**SQL Analysis Documentation: Solving Problem Statement with Data Analysis**

**Overview**

This document provides a comprehensive explanation of the SQL queries used to analyze demographic and literacy data.

The script employs advanced SQL techniques such as joins, derived formulas, and window functions to address various analytical objectives.

**Sections of Analysis**

**1. Data Exploration** - Queries to inspect the contents of two datasets:

SELECT \* FROM project..Data1;

SELECT \* FROM project..Data2;

**2. Gender and Population Analysis**

**- Objective**: Calculate the male and female population based on population size and sex ratio.

**- Key Formulas**:

- `Females / Males = Sex\_Ratio`

**- Derived:**

- `Males = Population / (Sex\_Ratio + 1)`

- `Females = Population - Males`

**- Implementation:**

**- At the district level:**

SELECT District, State, ROUND(Population / (Sex\_Ratio + 1), 0) AS Male,

ROUND(Population - Population / (Sex\_Ratio + 1), 0) AS Female

FROM Gender\_Ratio;

**- At the state level (aggregated):**

SELECT State, SUM(Male) AS Total\_Male, SUM(Female) AS Total\_Female

FROM State\_Level GROUP BY State;

**3. Literacy Distribution**

- **Objective**: Compute literate and illiterate populations using the literacy ratio.

**- Key Formulas:**

- `Total\_Literate = Literacy\_Ratio \* Population`

- `Total\_Illiterate = (1 - Literacy\_Ratio) \* Population`

**- Implementation:**

**- At the district level:**

SELECT District, State, ROUND(Literacy\_Ratio \* Population, 0) AS Literate,

ROUND((1 - Literacy\_Ratio) \* Population, 0) AS Illiterate

FROM Literacy\_Distribution;

**- At the state level:**

SELECT State, SUM(Literate) AS Total\_Literate\_Population,

SUM(Illiterate) AS Total\_Illiterate\_Population

FROM State\_Level GROUP BY State;

**4. Census Comparisons**

- **Objective**: Analyze growth trends by comparing current and previous census data.

**- Key Formula:**

- `Previous\_Census = Population / (1 + Growth\_Ratio)`

**- Implementation:**

**- District-level calculation:**

SELECT District, State, ROUND(Population / (1 + Growth\_Ratio), 0) AS Previous\_Census

FROM Census;

**- Aggregated state-level data:**

SELECT State, SUM(Previous\_Census) AS Previous\_Census,

SUM(Current\_Census) AS Current\_Census FROM State\_Level;

**- National population:**

SELECT SUM(Previous\_Census) AS Previous\_Census, SUM(Current\_Census) AS Current\_Census

FROM India;

**5. Population vs. Area**

- **Objective**: Calculate population density using census data and total area.

**- Key Steps:**

**- Assign a common key to census and area datasets:**

SELECT '1' AS Key, SUM(Area\_km2) AS Total\_Area FROM project..Data2;

**- Join the datasets and calculate densities:**

SELECT (Total\_Area / Previous\_Census) AS Previous\_Census\_Population,

(Total\_Area / Current\_Census) AS Current\_Census\_Population

FROM Decreased\_Area;

**6. Window Functions for Literacy Rates**

- **Objective**: Identify districts with top and bottom literacy rates within each state.

**- Implementation:**

**- Top 3 districts:**

SELECT State, District, Literacy, RANK() OVER (PARTITION BY State ORDER BY Literacy DESC) AS Ranks

FROM project..Data1 WHERE Ranks <= 3;

**- Bottom 3 districts:**

SELECT State, District, Literacy, RANK() OVER (PARTITION BY State ORDER BY Literacy ASC) AS Ranks

FROM project..Data1 WHERE Ranks <= 3;

**- Combine results for top and bottom:**

SELECT State, District, Literacy, 'Top' AS Rank\_Type FROM Highest

UNION ALL

SELECT State, District, Literacy, 'Bottom' AS Rank\_Type FROM Lowest;

**Recommendations**

- Validate the data consistency (e.g., ensure no null or invalid values in `Population` or `Sex\_Ratio` columns).

- Incorporate visualization tools for deeper insights into the results (e.g., heatmaps, bar charts).

- Extend analysis to include additional demographic factors like age distribution.

**This documentation aims to provide a clear understanding of the SQL file's logic and its application in solving the problem statement.**