**Census Dataset**

Abstract:

The Indian Census 2011 is a comprehensive dataset providing valuable demographic and geographical information about various districts and states across India. This census dataset contains 641 rows, each representing a specific administrative district or region. The objective of this analysis is to explore and understand key population demographics and geographical characteristics based on the Indian Census 2011 data. By utilizing SQL queries, we aim to gain insights into population growth, literacy rates, sex ratio, and other crucial factors contributing to the diverse landscape of India. Through rigorous data analysis and visualization, this project seeks to shed light on the distribution of demographic variables across different regions and provide valuable insights for decision-making and policy formulation.

Introduction:

The Indian Census 2011 is one of the largest demographic exercises in the world, providing comprehensive data on various aspects of the Indian population. With 641 rows representing districts and regions, this dataset encompasses essential information such as population, literacy rates, sex ratio, and growth rates for each administrative division. Understanding this vast dataset is crucial for policymakers, researchers, and organizations seeking to comprehend India's diverse demographic landscape.

The primary objective of this analysis is to perform a data-driven examination of the Indian Census 2011 data using SQL queries. By leveraging SQL's powerful capabilities, we aim to delve into demographic trends and patterns to gain deeper insights into the socio-economic and geographical characteristics of the Indian population. We will explore the growth rates of different regions, assess literacy levels, investigate the sex ratio, and analyze population density in various districts and states.

This project seeks to present the findings in a clear and organized manner, employing SQL queries to extract and manipulate the relevant data. We will explore trends, correlations, and variations across the diverse regions of India. Additionally, where possible, data visualizations will be employed to enhance the presentation of the results.

Through this census dataset analysis, we hope to contribute valuable insights to the understanding of India's demographic diversity and provide actionable information for policymakers and stakeholders. By unraveling demographic patterns and geographic trends, we aim to offer a comprehensive and data-backed perspective on the Indian population's composition and characteristics.

Objectives:

* Explore Population Growth:

Analyze population growth rates for different districts and states in India based on the Indian Census 2011 data.

Identify regions with high population growth and regions experiencing slower growth.

* Assess Literacy Rates:

Evaluate literacy rates for various districts and states to understand the educational status of the population.

Identify areas with high literacy rates and regions that require attention in terms of education.

* Investigate Sex Ratio:

Analyze the sex ratio (the ratio of males to females) for each district and state.

Identify regions with significant gender imbalances and regions with a balanced sex ratio.

* Examine Population Density:

Calculate population density (population per square kilometer) for each district.

Identify densely populated regions and regions with lower population density.

* Understand Demographic Diversity:

Gain a deeper understanding of the demographic diversity across different districts and states in India.

Identify variations in population characteristics and demographic trends.

* Enhance Data Literacy and Analysis Skills:

Enhance proficiency in SQL and data analysis techniques by working with a real-world census dataset.

Strengthen data management and data manipulation skills.

The above objectives outline the key areas of focus for the census dataset analysis project. By accomplishing these objectives, the project aims to uncover valuable insights about the Indian population's demographic and geographical characteristics and contribute to data-backed decision-making and policy formulation.

**Analysis**

---- **avg-growth**

select avg(Growth) as avg\_Growth from project.dataset1;

select State, avg(Growth) as avg\_Growth from project.dataset1

group by state;

|  |  |
| --- | --- |
| State | avg\_Growth |
| Nagaland | 82.28 |
| Dadra and Nagar Haveli | 55.88 |
| Daman and Diu | 42.74 |
| Puducherry | 34.30 |
| Meghalaya | 30.39 |
| Arunachal Pradesh | 27.81 |
| Manipur | 25.54 |
| Bihar | 25.23 |
| Mizoram | 24.35 |
| Jammu and Kashmir | 24.27 |

--- Average Population Density by State

SELECT

State,

AVG(Population / Area\_km2) AS Avg\_Population\_Density

FROM project.dataset2

GROUP BY State

order by Avg\_Population\_Density desc ;

|  |  |
| --- | --- |
| State | Top 5 Avg\_Population\_Density |
| Goa | 729 |
| Punjab | 276.92 |
| Himachal Pradesh | 227.71 |
| Assam | 226.15 |
| Tripura | 207 |

-- **avg sex ratio**

select State, round(avg(Sex\_Ratio),0) as avg\_sex\_Ratio from project.dataset1

group by state

order by avg\_sex\_Ratio desc;

-- **avg literacy rate**

select State, round(avg(literacy),0) as avg\_literacy\_Ratio from project.dataset1

group by state having round(avg(literacy),0)> 90

order by avg\_literacy\_Ratio desc;

--- **top 3 state showing highest growth**

select State, avg(Growth) as avg\_Growth from project.dataset1

group by state

order by avg\_Growth desc limit 3 ;

--- **bottom 3 state showing lowest sex ratio**

select State, round(avg(Sex\_Ratio),0) as avg\_sex\_Ratio from project.dataset1

group by state

order by avg\_sex\_Ratio asc limit 3;

--- **states starting with letter a**

select distinct state from project.dataset1

where lower(state) like 'a%'or lower(state) like 'b%';

---**total males and females**

select d.state,sum(d.males) total\_males, sum(d.females) total\_females from

(select c.district,c.state,round(c.population/(c.sex\_ratio+1),0)males,round((c.population\*c.sex\_ratio)/(c.sex\_ratio+1),0) females from

(select a.district,a.state,a.sex\_ratio/1000 sex\_ratio, b.population from project.dataset1 a

inner join project.dataset2 b on a.district=b.district)c)d

group by d.state;

select c.state, sum(literarte\_people) total\_literate\_pop,sum(illiterate\_people) total\_illiterate\_pop

from

(select d.district,d.state,round(d.literacy\_ratio\*d.population,0)literate\_people,

round(1-d.literacy\_ratio)\* d.population,0) illiterate\_people from

(select a.district,a.state,a.literacy/100 literacy\_ratio,b.population from

project.dataset1 a

inner join project.dataset2 b on a.district=b.district)d)c

group by c.state;

select a.\* from

(select district,state,literacy,rank() over(partition by state order by literacy desc) rnk

from project.dataset1)a

where a.rnk in (1,2,3) order by state;

SELECT state, district, population,

SUM(population) OVER (PARTITION BY state) AS state\_population,

AVG(population) OVER (PARTITION BY state) AS state\_avg\_population

FROM project.dataset2;

ff