**SQL\_Query\_File**

CREATE DATABASE Census\_2011

USE Census\_2011

SELECT \* FROM [dbo].[Data1$]

--As there were duplicates, removing duplicates here:

SELECT District,State,Growth,Sex\_ratio,Literacy,COUNT(\*) as count1 into New\_data\_1 from [dbo].[Data1$]

Group BY District,State,Growth,Sex\_ratio,Literacy

-- Now, two databases tables created as Data1 and Data2

SELECT \* FROM Data1;

SELECT \* FROM Data2;

-- Number of rows in the dataset

SELECT COUNT(\*) as count\_of\_data1 FROM Data1;

SELECT COUNT(\*) as count\_of\_data2 FROM Data2;

--Data from Jharkhand and Bihar

SELECT \* FROM Data1

WHERE State = 'Jharkhand' or State ='Bihar';

--Population of India

SELECT SUM(population) as Population FROM Data2

--Average Growth Rate

SELECT AVG(Growth)\*100 as Avg\_gwt\_rate FROM Data1

--Average Growth percentage - statewise

SELECT State,AVG(Growth)\*100 as Avg\_gwt\_rate FROM Data1

GROUP BY State

-- Average Sex ratio

SELECT State,Round(AVG(Sex\_ratio),0) as Avg\_sex\_rate FROM Data1

GROUP BY State

Order By Avg\_sex\_rate desc;

--Average Literacy rate

SELECT State,Round(AVG(Literacy),0) as Avg\_literacy\_rate FROM Data1

GROUP BY State

Having Round(AVG(Literacy),0) > 90

Order By Avg\_literacy\_rate desc;

--Top 3 states with highest average growth state

SELECT TOP 3 State,AVG(Growth)\*100 as Avg\_gwt\_rate FROM Data1

GROUP BY State

Order BY Avg\_gwt\_rate desc;

--Bottom 3 states with highest average growth state

SELECT TOP 3 State,Round(AVG(Sex\_ratio),0) as Avg\_sex\_rate FROM Data1

GROUP BY State

Order BY Avg\_sex\_rate asc;

--Top and bottom 3 states in literacy rate

create table #topstates

( state nvarchar(225),

topstates float,

)

insert into #topstates

SELECT TOP 3 State,Round(AVG(Literacy),0) as Avg\_literacy\_rate FROM Data1

GROUP BY State

Order BY Avg\_literacy\_rate desc

SELECT \* FROM #topstates

create table #bottomstates

( state nvarchar(225),

bottomstates float,

)

insert into #bottomstates

SELECT TOP 3 State,Round(AVG(Literacy),0) as Avg\_literacy\_rate FROM Data1

GROUP BY State

Order BY Avg\_literacy\_rate asc

SELECT \* FROM #bottomstates

-- Using union operator to combine both tables of top and bottom

SELECT \* FROM #topstates

UNION ALL

SELECT \* FROM #bottomstates

--States starting with letter "a"

SELECT distinct State FROM Data1

WHERE lower(State) like 'a%'

--States starting with both 'a' or 'b'

SELECT distinct State FROM Data1

WHERE lower(State) like 'a%' or lower(State) like 'b%'

--States starting with both 'a' or ending with 'd'

SELECT distinct State FROM Data1

WHERE lower(State) like 'a%' or lower(State) like '%d'

--States starting with both 'a' and ending with 'm'

SELECT distinct State FROM Data1

WHERE lower(State) like 'a%' and lower(State) like '%m'

-- Joining both the Data tables 1 and 2 to find number of Males and females

SELECT a.District,a.State,a.Sex\_ratio,b.Population FROM Data1 as a

LEFT JOIN Data2 as b on a.District = b.District

--Formulae:

--Sex\_ratio = number of females / numbers of males \*100 ... eq.1

--females + males = population

--therefore, females = population - males

--and males = population - females

--now, using eq.1 - (population - males) = sex\_ratio \* males

--population = males(sex\_ratio + 1)

--so, males = population / (sex\_ratio + 1)

--females = population - (population/(sex\_ratio + 1))

--Deriving the males and females using sub-query

--Districtwise data

SELECT district,state,Round((a.Population/(a.sex\_ratio+1)),0) as no\_of\_males,Round((a.Population - (a.Population/(a.sex\_ratio + 1))),0)as no\_of\_females from

(SELECT a.District,a.State,a.Sex\_ratio/1000 as sex\_ratio,b.Population FROM Data1 as a

LEFT JOIN Data2 as b on a.District = b.District)a

--Statewise data

SELECT State,sum(no\_of\_males) as males,sum(no\_of\_females) as females from

(SELECT district,state,Round((a.Population/(a.sex\_ratio+1)),0) as no\_of\_males,Round((a.Population - (a.Population/(a.sex\_ratio + 1))),0)as no\_of\_females from

(SELECT a.District,a.State,a.Sex\_ratio/1000 as sex\_ratio,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District)a)c

Group By State;

--total literacy rate

SELECT a.District,a.State,a.Literacy as literacy\_ratio,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District

--Formulae:

--total literate people/population = literacy\_ratio

--District-wise literate and illiterate people

SELECT District,State,literacy\_ratio,Population,total\_literate\_people,(Population - total\_literate\_people) as total\_illiterate\_people FROM

(SELECT District,State,literacy\_ratio,Population,Round((Population\*literacy\_ratio/100),0) as total\_literate\_people FROM

(SELECT a.District,a.State,a.Literacy as literacy\_ratio,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District)a)c

--State-wise literate and illiterate people

SELECT State,sum(total\_literate\_people) as total\_literate\_people,sum(total\_illiterate\_people) as total\_illiterate\_people FROM

(SELECT District,State,literacy\_ratio,Population,total\_literate\_people,(Population - total\_literate\_people) as total\_illiterate\_people FROM

(SELECT District,State,literacy\_ratio,Population,Round((Population\*literacy\_ratio/100),0) as total\_literate\_people FROM

(SELECT a.District,a.State,a.Literacy as literacy\_ratio,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District)a)c)d

GROUP BY State;

-- Finding the population in the previous Census

SELECT a.District,a.State,a.Growth as Growth\_rate,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District

--Formulae:

--previous\_census + growth\*previous\_census = population

--previous\_census = population / (1+growth)

--District-wise

SELECT District,State,Round((Population/(1+Growth\_rate)),0) as previous\_census,Population FROM

(SELECT a.District,a.State,a.Growth as Growth\_rate,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District)a

--State-wise

SELECT State, sum(previous\_census) as previous\_census,sum(population) as current\_census FROM

(SELECT District,State,Round((Population/(1+Growth\_rate)),0) as previous\_census,Population FROM

(SELECT a.District,a.State,a.Growth as Growth\_rate,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District)a)b

GROUP BY STATE

--Total previous census population and current census population

SELECT SUM(previous\_census) as previous\_census\_population , SUM(current\_census) as current\_census\_population FROM

(SELECT State, sum(previous\_census) as previous\_census,sum(population) as current\_census FROM

(SELECT District,State,Round((Population/(1+Growth\_rate)),0) as previous\_census,Population FROM

(SELECT a.District,a.State,a.Growth as Growth\_rate,b.Population FROM Data1 as a

INNER JOIN Data2 as b on a.District = b.District)a)b

GROUP BY STATE)f

--TOP 3 districts with each state with highest literacy rate

SELECT \* FROM

(SELECT District,State,Literacy,rank() over(partition by State Order by literacy desc) as rank1 FROM Data1)a

WHERE rank1 in (1,2,3)

ORDER BY State