***Indian\_Population\_Analysis***

**The Following are the SQL queries fired for the problem statements**

**Creating seperate tables for both the datasets, i.e., Dataset1 and Dataset2\*/**

**Creating table 'Dataset1'**

CREATE TABLE dataset1

(

district VARCHAR(255),

state VARCHAR(255),

growth float,

sex\_ratio float,

literacy float

)

**Creating table 'Dataset2'**

CREATE TABLE dataset2

(

district VARCHAR(255),

state VARCHAR(255),

area\_km2 float,

population int8

);

**Importing data from the CSV file to the table 'Dataset1'**

COPY dataset1(district,state,growth,sex\_ratio,literacy)

FROM 'D:\Sunil\Dataset1.csv'

DELIMITER ','

CSV HEADER;

**Importing data from the CSV file to the table 'Dataset2'**

COPY dataset2(district,state,area\_km2,population)

FROM 'D:\Sunil\Dataset2.csv'

DELIMITER ','

CSV HEADER;

**Checking how the data looks like**

SELECT \* FROM dataset1

SELECT \* FROM dataset2

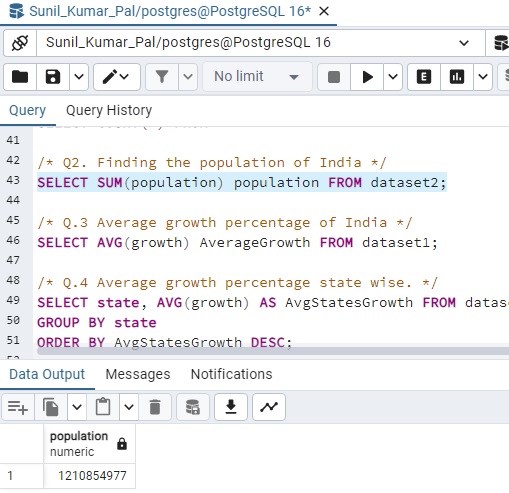
**Q1. Calculation of total number of rows in both the dataset**

SELECT COUNT(\*) FROM dataset1

SELECT COUNT(\*) FROM dataset2

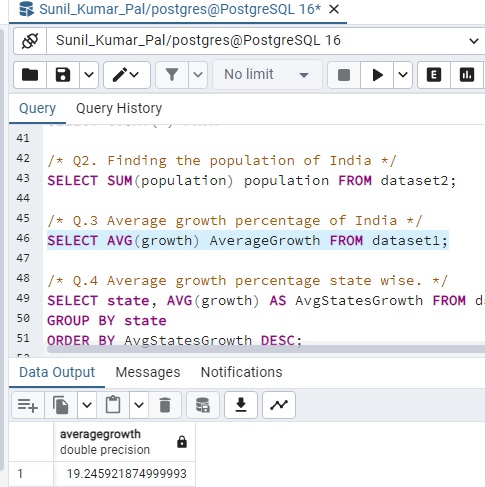
**Q2. Finding the population of India**

SELECT SUM(population) population FROM dataset2;



**Q.3 Average growth percentage of India**

SELECT AVG(growth) AverageGrowth FROM dataset1;

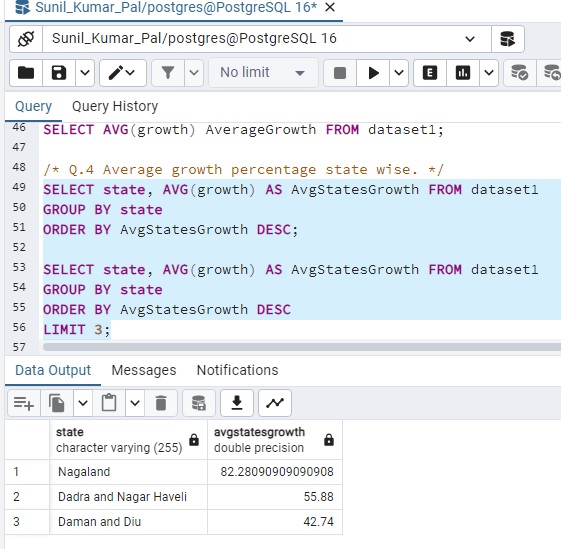


**Q.4 Average growth percentage state wise.**

SELECT state, AVG(growth) AS AvgStatesGrowth FROM dataset1

GROUP BY state

ORDER BY AvgStatesGrowth DESC;



**Q.5 Average sex ratio of different states and find the worst 3 performers.**

SELECT state, ROUND(AVG(sex\_ratio)) AS sex\_ratio FROM dataset1

GROUP BY state

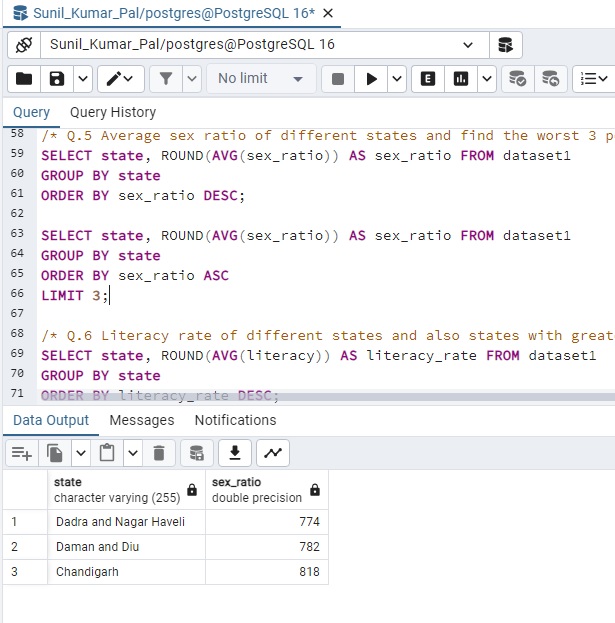
ORDER BY sex\_ratio DESC;

SELECT state, ROUND(AVG(sex\_ratio)) AS sex\_ratio FROM dataset1

GROUP BY state

ORDER BY sex\_ratio ASC

LIMIT 3;



**Q.6 Literacy rate of different states and also states with greater than 90%**

SELECT state, ROUND(AVG(literacy)) AS literacy\_rate FROM dataset1

GROUP BY state

ORDER BY literacy\_rate DESC;

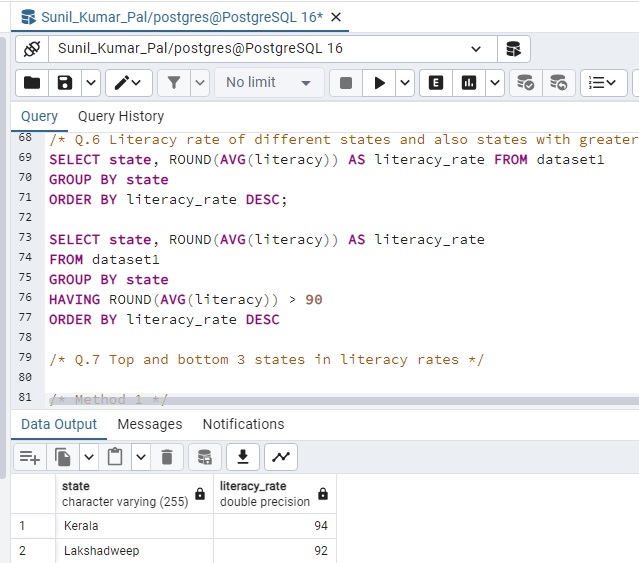
SELECT state, ROUND(AVG(literacy)) AS literacy\_rate

FROM dataset1

GROUP BY state

HAVING ROUND(AVG(literacy)) > 90

ORDER BY literacy\_rate DESC



**Q.7 Top and bottom 3 states in literacy rates**

(SELECT state, ROUND(AVG(literacy)) AS literacy\_rate

FROM dataset1

GROUP BY state

ORDER BY literacy\_rate ASC

LIMIT 3)

UNION

(SELECT state, ROUND(AVG(literacy)) AS literacy\_rate

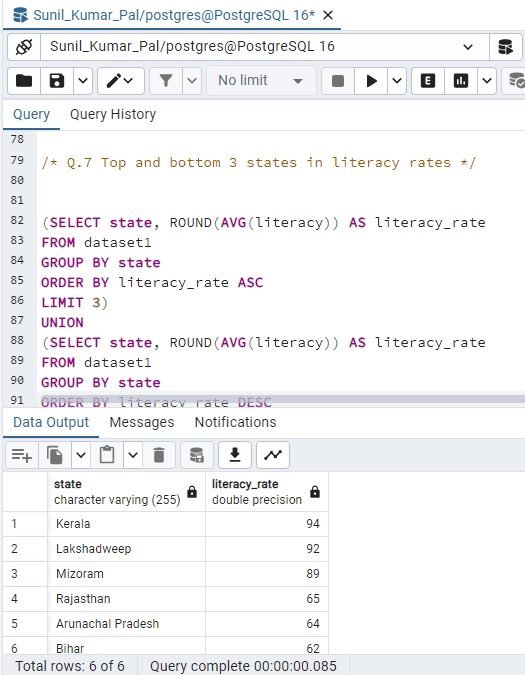
FROM dataset1

GROUP BY state

ORDER BY literacy\_rate DESC

LIMIT 3)

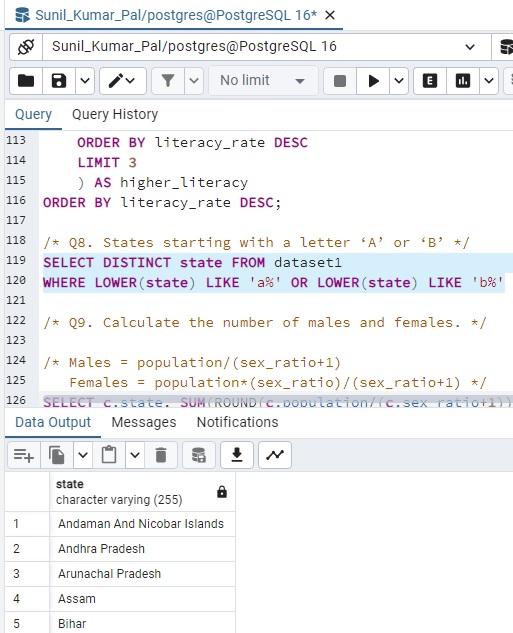
ORDER BY literacy\_rate DESC



**Q8. States starting with a letter ‘A’ or ‘B’**

SELECT DISTINCT state FROM dataset1

WHERE LOWER(state) LIKE 'a%' OR LOWER(state) LIKE 'b%'



**Q9. Calculate the number of males and females.**

Males = population/(sex\_ratio+1)

Females = population\*(sex\_ratio)/(sex\_ratio+1)

SELECT c.state, SUM(ROUND(c.population/(c.sex\_ratio+1))) AS male, SUM(ROUND(c.population\*(c.sex\_ratio)/(c.sex\_ratio+1))) AS female

FROM

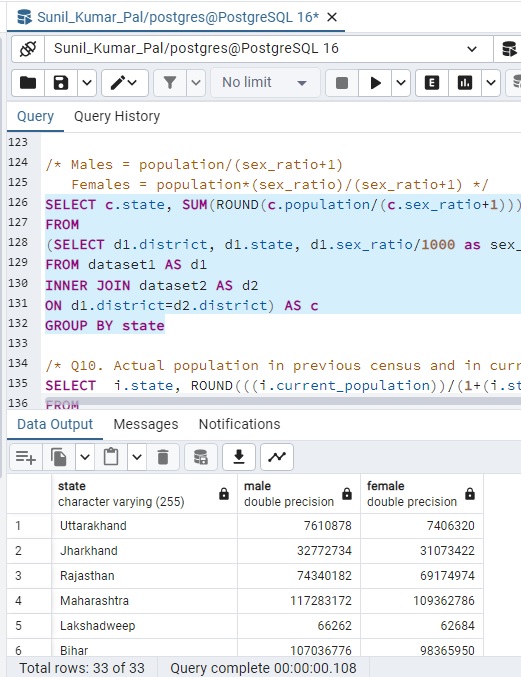
(SELECT d1.district, d1.state, d1.sex\_ratio/1000 as sex\_ratio, d2.population

FROM dataset1 AS d1

INNER JOIN dataset2 AS d2

ON d1.district=d2.district) AS c

GROUP BY state



**Q10. Actual population in previous census and in current census.**

SELECT i.state, ROUND(((i.current\_population))/(1+(i.states\_growth/100))) AS previous\_population, i.current\_population

FROM

(SELECT d1.state,

(SUM(d1.growth)) / (COUNT(d1.growth)) AS states\_growth,

SUM(d2.population) AS current\_population

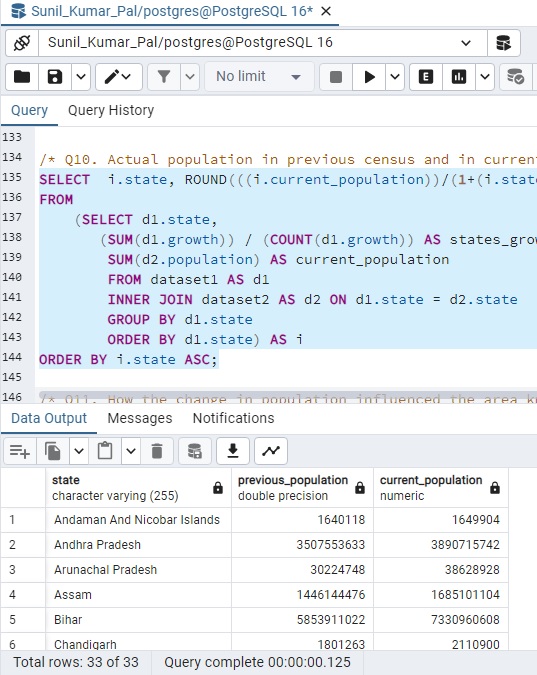
FROM dataset1 AS d1

INNER JOIN dataset2 AS d2 ON d1.state = d2.state

GROUP BY d1.state

ORDER BY d1.state) AS i

ORDER BY i.state ASC;



**Q11. How the change in population influenced the area km2 of the population.**

SELECT

(g.total\_area / g.previous\_census\_population) AS previous\_census\_population\_vs\_area,

(g.total\_area / g.current\_census\_population) AS current\_census\_population\_vs\_area

FROM (

SELECT q.\*, r.total\_area

FROM (

SELECT '1' AS keyy, n.\*

FROM (

SELECT

SUM(m.previous\_census\_population) AS previous\_census\_population,

SUM(m.current\_census\_population) AS current\_census\_population

FROM (

SELECT e.state,

SUM(e.previous\_census\_population) AS previous\_census\_population,

SUM(e.current\_census\_population) AS current\_census\_population

FROM (

SELECT d.district, d.state, ROUND(d.population / (1 + d.growth)) AS previous\_census\_population, d.population AS current\_census\_population

FROM (

SELECT a.district, a.state, a.growth, b.population

FROM dataset1 a

INNER JOIN dataset2 b ON a.district = b.district

) d

) e

GROUP BY e.state

) m

) n

) q

INNER JOIN (

SELECT '1' AS keyy, z.\*

FROM (

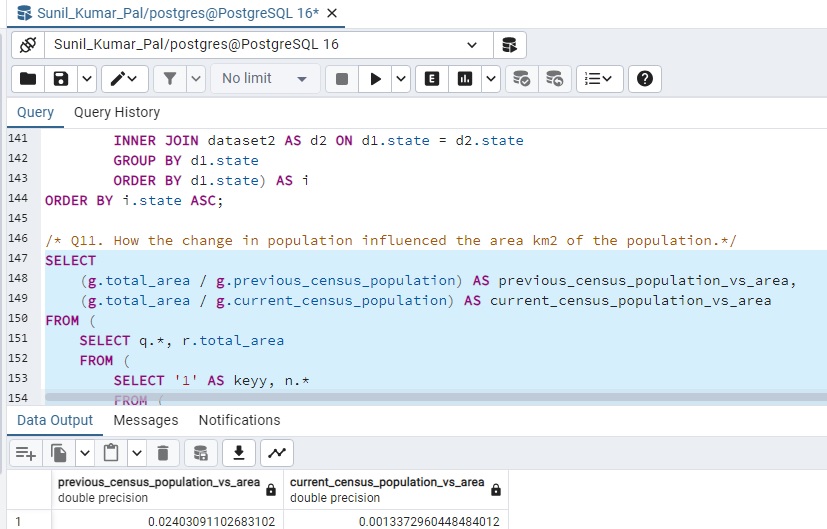
SELECT SUM(area\_km2) AS total\_area

FROM dataset2

) z

) r ON q.keyy = r.keyy

) g;



**Q12. Calculate the top 3 districts with highest literacy rates from each district**

SELECT a.\* FROM

(SELECT district, state, literacy, RANK() OVER(PARTITION BY state

ORDER BY literacy DESC) AS rnk FROM dataset1) AS a

WHERE a.rnk in (1,2,3) ORDER BY state

