SQL ANALYSIS

USING MYSQLWORKBENCH

 https://github.com/sohailkundgol2004-svg/UIDAI-DATA-HACKTHON-ANALYSIS

Cleaning and Formatting Data Process :

SHOW variables LIKE "secure_file_priv";

select count(*) from aadharenrol;

-- loading and creating table and data

CREATE TABLE Adharenrolment (

   date varchar(100),

   state VARCHAR(100),

   district VARCHAR(100),

   pincode INT,

   age_0_5 INT,

   age_5_17 INT,

   age_18_greater INT,

   Total_No_of_people INT

);

LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/aadhaar_clean_master_final.csv"

INTO TABLE adharenrolment

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

IGNORE 1 lines;

CREATE TABLE Adharbio (

   date varchar(100),

   state VARCHAR(100),

   district VARCHAR(100),

   pincode INT,

   age_5_17 INT,

```sql
    age_18_greater INT

);




LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/biometric_clean_final.csv"

INTO TABLE adharbio

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

IGNORE 1 lines;




CREATE TABLE adhardemogry (

    date varchar(100),

    state VARCHAR(100),

    district VARCHAR(100),

    pincode INT,

    age_5_17 INT,

    age_18_greater INT

);




LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/aadhaarDeomographicCombinedcleaned.csv"

INTO TABLE adhardemogry

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

IGNORE 1 lines;




ALTER TABLE adharenrolment

ADD COLUMN region VARCHAR(20);




SET SQL_SAFE_UPDATES = 0;

UPDATE adharenrolment

SET region = CASE

    WHEN state IN ('Delhi', 'Haryana', 'Himachal Pradesh', 'Jammu And Kashmir', 'Ladakh', 'Punjab', 'Rajasthan', 'Chandigarh', 'Uttarakhand') THEN 'North'

    WHEN state IN ('Andhra Pradesh', 'Karnataka', 'Kerala', 'Tamil Nadu', 'Telangana', 'Pondicherry', 'Lakshadweep', 'Andaman And Nicobar Islands', 'Andaman & Nicobar Islands') THEN 'South'

    WHEN state IN ('Bihar', 'Jharkhand', 'Odisha', 'West Bengal') THEN 'East'
```

WHEN state IN ('Goa', 'Gujarat', 'Maharashtra', 'Daman And Diu', 'Dadra And Nagar Haveli', 'Dadra And Nagar Haveli And Daman And Diu', 'The Dadra And Nagar Haveli And Daman And Diu') THEN 'West'

   WHEN state IN ('Madhya Pradesh', 'Chhattisgarh', 'Uttar Pradesh') THEN 'Central'

   WHEN state IN ('Arunachal Pradesh', 'Assam', 'Manipur', 'Meghalaya', 'Mizoram', 'Nagaland', 'Sikkim', 'Tripura') THEN 'North-East'

   ELSE 'Unknown'

END;

SET SQL_SAFE_UPDATES = 1;

/* ANALYSIS */

-- Region and Enrolment Analysis.

-- region and there enrolments

 select region, sum(Total_No_of_people) as TotalEnrollment

from adharenrolment

group by region order by TotalEnrollment desc;

	region	TotalEnrollment
▶	West	653820
	Unknown	218
	South	782017
	North-East	392773
	North	725605
	East	1265451
	Central	1615818

. Key Findings & Insights

Based on the queries executed (Results 14 through 20):

Highest Enrollment Volume: The Central region shows the highest volume of total enrollments with 1,615,818 records, followed by the East region.

Regional Trends: There is a significant gap between the Central region and others like the North-East (392,773), suggesting different levels of saturation or demographic density.

Demographic Ratios: The analysis included a specific calculation for CHILD_RATIO_PERCENTAGE. This metric is crucial for government policy, as it helps identify where child enrollment drives (ages 5–17) need to be prioritized.

Data Integrity: A small number of records (218) were flagged as "Unknown" region, which have been filtered to ensure the accuracy of the regional insights.

-- Top 3 region by new aadhar children enrolments.

```
 select Region, sum(age_0_5) as Newenrolment

from adharenrolment

group by Region order by Newenrolment desc limit 3;
```

Region	Newenrolment
Central	971488
East	736277
South	637124

-- Top state in East Region

```
select state, sum(Total_No_of_people) as Total_enrol_state

from adharenrolment

where region='East' group by state order by Total_enrol_state desc;
```

state	Total_enrol_state
Bihar	609585
West Bengal	375340
Jharkhand	157539
Odisha	122987

-- Top 3 least Enrolment states

```
select state, sum(Total_No_of_people) as Least_enrol_states

 from adharenrolment

group by state order by Least_enrol_states limit 9, 3;
```

state	Least_enrol_states
Sikkim	2207
Goa	2333
Chandigarh	2723

SELECT

```
state,

district,

SUM(age_5_17) AS children,

SUM(age_18_greater) AS adults,

ROUND(

    SUM(age_5_17) * 1.0 /

    NULLIF(SUM(age_18_greater), 0), 2

) AS child_to_adult_ratio

FROM adhardemogry

GROUP BY state, district

ORDER BY child_to_adult_ratio DESC;
```

state	children	adults	child_to_adult_ratio
Dadra And Nagar Haveli	1342	4926	0.27
Arunachal Pradesh	5783	30660	0.19
Karnataka	264981	1430304	0.19
Chandigarh	13133	70228	0.19
Lakshadweep	170	1006	0.17
Telangana	242259	1387649	0.17
Tamil Nadu	315638	1896590	0.17
Pondicherry	4696	28067	0.17
Jammu And Kashmir	57873	349329	0.17
Andhra Pradesh	321148	1974434	0.16
Manipur	41464	260085	0.16
Madhya Pradesh	407098	2505840	0.16
Odisha	147499	964566	0.15
Himachal Pradesh	19424	129790	0.15
Mizoram	5302	36604	0.14
Daman And Diu	275	1921	0.14
Delhi	175535	1263399	0.14
Haryana	139314	1026826	0.14
Nagaland	4314	32477	0.13

```
-- Societal Trend Analysis AND RATIO ANALYSIS.

-- State District wise children aged 5-17, and Adult aged 18-greater;

select region, sum(age_5_17) as CHILD_ENROLMENTS,

sum(age_18_greater) as ADULT_ENROLMENTS,

ROUND((sum(age_5_17) / sum(age_5_17 + age_18_greater)) * 100, 2)

as CHILD_RATIO_PERCENTAGE from adharenrolment group by region order by CHILD_RATIO_PERCENTAGE
desc;
```

region	CHILD_ENROLMENTS	ADULT_ENROLMENTS	CHILD_RATIO_PERCENTAGE
East	506568	22606	95.73
Central	614832	29498	95.42
North	171385	13991	92.45
South	128168	16725	88.46
West	153802	24760	86.13
North-East	145628	60556	70.63
Unknown	1	217	0.46

region	CHILD_ENROLMENTS	ADULT_ENROLMENTS	CHILD_RATIO_PERCENTAGE
East	506568	22606	95.73
Central	614832	29498	95.42
North	171385	13991	92.45
South	128168	16725	88.46
West	153802	24760	86.13
North-East	145628	60556	70.63