**PORTOFOLIO PROJECT ON INDIAN CENSUS DATA**

**LET’S FIRST ESTABLISH CONNECTION:**

* SHOW GLOBAL VARIABLES LIKE 'LOCAL\_INFILE';
* SET GLOBAL LOCAL\_INFILE=TRUE;

C:\PROGRAM FILES\MYSQL\MYSQL SERVER 8.0\BIN>MYSQL --LOCAL\_INFILE=1 -U ROOT -PAKSHAY@1998 ARYAN

**CREATING TABLE IN DATABASE:**

CREATE TABLE DATAONE(DISTRICT NVARCHAR(255),STATE NVARCHAR(255),GROWTH FLOAT,SEX\_RATIO FLOAT,LITERACY FLOAT);

CREATE TABLE DATA2(DISTRICT NVARCHAR(255),STATE NVARCHAR(255),AREA\_KM2 FLOAT,POPULATION FLOAT);

**LOADING EXCEL DATA INTO TABLE FIELDS:**

load data local infile 'C:\\Users\\AKSHAY AKHADE\\Desktop\\new challenges\\DTWO.CSV' into table DATA2 FIELDS TERMINATED BY “,” LINES TERMINATED BY “\n” IGNORE 1 LINES(DISTRICT,STATE,AREA\_KM2,POPULATION);

load data local infile 'C:\\Users\\AKSHAY AKHADE\\Desktop\\new challenges\\DONE.CSV' into table DATAONE FIELDS TERMINATED BY “,” LINES TERMINATED BY “\n” IGNORE 1 LINES(DISTRICT,STATE,GROWTH,SEX\_RATIO,LITERACY);

**SQL COMMANDS TO PERFORM ANALYSIS:**

# WHAT IS GROWTH, LITERACY,SEX\_RATIO FOR STATE GOA AND PUNJAB

SELECT STATE,GROWTH,LITERACY,SEX\_RATIO FROM DATA1 WHERE STATE IN('GOA','PUNJAB') ORDER BY STATE;

# DISPLAY ALL RECORDS OF TABLE DATA1 & TABLE DATA2

SELECT \* FROM DATA1 LIMIT ; AND

SELECT \* FROM DATA2;

# SHOW TOTAL POPULATION

SELECT SUM(POPULATION) AS TOTAL\_POPULATION FROM DATA2;

# FIND OUT AVG GROWTH IN % BY STATE

SELECT STATE,ROUND(AVG(GROWTH\*100),0) AS AVERAGE\_GROWTH\_IN\_PERCENTAGE FROM DATA1 GROUP BY STATE ORDER BY AVERAGE\_GROWTH\_IN\_PERCENTAGE;

#AVG SEX\_RATIO BY STATE

SELECT STATE,ROUND(AVG(SEX\_RATIO),0) AS AVERAGE\_SEX\_RATIO FROM DATA1 GROUP BY STATE ORDER BY AVERAGE\_SEX\_RATIO;

# AVG LITERCY BY STATE

SELECT STATE,ROUND(AVG(LITERACY),0) AS AVERAGE\_LITERACY FROM DATA1 GROUP BY STATE ORDER BY AVERAGE\_LITERACY ;

# TOP 5 STATE WITH HIGHEST AVERAGE LITERACY

SELECT STATE,ROUND(AVG(LITERACY),0) AS AVG\_LITERACY FROM DATA1 GROUP BY STATE HAVING AVG\_LITERACY>80 ORDER BY AVG\_LITERACY DESC LIMIT 5;

# WHICH ARE LOWEST 5 STATE WITH BOTTOM 5 AVERAGE GROWTH?

SELECT STATE,ROUND(AVG(GROWTH)\*100,0) AS AVERAGE\_GROWTH\_IN\_PERCETAGE FROM DATA1 GROUP BY STATE ORDER BY AVERAGE\_GROWTH\_IN\_PERCENTAGE DESC LIMIT 5;

# WHICH ARE TOP 5 STATE WITH HIGHEST AVERAGE SEX\_RATIO?

SELECT STATE,ROUND(AVG(SEX\_RATIO)\*100,0) AS AVERAGE\_SEX\_RATIO FROM DATA1 GROUP BY STATE ORDER BY AVERAGE\_SEX\_RATIO DESC LIMIT 5;

# STATE WITH STARTING LETTER ‘A’

SELECT \* FROM DATA1 WHERE state LIKE 'a%' group by state order by state;

# STATE WITH STARTING LETTER ‘M’ AND LAST LETTER ‘A’

SELECT \* FROM DATA1 WHERE STATE LIKE 'M%' and state like '%a' GROUP BY STATE ORDER BY STATE;

# WHICH ARE TOP 3 AND BOTTOM 3 STATE WITH AVERAGE LITERACY?

CREATE TABLE BOTTOMUP(STATE VARCHAR(255),LITERACY FLOAT);

CREATE TABLE TOPUP(STATE VARCHAR(255),LITERACY FLOAT);

INSERT INTO BOTTOMUP(STATE,LITERACY)(SELECT STATE,ROUND(AVG(LITERACY),0) AS AVG\_LITERACY FROM DATA1 GROUP BY STATE ORDER BY AVG\_LITERACY);

INSERT INTO TOPUP(STATE,LITERACY)(SELECT STATE,ROUND(AVG(LITERACY),0) AS AVG\_LITERACY FROM DATA1 GROUP BY STATE ORDER BY AVG\_LITERACY DESC);

SELECT \* FROM (SELECT \* FROM BOTTOMUP GROUP BY STATE ORDER BY LITERACY ASC LIMIT 3;)A UNION

SELECT \* FROM (SELECT \* FROM BOTTOMUP GROUP BY STATE ORDER BY LITERACY DESC LIMIT 3;)B

# FIND OUT DISTRICT WISE NUMBER OF MALES AND FEMALES

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 DATA1.DISTRICT,DATA1.STATE,((DATA1.SEX\_RATIO)/1000) SEX\_RATIO,DATA2.POPULATION FROM DATA1 INNER JOIN DATA2 ON DATA1.STATE=DATA2.STATE)C;

# FIND STATE WISE NUMBER OF MALES AND FEMALES

SELECT D.STATE,D.MALES,D.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 DATA1.DISTRICT,DATA1.STATE,((DATA1.SEX\_RATIO)/1000) SEX\_RATIO,DATA2.POPULATION FROM DATA1 INNER JOIN DATA2 ON DATA1.STATE=DATA2.STATE)C)D GROUP BY D.STATE LIMIT;

# FIND OUT TOTAL LITERATE AND ILITARATE PEPOPLE BY STATE

SELECT C.STATE,SUM(ROUND((C.LITERACY\_RATIO\*C.POPULATION),0))TOTAL\_LITERATE\_PEOPLE,SUM(ROUND((1-C.LITERACY\_RATIO)\*(C.POPULATION) ,0))TOTAL\_ILITERATE\_PEOPLE,C.POPULATION FROM (SELECT A.DISTRICT,A.STATE,ROUND((A.LITERACY/100),4) LITERACY\_RATIO,B.POPULATION FROM DATA1 A INNER JOIN DATA2 B ON A.STATE=B.STATE)C GROUP BY STATE;

# WHAT IS THE PREVIOUS CENSUS POPULATION?

SELECT SUM(E. CURRENT\_CENSUS) TOTAL\_CURRENT\_CENSUS,SUM(E. PREVIOUS\_CENSUS) TOTAL\_PREVIOUS\_CENSUS FROM (SELECT D.STATE,SUM(D. CURRENT\_CENSUS) CURRENT\_CENSUS,SUM(D. PREVIOUS\_CENSUS) PREVIOUS\_CENSUS FROM (SELECT C.STATE,C.POPULATION CURRENT\_CENSUS,ROUND((C.POPULATION/(1+C.GROWTH)),0) PREVIOUS\_CENSUS FROM (SELECT A.DISTRICT,A.STATE,A.GROWTH,B.POPULATION FROM DATA1 A INNER JOIN DATA2 B ON A.DISTRICT=B.DISTRICT)C)D GROUP BY D.STATE)E;

# FIND OUT TOAL\_AREA VS CURRENT\_CENSUS\_POPULATION AS WELL AS TOTAL\_AREA VS PREVIOUSE\_CENSUS\_POPULATION

SELECT Z.TOTAL\_AREA/ Z.TOTAL\_CURRENT\_CENSUS TOTAL\_CURRENT\_CENSUS\_VS\_AREA, Z.TOTAL\_AREA/ Z.TOTAL\_PREVIOUS\_CENSUS TOTAL\_PREVIOUS\_CENSUS\_VS\_AREA FROM (SELECT X.\*, Y.TOTAL\_AREA FROM (SELECT '1' AS KEYY,R.\* FROM(SELECT SUM(E. CURRENT\_CENSUS) TOTAL\_CURRENT\_CENSUS,SUM(E. PREVIOUS\_CENSUS) TOTAL\_PREVIOUS\_CENSUS FROM (SELECT D.STATE,SUM(D. CURRENT\_CENSUS) CURRENT\_CENSUS,SUM(D. PREVIOUS\_CENSUS) PREVIOUS\_CENSUS FROM (SELECT C.STATE,C.POPULATION CURRENT\_CENSUS,ROUND((C.POPULATION/(1+C.GROWTH)),0) PREVIOUS\_CENSUS FROM (SELECT A.DISTRICT,A.STATE,A.GROWTH,B.POPULATION FROM DATA1 A INNER JOIN DATA2 B ON A.DISTRICT=B.DISTRICT)C)D GROUP BY D.STATE)E)R )X INNER JOIN (SELECT '1' KEYY,N.\* FROM(SELECT SUM(AREA\_KM2)TOTAL\_AREA FROM DATA2)N)Y ON X.KEYY=Y.KEYY)Z;

# TOP 3 DISTRICT FROM EACH STATE WITH HIGHEST LITERACY

SELECT A.\* FROM(SELECT DISTRICT,STAT E,LITERACY, RANK() OVER (PARTITION BY STATE ORDER BY LITERACY)RNK FROM DATA1)A WHERE RNK IN (1,2,3) ORDER BY A.STATE;