**INDIAN CENSUS 2011 DATA ANALYSIS**

**Data Overview**

*The dataset of the Indian census of 2011, structured into two distinct tables. The first dataset comprises columns encompassing geographical and demographic aspects of the population, including information such as district, state, sex ratio, population growth rate, and literacy rate. Meanwhile, the second dataset contains columns like district, state, area in square kilometres, and population count.*

*In my analysis, I formulated various questions to explore the dataset comprehensively, aiming to uncover insights into the population trends and characteristics. Each question was selected with a specific purpose in mind, based on its relevance to demographic patterns and regional dynamics. The conclusions drawn from these questions are presented, shedding light on notable findings and contributing to a deeper understanding of the data's implications.*

**Questions and Reasons**

1. **Calculation of total number of rows in both the dataset.**

**Reason:** Verifying the number of rows in an SQL dataset is crucial as it offers an essential measure of data volume and completeness. This information helps ensure data integrity, aids in identifying potential data discrepancies, and provides a fundamental understanding of the dataset's scale and scope.

**Finding:** Both the dataset has exactly **640 rows** of data.

1. **Finding the population of India.**

Reason: Before we proceed with the in-depth analysis of the Indian Census 2011 data, it's essential to establish the total population figure that our analysis encompasses. The analysis utilized Dataset2 because it contained the necessary information to determine the population of each district.

**Finding:** The sum of the population of every district is, **1210854977**.

1. **Average growth percentage of India.**

**Reason:** Calculating the average growth percentage of India aids in comprehending the pace of the country's overall population expansion. This information assists in estimating future population sizes after a specific number of years, enabling informed projections and planning.

**Finding:** The average rate of growth of India’s population is **19.24%.**

1. **Average growth percentage state wise and also display the top 3.**

**Reason:** In dataset1, the growth percentage is presented on a district level, offering a detailed perspective of the data. However, for a broader understanding and to formulate more effective strategies, a more comprehensive overview might be more beneficial. Zooming out to view the data on a larger scale could provide a clearer insight into the trends and help in identifying actionable steps.

**Finding:** Highest growth% state is **Nagaland,** followed by **Dadra**, and **Daman** with **82.28%, 55.88%,** and **42.74%** respectively.

1. **Average sex ratio of different states and find the worst 3 performers.**

**Reason:** Determining the average gender distribution across various states can aid in tailoring product offerings to suit specific regional demographics. This approach ensures that products are aligned with the preferences of different states' populations, enhancing the potential for successful market penetration.

**Finding:** The highest ratio is of Kerala’s with 1080 Females per 1000 Males. The worst performers are Dadra, Daman and Chandigarh.

1. **Literacy rate of different states and also states with greater than 90%.**

**Reason:** The literacy rate serves as a significant parameter for determining the most effective marketing approach. This factor ensures that marketing materials resonate better with the audience by considering their level of understanding and engagement.

**Finding:** **Kerala** again comes on the top with the highest literacy rate in India, with **94%,** followed by **Lakshadweep** with **92%.**

1. **Top and bottom 3 states in literacy rates.**

**Reason:** Finding the extreme edges helps us in understanding the spread of the data that we are dealing with.

**Finding:** The top 3 are, Kerala, Lakshadweep and Mizoram with 94%, 92%, 89%, respectively, and the bottom 3 are Rajasthan, Arunachal Pradesh, and Bihar with 65%, 64% and 62% respectively.

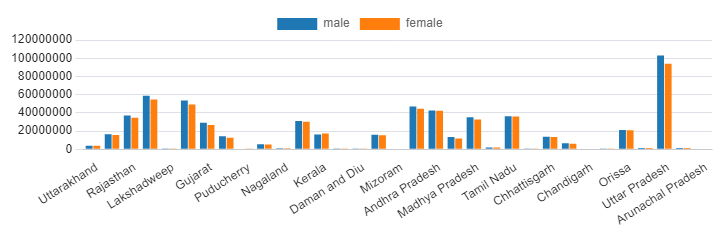
1. **States starting with a letter ‘A’ or ‘B’**.

**Reason:** This question helps to display the power of LIKE function

**Finding:** States that starts with the letter ‘A’ are, Andaman and Nicobar Islands, Andhra Pradesh, Arunachal Pradesh, Assam. For letter ‘B’ is only Bihar.

1. **Calculate the number of males and females.**

**Reason:** In our earlier analysis, we focused solely on calculating the average sex ratio, which provided a percentage-based perspective. However, this approach didn't offer a detailed understanding of the actual male and female populations in different states. To address this limitation, I've now incorporated the real male and female population figures for each state, allowing for a more comprehensive and accurate assessment.

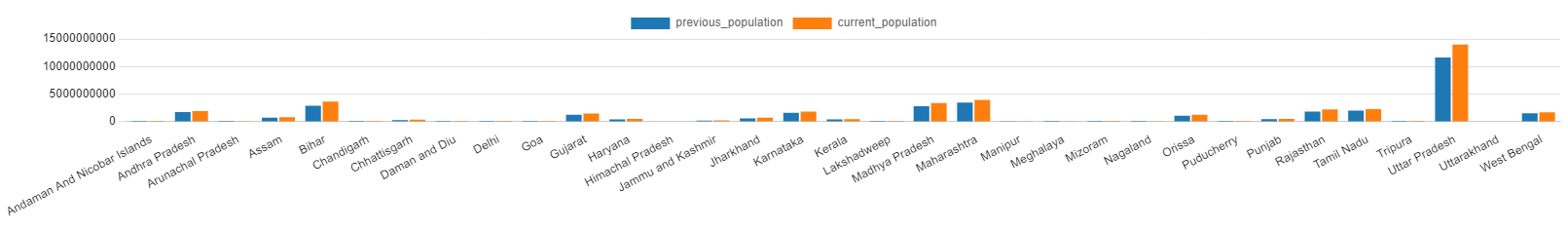
**Finding:**

**State Wise Gender Distribution**

1. **Actual population in previous census and in current census.**

Reason: The difference in values will help us understand at which pace the population is growing at. To calculate the previous census, I have subtracted the growth percentage from the current census data.

Finding:



**State Wise Population Change**

1. **How the change in population influenced the area km2 of the population.**

**Reason:** As the country's population grows, the available land area per person is likely to decrease. This could lead to a more condensed living space, potentially resulting in the construction of skyscrapers and tall buildings to accommodate the increasing population within limited land resources.

**Finding:**

|  |  |
| --- | --- |
| **Area km2 (Previous Census)** | **Area km2 (Current Census)** |
| 0.04806182205366204 | 0.0026745920896968024 |

1. **Calculate the top 3 districts with highest literacy rates from each district.**

**Reason:** The primary objective of this calculation is to showcase the effectiveness of window functions in SQL. These functions simplify complex coding tasks by allowing us to achieve significant results through straightforward steps.

**End of report**