Analysis of Population Demographics Using SQL Queries

Contributor: Debjyoti Saha

Executive Summary

This project explores demographic data to calculate male and female populations using SQL. The focus was on leveraging relational database principles, including joins and mathematical computations, to derive insights from datasets. It also outlines the process of setting up a SQL database and conducting exploratory data analysis.

Introduction

Background:

Analyzing demographic data is essential for understanding population structures and making informed policy decisions. This project aimed to demonstrate the use of SQL in solving real-world problems.

Objective:

To calculate male and female populations from given data using the sex ratio and population statistics while exploring data setup and visualization techniques in SQL.

Scope:

The project included creating databases, importing datasets, and writing analytical queries.

Methodology

- **Database Setup**:
    - Created a database and imported demographic datasets.
    - Visualized and explored the data to understand its structure.
- **Key Techniques**:
    - Joins: To combine datasets on common attributes like district names.
    - SQL Functions: Used mathematical operations to calculate demographic metrics.
    - Filtering: Limited data to specific states for focused analysis.

Project Implementation

Step 1: Setting up the Database

- Created a database following step-by-step instructions: naming the database, importing data, and verifying its structure.
- Queried basic dataset information (e.g., row counts).

Step 2: Solving the Problem Statement

- Applied mathematical formulas to compute male and female populations using the given relationships:
    - Males = Population / (Sex Ratio + 1)
    - Females = Population - Males
- **Example Query:**

```
SELECT Gender_Ratio.District, Gender_Ratio.State,
    ROUND(Gender_Ratio.Population/(Gender_Ratio.Sex_Ratio + 1), 0) AS Male,
    ROUND(((Gender_Ratio.Population) -((Gender_Ratio.Population)/(Gender_Ratio.Sex_Ratio + 1))), 0) AS Female
FROM ...
```

Step 3: Data Exploration

- Queried datasets for basic statistics like row counts and subsets for specific states (e.g., Maharashtra and West Bengal).

Results and Findings

- **Male and Female Populations**: Computed populations for each district using sex ratio and total population.
- **Descriptive Statistics**: Verified row counts and analyzed subsets for specific states to validate the data integrity.

Discussion

- **Significance**: The project demonstrates the power of SQL in deriving meaningful insights from relational data.
- **Challenges**: Managing data consistency across tables and ensuring correct formulas were used.
- **Limitations**: The analysis relied heavily on the accuracy of the provided sex ratio and population data.

Conclusion

This SQL project provided a hands-on demonstration of analyzing demographic data. The calculated populations and exploratory queries showcased SQL's versatility in solving complex problems. Future work could integrate visualization tools or more extensive datasets for deeper insights.

References

- SQL Queries provided in uploaded files.
- Documentation for SQL functions and joins.