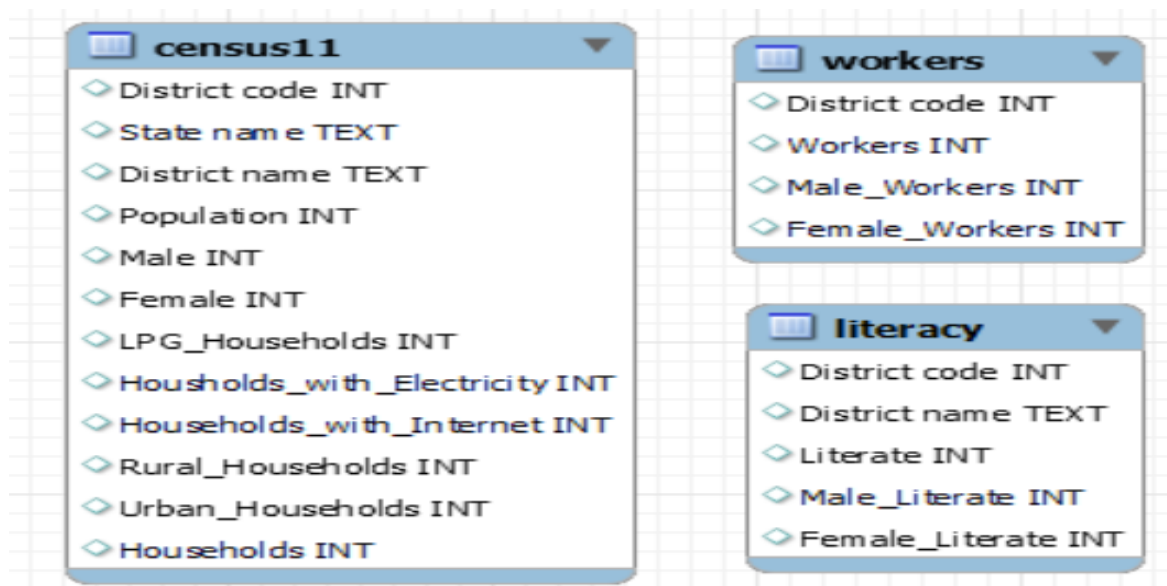


Analysing India's Census(2011) with SQL: Unveiling Socio-Demographic Insights

Problem Statement: This project employs SQL to derive valuable insights from the 2011 Indian census, exploring factors like literacy rate, sex ratio, workforce composition, electricity/internet connectivity, and urban-rural household dynamics. The analysis aims to shed light on socio-economic trends, providing a comprehensive understanding of India's population and enabling informed policy-making, optimized resource allocation, and development strategies.

Technique Used: SQL query writing, aggregate functions, joins, subqueries, window-functions, views

Tables Used:



The image shows three database tables in a software interface. The 'census11' table has 13 columns: District code (INT), State name (TEXT), District name (TEXT), Population (INT), Male (INT), Female (INT), LPG_Households (INT), Housholds_with_Electricity (INT), Households_with_Internet (INT), Rural_Households (INT), Urban_Households (INT), and Households (INT). The 'workers' table has 5 columns: District code (INT), Workers (INT), Male_Workers (INT), and Female_Workers (INT). The 'literacy' table has 6 columns: District code (INT), District name (TEXT), Literate (INT), Male_Literate (INT), and Female_Literate (INT).

Query1. Total number of records in the census table?

```
select count(*) from census11;
```

count(*)
640

Query2. Total population of India as per the census?

```
select sum(population) from census11;
```

total_population
1210854977

Query3. What is the Male to Female ratio of population as per census?

```
select sum(male) as total_male, sum(female) as total_female, sum(male)/sum(female) as  
ratio, round(sum(male)/sum(female)*1000,0) as male_per_1000_female  
from census11;
```

total_male	total_female	ratio	male_per_1000_female
623270258	587584719	1.0607	1061

Query4. Which State has the highest population in India

```
select `state name`, sum(population) as Total_Population  
from census11  
group by `state name`  
order by total_population desc limit 1;
```

state name	Total_Population
UTTAR PRADESH	199812341

Query5. Which State/UT has the lowest population in India as per census?

```
Select `state name`, sum(population) as Total_Population  
from census11  
group by `state name`  
order by total_population limit 1;
```

state name	Total_Population
LAKSHADWEEP	64473

Query6. What is the literacy rate of India as per census?

```
select sum(literate)/sum(population)*100 as Literacy_rate  
from literacy as l join census11 as c  
using(`District code`);
```

Literacy_rate
63.0661

Query7. What is the Male literacy rate of India as per census ?

```
select sum(male_literate)/sum(male)*100 as Male_LiteracyRate
from literacy as l join census11 as c
using(`District code`);
```

Male_LiteracyRate
69.7552

Query8. What is the Female literacy rate of India as per census?

```
select sum(female_literate)/sum(female)*100 as Female_LiteracyRate
from literacy as l join census11 as c
using(`District code`);
```

Female_LiteracyRate
55.9707

Query9. Which are the 5 most literate State/UT in India ?

```
select `State name`,sum(literate)/sum(population)*100 as Literacy_rate
from literacy as l join census11 as c
using(`District code`)
group by `state name`
order by Literacy_rate desc limit 5;
```

State name	Literacy_rate
KERALA	84.2237
LAKSHADWEEP	81.5116
GOA	79.9075
DAMAN AND DIU	77.4546
ANDAMAN AND NICOBAR ISLANDS	77.3241

Query10. Which are the 5 least literate State/UT in India ?

```
select `State name`,sum(literate)/sum(population)*100 as Literacy_rate
from literacy as l join census11 as c
```

using(`District code`) group by `state name` order by Literacy_rate limit 5;

State name	Literacy_rate
BIHAR	50.4369
ARUNACHAL PRADESH	55.3581
JHARKHAND	55.5596
RAJASTHAN	55.8368
JAMMU AND KASHMIR	56.3517

Query11. Which District in India has the highest literate population of India?

```
select c.`district name`,sum(literate)/sum(population)*100 as Literacy_rate, `state name`  
from literacy as l join census11 as c  
using(`District code`)  
group by c.`district name`  
order by Literacy_rate desc limit 1;
```

district name	Literacy_rate	state name
Pathanamthitta	88.7375	KERALA

Query12. Which District in India has the lowest literate population of India?

```
Select c.`district name`,sum(literate)/sum(population)*100 as Literacy_rate, `state name`  
from literacy as l join census11 as c  
using(`District code`)  
group by c.`district name`  
order by Literacy_rate limit 1;
```

district name	Literacy_rate	state name
Alirajpur	28.7729	MADHYA PRADESH

Query13. What percentage of India's total population is working population?

```
select round(sum(workers)/sum(population)*100,0) as working_population  
from census11 join workers using(`district code`);
```

working_population
40

Query14. What percentage of workers in India are women?

```
Select sum(female_workers)/sum(workers)*100 as `Percentage of Female worker`  
from workers;
```

Percentage of Female worker
31.1169

Query15. What percentage of workers in India are men?

```
select sum(male_workers)/sum(workers)*100 as `Percentage of Male worker`  
from workers;
```

Percentage of Male worker
68.8831

Query16. Which are the top 3 States/UTs with highest percentage of female workers?

```
Select `state name`, sum(female_workers)/sum(workers)*100 as `percentage of female workers`  
from census11 join workers using(`district code`)  
group by `state name`  
order by `percentage of female workers` desc  
limit 3;
```

state name	percentage of female workers
NAGALAND	43.8102
MANIPUR	43.3234
HIMACHAL PRADESH	42.5926

Ques17. Which are the top 3 States/UTs with highest percentage of Male workers

```
select `state name`, sum(male_workers)/sum(workers)*100 as `percentage of male workers`  
from census11 join workers  
using(`district code`)  
group by `state name`
```

order by `percentage of male workers` desc limit 3;

state name	percentage of male workers
DAMAN AND DIU	88.5900
NCT OF DELHI	85.2333
LAKSHADWEEP	81.6829

Query18. what percentage of households in India has LPG connection ?

Select sum(LPG_Households)/sum(households)*100

from census11;

Percentage of households with LPG
21.2864

Query19. Rank the districts from each State based on highest to lowest percentage of households with LPG connection

with subq as (select `state name`,`district name`,
sum(LPG_Households)/sum(households)*100 as 'Percentage of household with LPG'

from census11

group by `district name`)

select `state name`,`district name`,`Percentage of household with LPG`,dense_rank()

over(partition by `state name` order by `Percentage of household with LPG` desc) as ranking

from subq;

state name	district name	Percentage of household with LPG	ranking
ANDAMAN AND NICOBAR ISLANDS	South Andaman	37.4106	1
ANDAMAN AND NICOBAR ISLANDS	Nicobars	22.8349	2
ANDAMAN AND NICOBAR ISLANDS	North AND Middle Andaman	14.6746	3
ANDHRA PRADESH	Hyderabad	54.5358	1
ANDHRA PRADESH	Rangareddy	49.5251	2
ANDHRA PRADESH	Krishna	42.8690	3
ANDHRA PRADESH	Guntur	40.9711	4
ANDHRA PRADESH	Visakhapatnam	35.9149	5
ANDHRA PRADESH	West Godavari	34.6706	6
ANDHRA PRADESH	East Godavari	29.8855	7

Query20. Find top 3 districts from each state with highest percentage of households with LPG Connection

```
SELECT `State Name`, `District Name`, (LPG_households / Households) * 100 AS
Percentage_LPG, ranking

FROM (SELECT `State Name`, `District Name`, Households, LPG_households,

        ROW_NUMBER() OVER (PARTITION BY `State Name` ORDER BY (LPG_households /
Households) DESC) AS Ranking

FROM census11) AS RankedData

WHERE Ranking <= 3;
```

State Name	District Name	Percentage_LPG	ranking
ANDAMAN AND NICOBAR ISLANDS	South Andaman	37.4106	1
ANDAMAN AND NICOBAR ISLANDS	Nicobars	22.8349	2
ANDAMAN AND NICOBAR ISLANDS	North AND Middle Andaman	14.6746	3
ANDHRA PRADESH	Hyderabad	54.5358	1
ANDHRA PRADESH	Rangareddy	49.5251	2
ANDHRA PRADESH	Krishna	42.8690	3
ARUNACHAL PRADESH	Papum Pare	45.4880	1
ARUNACHAL PRADESH	West Kameng	29.4813	2
ARUNACHAL PRADESH	Tawang	28.7817	3
ASSAM	Kamrup Metropolitan	52.1654	1
ASSAM	Dibrugarh	20.2358	2

Query21. Which state has the highest and lowest percentage of households with LPG connection

```
SELECT (SELECT `State Name`

FROM census11

GROUP BY `State Name`

ORDER BY (SUM(LPG_households) / SUM(Households))*100 DESC

LIMIT 1) AS State_With_Highest_LPG,

(SELECT `State Name`

FROM census11

GROUP BY `State Name` ORDER BY (SUM(LPG_households) / SUM(Households))*100 ASC

LIMIT 1) AS State_With_Lowest_LPG;
```

State_With_Highest_LPG	State_With_Lowest_LPG
NCT OF DELHI	BIHAR

Query22. What percentage of households in India has Electricity connection ?

```
select round(sum(Housholds_with_Electricity)/sum(households)*100,2) as
```

```
'Percentage of households with electricity'
```

```
from census11;
```

Percentage of households with electricity
50.14

Query23. Which are top 5 State/UT with highest percentage of households having Electricity connection ?

```
Select `state name`, round(sum(Housholds_with_Electricity)/sum(households)*100,2) as
```

```
`Percentage of households with electricity`
```

```
from census11
```

```
group by `state name` order by `Percentage of households with electricity` desc
```

```
limit 5;
```

state name	Percentage of households with electricity
CHANDIGARH	77.49
PONDICHERRY	75.83
ANDHRA PRADESH	75.71
TAMIL NADU	74.53
NCT OF DELHI	71.89

Query24. Which are top 5 State/UT with lowest percentage of households having Electricity connection

```
Select `state name`, round(sum(Housholds_with_Electricity)/sum(households)*100,2) as
```

```
`Percentage of households with electricity` from census11
```

```
group by `state name` order by `Percentage of households with electricity`
```

```
limit 5;
```


state name	Percentage of households with electricity
BIHAR	13.23
ASSAM	25.99
UTTAR PRADESH	26.83
ORISSA	32.57
JHARKHAND	36.30

Query25. What percentage of households in India is with Internet connection ?

Select round(sum(Households_with_Internet)/sum(households)*100,2)

as `Percentage of households with Internet`

from census11;

Percentage of households with Internet
2.33

Query26. What percentage of total households in India are Rural households ?

Select round(sum(Rural_Households)/sum(households)*100,2)

as `Percentage of rural households`

from census11;

Percentage of rural households
66.71

Query27. which State/UT has the highest and lowest percentage of rural households in India ?

SELECT (SELECT `State Name`

FROM census11

GROUP BY `State Name`

ORDER BY (SUM(Rural_Households) / SUM(Households))*100 DESC

LIMIT 1) AS State_With_Highest_Rural_population,

(SELECT `State Name`

FROM census11

GROUP BY `State Name`

ORDER BY (SUM(Rural_Households) / SUM(Households))*100 ASC

LIMIT 1) AS State_With_Lowest_Rural_population;

State_With_Highest_Rural_population	State_With_Lowest_Rural_population
HIMACHAL PRADESH	NCT OF DELHI

Query28. What percentage of total households in India are Urban households ?

Select round(sum(urban_Households)/sum(households)*100,2)

as `Percentage of Uural households`

from census11;

Percentage of Urban households
33.29

Ques29. Create a view of States/UTs having literacy rate is above 60%

create view Top_Literate_States as

select `state name`, sum(literate)/sum(population)*100 as Literacy_rate

from literacy as l join census11 as c

using(`District code`)

group by `state name`

having Literacy_rate >60;

state name	Literacy_rate
HIMACHAL PRADESH	73.4163
PUNJAB	67.4293
CHANDIGARH	76.3123
UTTARAKHAND	68.2208
HARYANA	65.4755
NCT OF DELHI	75.8745
SIKKIM	72.8740
NAGALAND	67.8510
MANIPUR	66.8282
MIZORAM	77.3032
TRIPURA	76.3431

Summary and Conclusion

- The analysis of the 2011 Indian census data using SQL revealed several significant insights. The population of India in 2011 was 1,210,854,977 (around 1.2 billion), with

a male-to-female sex ratio of 1,061, meaning that for every 1,061 males, there are approximately 1,000 females in the population. It indicates that there is a higher proportion of males compared to females in the given population.

- Uttar Pradesh had the highest population, while Lakshadweep had the lowest.
- The overall literacy rate was 63%, with males at 69% and females at 56%. Kerala had the highest literacy rate, while Bihar had the lowest.
- The workforce constituted 40% of the population, with 31% women and 69% men. Only 21% of households had an LPG connection, while electricity and internet connectivity were low at 50% and 2%, respectively. Rural households accounted for 67% of the total, while urban households comprised 33%.

These insights can contribute to informed policy-making, optimized resource allocation, and development strategies.

By comparing the insights from the 2011 census with the figures of from 2021 census, we can observe the progress and changes that have occurred over the decade. It allows us to assess the effectiveness of policies and interventions implemented during this period. We can analyse trends in population growth, literacy rates, gender ratios, workforce composition, connectivity, and urban-rural dynamics. Any improvements or challenges identified can guide future decision-making, enabling targeted interventions to further enhance socio-economic development and address emerging needs for the well-being of the population.