

**2. Second Normal Form (2NF)**

**Definition**: A table is in 2NF if it is in 1NF and all non-key attributes are fully functionally dependent on the entire primary key.

**Application to Our Schema**:

For each entity, we will list the primary key and functional dependencies (FDs):

**Country**

- **Primary Key**: CountryID
- **Functional Dependencies**:
    - CountryID → all other attributes in the Country table.

Since there is only one attribute in the primary key and all other attributes depend on it, there are no partial dependencies.

**Climate**

- **Primary Key**: CountryID
- **Functional Dependencies**:

    - CountryID → AgriculturalLandPercent, ForestedAreaPercent, CO2Emissions

All non-key attributes depend on the primary key.

**Economy**

- **Primary Key**: CountryID
- **Functional Dependencies**:
    - CountryID → all other attributes in the Economy table.

All non-key attributes depend on the primary key.

**Energy**

- **Primary Key**: (CountryID, EnergyType)
- **Functional Dependencies**:
    - (CountryID, EnergyType) → EnergyConsumption, EnergyProduction

All non-key attributes depend on the entire composite primary key.

**UserInfo**

- **Primary Key**: UserID
- **Functional Dependencies**:
    - UserID → Username, Password, Email, PrimaryCitizenshipID

All non-key attributes depend on the primary key.

**UserInput**

- **Primary Key**: UserInputID
- **Functional Dependencies**:
    - UserInputID → all other attributes in the UserInput table.

All non-key attributes depend on the primary key.

**Conclusion**: All tables are in 2NF as there are no partial dependencies.

**3. Third Normal Form (3NF)**

**Definition**: A table is in 3NF if it is in 2NF and all the attributes are dependent only on the primary key, not on any other non-key attributes

**Application to Our Schema**:

We need to check for transitive dependencies in each table.

**Country**

- **Functional Dependencies**:
  - CountryID → Name, Abbreviation, LandAreaKm2, DensityPerKm2, Population, CapitalCity, LargestCity, OfficialLanguage, LaborForceParticipationPercent, BirthRate, FertilityRate, InfantMortality, LifeExpectancy, MaternalMortalityRatio, UrbanPopulationPercent, PhysiciansPerThousand, ArmedForcesSize, Latitude, Longitude, CallingCode
  - **Assumption**: Name and Abbreviation are unique and could serve as candidate keys.
- **Transitive Dependencies**:
  - None, as all attributes depend directly on CountryID.

**Climate**

- **Functional Dependencies**:
  - CountryID → AgriculturalLandPercent, ForestedAreaPercent, CO2Emissions
- **Transitive Dependencies**:
  - None.

**Economy**

- **Functional Dependencies**:
  - CountryID → GDP, CPI, CPIChangePercent, CurrencyCode, MinimumWage, UnemploymentRate, TaxRevenuePercent, TotalTaxRate, GasolinePrice, OutOfPocketHealthExpenditurePercent, GrossPrimaryEducationEnrollmentPercent, GrossTertiaryEducationEnrollmentPercent
- **Transitive Dependencies**:
  - None.

**Energy**

- **Functional Dependencies**:
  - (CountryID, EnergyType) → EnergyConsumption, EnergyProduction
- **Transitive Dependencies**:
  - None.

**UserInfo**

- **Functional Dependencies**:

    - UserID → Username, Password, Email, PrimaryCitizenshipID
    - **Assumptions**:
        - Username and Email are unique and could serve as candidate keys.
- **Transitive Dependencies**:
    - None.

**UserInput**

- **Functional Dependencies**:
    - UserInputID → UserID, CountryID, DateVisitedFrom, DateVisitedTo, FoodRating, HospitalityRating, ClimateRating, TourismRating, SafetyRating, CostOfLivingRating, CultureEntertainmentRating, InfrastructureRating, HealthcareRating, Comments
- **Transitive Dependencies**:
    - None.

All tables are in 3NF as there are no transitive dependencies.

**4. Boyce-Codd Normal Form (BCNF)**

**Definition**: A table is in BCNF if it is in 3NF and, for every non-trivial functional dependency (X → Y), X is a superkey.

**Application to Our Schema**:

**Country**

- **Candidate Keys**: CountryID, Name, Abbreviation
- **Functional Dependencies**:
    - CountryID → all other attributes
    - Name → all other attributes
    - Abbreviation → all other attributes
- **Conclusion**:
    - All determinants (CountryID, Name, Abbreviation) are candidate keys.
    - No other functional dependencies violate BCNF.

**Climate**

- **Candidate Key**: CountryID
- **Functional Dependencies**:
    - CountryID → AgriculturalLandPercent, ForestedAreaPercent, CO2Emissions
- **Conclusion**:
    - The determinant CountryID is a candidate key.

    - No violations of BCNF.

**Economy**

- **Candidate Key**: CountryID
- **Functional Dependencies**:
    - CountryID → all other attributes
- **Conclusion**:
    - The determinant CountryID is a candidate key.
    - No violations of BCNF.

**Energy**

- **Candidate Key**: (CountryID, EnergyType)
- **Functional Dependencies**:
    - (CountryID, EnergyType) → EnergyConsumption, EnergyProduction
    - **Possible Additional Dependency**: EnergyType → EnergyProductionMethod (if such an attribute exists)
- **Conclusion**:
    - All determinants are superkeys.
    - No violations of BCNF.

**UserInfo**

- **Candidate Keys**: UserID, Username, Email
- **Functional Dependencies**:
    - UserID → Username, Password, Email, PrimaryCitizenshipID
    - Username → UserID, Password, Email, PrimaryCitizenshipID
    - Email → UserID, Username, Password, PrimaryCitizenshipID
- **Conclusion**:
    - All determinants are candidate keys.
    - No violations of BCNF.

**UserInput**

- **Candidate Key**: UserInputID
- **Functional Dependencies**:
    - UserInputID → all other attributes
    - **Composite Candidate Key Consideration**: (UserID, CountryID, DateVisitedFrom, DateVisitedTo) could potentially serve as a composite key if we assume a user cannot have multiple inputs for the same country and date range.
- **Conclusion**:

- The determinant UserInputID is a candidate key.
- No violations of BCNF.

All tables are in BCNF as all determinants of non-trivial functional dependencies are superkeys.

**Relational Schema**

**Country**

Country(
  CountryID: INT [PK],
  Name: VARCHAR(100),
  Abbreviation: VARCHAR(10),
  LandAreaKm2: DECIMAL,
  DensityPerKm2: DECIMAL,
  Population: INT,
  CapitalCity: VARCHAR(100),
  LargestCity: VARCHAR(100),
  OfficialLanguage: VARCHAR(100),
  LaborForceParticipationPercent: DECIMAL,
  BirthRate: DECIMAL,
  FertilityRate: DECIMAL,
  InfantMortality: DECIMAL,
  LifeExpectancy: DECIMAL,
  MaternalMortalityRatio: DECIMAL,
  UrbanPopulationPercent: DECIMAL,
  PhysiciansPerThousand: DECIMAL,
  ArmedForcesSize: INT,
  Latitude: DECIMAL,
  Longitude: DECIMAL,
  CallingCode: VARCHAR(10)
)

**Climate**

Climate(
  CountryID: INT [PK, FK to Country.CountryID],
  AgriculturalLandPercent: DECIMAL,
  ForestedAreaPercent: DECIMAL,
  CO2Emissions: DECIMAL
)

**Economy**

Economy(
  CountryID: INT [PK, FK to Country.CountryID],
  GDP: DECIMAL,
  CPI: DECIMAL,
  CPIChangePercent: DECIMAL,
  CurrencyCode: VARCHAR(10),
  MinimumWage: DECIMAL,
  UnemploymentRate: DECIMAL,
  TaxRevenuePercent: DECIMAL,
  TotalTaxRate: DECIMAL,
  GasolinePrice: DECIMAL,
  OutOfPocketHealthExpenditurePercent: DECIMAL,
  GrossPrimaryEducationEnrollmentPercent: DECIMAL,
  GrossTertiaryEducationEnrollmentPercent: DECIMAL
)

**Energy**

Energy(
  CountryID: INT [FK to Country.CountryID],
  EnergyType: VARCHAR(50),
  EnergyConsumption: DECIMAL,
  EnergyProduction: DECIMAL,
  [PK: CountryID, EnergyType]
)

**UserInfo**

UserInfo(
  UserID: INT [PK],
  Username: VARCHAR(50),
  Password: VARCHAR(50),
  Email: VARCHAR(100),
  PrimaryCitizenshipID: INT [FK to Country.CountryID]
)

**UserInput**

UserInput(
  UserInputID: INT [PK],
  UserID: INT [FK to UserInfo.UserID],
  CountryID: INT [FK to Country.CountryID],
  DateVisitedFrom: DATE,
  DateVisitedTo: DATE,
  FoodRating: INT,
  HospitalityRating: INT,
  ClimateRating: INT,
  TourismRating: INT,
  SafetyRating: INT,
  CostOfLivingRating: INT,
  CultureEntertainmentRating: INT,
  InfrastructureRating: INT,
  HealthcareRating: INT,
  Comments: TEXT
)

**Summary**

- **Entities:** The database includes six entities—Country, Climate, Economy, Energy, UserInfo, and UserInput—each serving a specific purpose in the application.
- **Relationships:** The schema includes various relationships with cardinalities such as 1-to-1 and 1-to-many, satisfying the requirement of having at least two types of relationships.
- **Normalization:** The database schema is normalized to BCNF, ensuring minimal redundancy and optimal data integrity.
- **Relational Schema:** The logical design translates the conceptual ERD into a relational schema, formatted as per the specified guidelines.