INDIAN CENCUS 2011 ANALYSIS

**CODE**

select \* from Data1;

select \* from Data2;

-- 1. No. of Rows Present in the Dataset

select COUNT(\*) from Data1;

select COUNT(\*) from Data2;

--2. Dataset for delhi, WB

select \* from Data1 where state in ('delhi','West Bengal')

order by state;

--3. Population of India

select SUM(population) as 'total population' from Data2

--4. Avg Growth for each state

select state, AVG(growth)\*100 as 'Avg\_growth' from Data1

group by state

order by Avg\_growth;

--5. Avg sex-ratio for each state

select state, round(AVG(sex\_ratio),0) as 'Avg\_sex\_ratio' from Data1

group by State

order by Avg\_sex\_ratio desc;

--6. Avg litracy rate for each state

select state, round(AVG(Literacy),1) as 'Avg\_literacy' from Data1

group by State

having round(AVG(Literacy),1) >85

order by Avg\_literacy desc;

--7. Top 3 states having highest growth

select top 3 state, AVG(growth)\*100 as 'Avg\_growth' from Data1

group by state

order by Avg\_growth desc;

--8. bottom 3 states having lowest litrecy

select top 3 state, round(AVG(Literacy),1) as 'Avg\_literacy' from Data1

group by State

order by Avg\_literacy asc;

--9. Top and bottom 3 states in literacy rate into 1 single table (using temporary tables and union)

drop table top\_states

create table top\_states

(state nvarchar(255),

topstates float

)

insert into top\_states

select state, round(AVG(Literacy),0) as 'Avg\_literacy' from Data1

group by State

order by Avg\_literacy desc;

select top 3\* from top\_states

--bottom 3

drop table bottom\_states

create table bottom\_states

(state nvarchar(255),

bottomstates float

)

insert into bottom\_states

select state, round(AVG(Literacy),0) as 'Avg\_literacy' from Data1

group by State

order by Avg\_literacy asc;

select top 3\* from bottom\_states

--combining the top and bottom tables

select \* from(

select top 3\* from top\_states

) a

union

select \* from(

select top 3\* from bottom\_states

) b

--10. names of states starting with a or d

select distinct state from Data1 where lower(State) like 'a%' or lower(State) like 'd%'

select distinct state from Data1 where lower(State) like '\_\_a%'

select distinct district from Data2 where lower(District) like '%est%' LIKE '[bsp]%'

select distinct district from Data2 where lower(District) LIKE '[bs]%'

select distinct district from Data2 where lower(District) like '[d-f]%'

select distinct district from Data2 where lower(District) like 'b%e' -- starts with b and ends with e

select distinct district from Data2 where lower(District) like '[aeiuo]%' -- starts with a vowel

select distinct district from Data2 where left(lower(District),1) in ('a','e','i','o','u')

and right(lower(District),1) not in ('a','e','i','o','u')

--starts with vowels and does not end with vowels...usimg substring

--PART 2

--1. Joining tables and male female on district level

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 Data1 a inner join Data2 b on a.District=b.district) c

--2. Joining tables and male female on state level

select d.state,sum(d.males) Males, sum(d.females) 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 Data1 a inner join Data2 b on a.District=b.district) c)d

group by d.state

--3. Calculating total litracy rate

select d.state, sum(d.literate\_people) tot\_lit\_pop, sum(d.illiterate\_people) tot\_illit\_pop from

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

round((1-c.literacy\_ratio)\*c.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)c)d

group by State

-- 4. population in previous census and sum

select sum(m.prev\_census\_popln) as prev,sum(current\_census\_popln) as curr from

(select d.state, sum(d.prev\_census\_popln) prev\_census\_popln ,

sum(d.current\_census\_popln) current\_census\_popln from

(select c.district, c.state, c.growth, round(c.population/(1+c.growth),0) prev\_census\_popln,

c.population current\_census\_popln from

(select a.district, a.state, a.growth growth, b.population

from Data1 a inner join Data2 b on a.District=b.district)c)d

group by State)m

--5. Population vs Area

select (g.total\_area/g.prev) as prev\_cen\_pop, (g.total\_area/g.curr) as curr\_cen\_pop from

(select q.\*,r.total\_area from(

select '1' as keyy, n.\* from

(select sum(m.prev\_census\_popln) as prev,sum(current\_census\_popln) as curr from

(select d.state, sum(d.prev\_census\_popln) prev\_census\_popln ,

sum(d.current\_census\_popln) current\_census\_popln from

(select c.district, c.state, c.growth, round(c.population/(1+c.growth),0) prev\_census\_popln,

c.population current\_census\_popln from

(select a.district, a.state, a.growth growth, b.population

from Data1 a inner join Data2 b on a.District=b.district)c)d

group by State)m)n)q inner join (

select '1' as keyy, z.\* from(

select SUM(area\_km2) as total\_area from Data2)z)r on q.keyy=r.keyy) g

--6. Top 3 districts from each state which has the highest literacy rate (using window functions)

select a.\* from

(select district, state, literacy, RANK() over(partition by state order by literacy desc) rnk from data1) a

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