

INDIAN CENSUS DATA ANALYSIS USING SQL

Description:

You, as a data analyst working for the Government of India, need to focus on developing insights from the Indian Census data so that the government can use your insights for taking important decisions such as fund allocation in different parts of the country. Your insights will also help the government to make important decisions on resource allocation for public welfare.

Dataset Description:

Dataset 1:

- District - Name of districts
- State - Names of different states of India
- Growth - Percentage of growth in population
- Sex_ratio - Number of females per 1000 males
- Literacy - Percentage of literate population

Dataset 2:

- District - Name of districts
- State - Names of different states of India
- Area_km2 - Area in KM²
- Population - Population of the given district

Operations to be performed:

1. Count number of rows in the dataset

Query:

```
SELECT COUNT(*) FROM Data1;  
SELECT COUNT(*) FROM Data2;
```

Output:

Results		Messages	
	(No column name)		
1	640		

	(No column name)		
1	640		

2. Show data for Jharkhand and Bihar only

Query:

```
SELECT * FROM Data1 WHERE State in ('Jharkhand', 'Bihar');
```

Output:

	District	State	Growth	Sex_Ratio	Literacy
11	Sitamarhi	Bihar	0.2762	899	52.05
12	Siwan	Bihar	0.227	988	69.45
13	Purnia	Bihar	0.2833	921	51.08
14	Katihar	Bihar	0.2835	919	52.24
15	Bhagalpur	Bihar	0.2536	880	63.14
16	Begusarai	Bihar	0.2644	895	63.87
17	Rohtas	Bihar	0.2078	918	73.37
18	Ranchi	Jharkhand	0.2398	949	76.06
19	Nalanda	Bihar	0.2139	922	64.43
20	Araria	Bihar	0.3025	921	53.53
21	Bhojpur	Bihar	0.2163	907	70.47
22	Dhanbad	Jharkhand	0.1199	909	74.52
23	Gopalganj	Bihar	0.1902	1021	65.47
24	Aurangabad	Bihar	0.2618	926	70.32
25	Giridih	Jharkhand	0.2841	944	63.14
26	Purbi Singhbhum	Jharkhand	0.1568	949	75.49
27	Supaul	Bihar	0.2866	929	57.67
28	Nawada	Bihar	0.2263	939	59.76
29	Bokaro	Jharkhand	0.1601	922	72.01
30	Banka	Bihar	0.2648	907	58.17
31	Madhepura	Bihar	0.3112	911	52.25
32	Palamu	Jharkhand	0.2617	928	63.63

3. Calculate the total population of India

Query:

```
SELECT SUM(population) Population from Data2;
```

Output:

Results		Messages	
	Population		
1	1210854977		

4. Calculate average population growth rate of India

Query:

```
SELECT AVG(growth)*100 AS avg_growth from Data1;
```

Output:

Results		Messages	
	avg_growth		
1	19.245921875		

5. Calculate average growth rate by individual states.

Query:

```
SELECT state,AVG(growth)*100 AS avg_growth  
FROM Data1 GROUP BY state;
```

Output:

	state	avg_growth
1	Andaman And Nicobar Islands	0.596666666666667
2	Andhra Pradesh	10.9239130434783
3	Arunachal Pradesh	27.805625
4	Assam	16.5237037037037
5	Bihar	25.2318421052632
6	Chandigarh	17.19
7	Chhattisgarh	20.035
8	Dadra and Nagar Haveli	55.88
9	Daman and Diu	42.74
10	Delhi	13.8877777777778
11	Goa	8.285
12	Gujarat	17.0823076923077
13	Haryana	20.4923809523809
14	Himachal Pradesh	11.3425
15	Jammu and Kashmir	24.2704545454546
16	Jharkhand	23.7816666666667

6. Calculate sex ratio by individual states

Query:

```
SELECT state, ROUND(AVG(sex_ratio),0) sex_ratio FROM Data1  
GROUP BY state;
```

Output:

	state	sex_ratio
1	Andaman And Nicobar Islands	858
2	Andhra Pradesh	995
3	Arunachal Pradesh	920
4	Assam	958
5	Bihar	918
6	Chandigarh	818
7	Chhattisgarh	995
8	Dadra and Nagar Haveli	774
9	Daman and Diu	783
10	Delhi	866
11	Goa	975
12	Gujarat	938
13	Haryana	880
14	Himachal Pradesh	953
15	Jammu and Kashmir	883
16	Jharkhand	956
17	Karnataka	984
18	Kerala	1080
19	Lakshadweep	946

7. Calculate average literacy rates by states.

Query:

```
SELECT state, ROUND(AVG(literacy),0) literacy_rate  
FROM Data1 GROUP BY state;
```

Output:

	state	literacy_rate
1	Andaman And Nicobar Islands	84
2	Andhra Pradesh	66
3	Arunachal Pradesh	64
4	Assam	72
5	Bihar	62
6	Chandigarh	86
7	Chhattisgarh	66
8	Dadra and Nagar Haveli	76
9	Daman and Diu	86
10	Delhi	87
11	Goa	89
12	Gujarat	76
13	Haryana	75
14	Himachal Pradesh	82
15	Jammu and Kashmir	65
16	Jharkhand	65
17	Karnataka	74
18	Kerala	94
19	Lakshadweep	92

8. Display states having literacy rates greater than 90%.

Query:

```
SELECT state, ROUND(AVG(literacy),0) literacy_rate  
FROM Data1 GROUP BY state
```

```
HAVING ROUND(AVG(literacy),0)>90;
```

Output:

Results Messages		
	state	literacy_rate
1	Kerala	94
2	Lakshadweep	92

9. Display the top 3 states having the highest population growth rate

Query:

```
SELECT TOP 3 state, AVG(growth)*100 avg_growth
FROM Data1 GROUP BY state
ORDER BY avg_growth DESC;
```

Output:

Results Messages		
	state	avg_growth
1	Nagaland	82.2809090909091
2	Dadra and Nagar Haveli	55.88
3	Daman and Diu	42.74

10. Display the top 3 states having the lowest sex ratio

Query:

```
SELECT TOP 3 state, ROUND(AVG(sex_ratio),0) avg_sex_ratio  
FROM Data1 GROUP BY state  
ORDER BY avg_sex_ratio ASC;
```

Output:

	state	avg_sex_ratio
1	Dadra and Nagar Haveli	774
2	Daman and Diu	783
3	Chandigarh	818

11. Display states starting with letter 'a'.

Query:

```
SELECT DISTINCT state FROM Data1 WHERE state like 'A%';
```

Output:

	state
1	Andaman And Nicobar Islands
2	Andhra Pradesh
3	Arunachal Pradesh
4	Assam

12. Display states starting with either letter 'A' or 'B';

Query:

```
SELECT DISTINCT state FROM Data1 WHERE state LIKE 'A%' or
state LIKE 'B%';
```

Output:

Results		Messages
	state	
1	Andaman And Nicobar Islands	
2	Andhra Pradesh	
3	Arunachal Pradesh	
4	Assam	
5	Bihar	

13. Count total number of males and females by state.

Query:

```
SELECT d.state, SUM(d.males) total_males, SUM(d.females)
total_females FROM
```

```
(SELECT c.district,c.state
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 Data1 a INNER JOIN Data2 b ON
a.district=b.district ) c) d GROUP BY d.state;
```

Output:

	state	total_males	total_females
1	Andaman And Nicobar Islands	148014	126970
2	Andhra Pradesh	42439617	42141160
3	Arunachal Pradesh	624752	582402
4	Assam	15939857	15265719
5	Bihar	53518388	49182975
6	Chandigarh	580556	474894
7	Chhattisgarh	13659264	13518218
8	Daman and Diu	150264	92983
9	Delhi	77939	64065
10	Goa	739239	719306
11	Gujarat	30423412	27889194
12	Haryana	13495175	11856287
13	Himachal Pradesh	5337119	5263833
14	Jammu and Kashmir	6640603	5900699
15	Jharkhand	16386367	15536711

14. Count total number of literate and illiterate people by state

Query:

```
SELECT c.state,SUM(literate_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 Data1 a
INNER JOIN Data2 b ON a.district=b.district) d) c
GROUP BY c.state
```

Output:

Results		Messages	
	state	total_literate_pop	total_illiterate_pop
1	Andaman And Nicobar Islands	241015	33969
2	Andhra Pradesh	56671677	27909100
3	Arunachal Pradesh	761557	445597
4	Assam	22484409	8721167
5	Bihar	63994271	38707092
6	Chandigarh	908215	147235
7	Chhattisgarh	18621154	8556328
8	Daman and Diu	211827	31420
9	Delhi	125446	16558
10	Goa	1293736	164809
11	Gujarat	45836339	12476267
12	Haryana	19123489	6227973
13	Himachal Pradesh	8883924	1717028
14	Jammu and Kashmir	8391150	4150152
15	Jharkhand	21132066	10791012

15. Calculate total population in previous census using growth data.

Query:

```
SELECT SUM(m.previous_census_population)
previous_census_population, SUM(m.current_census_population)
current_census_population
FROM(SELECT e.state, SUM(e.previous_census_population)
previous_census_population, SUM(e.current_census_population)
current_census_population
FROM(SELECT
d.district, d.state, ROUND(d.population/(1+d.growth),0)
previous_census_population, d.population
current_census_population
FROM(SELECT a.district, a.state, a.growth growth, b.population
FROM data1 a INNER JOIN data2 b ON a.district=b.district) d)
e GROUP BY e.state)m
```

Output:

Results Messages		
	previous_census_population	current_census_population
1	1007610297	1186354496

16. Output 3 districts from each state having the highest literacy rates.

Query:

```
SELECT a.* FROM
(SELECT district,state,literacy,RANK() OVER(PARTITION BY
STATE ORDER BY literacy DESC) literacy_rank FROM data1) a
WHERE a.literacy_rank in (1,2,3) ORDER BY STATE;
```

Output:

	district	state	literacy	literacy_rank
1	South Andaman	Andaman And Nicobar Islands	89.13	1
2	North And Middle Andaman	Andaman And Nicobar Islands	83.91	2
3	Nicobars	Andaman And Nicobar Islands	78.06	3
4	Hyderabad	Andhra Pradesh	83.25	1
5	Rangareddy	Andhra Pradesh	75.87	2
6	West Godavari	Andhra Pradesh	74.63	3
7	Papumpare	Arunachal Pradesh	79.95	1
8	Lower Subansiri	Arunachal Pradesh	74.35	2
9	East Siang	Arunachal Pradesh	72.54	3
10	Kamrup Metropolitan	Assam	88.71	1
11	Jorhat	Assam	82.15	2
12	Sivasagar	Assam	80.41	3
13	Rohtas	Bihar	73.37	1
14	Patna	Bihar	70.68	2
15	Bhojpur	Bihar	70.47	3
16	Chandigarh	Chandigarh	86.05	1
17	Durg	Chhattisgarh	79.06	1
18	Dhamtari	Chhattisgarh	78.36	2