

The graphic features a light gray background with a pattern of fine, white, curved lines that create a sense of motion. In the top-left and bottom-right corners, there are thick, sweeping curves in dark blue and gold. The text is centered and reads:

INDIAN CENSUS 2011

INSIGHTS

PROJECT GOALS

The goal of performing SQL analysis on the Indian Census data is to derive meaningful insights into demographic trends, such as population growth, literacy rates, and sex ratios across different states and districts. By organizing and querying the data, we can identify disparities, make comparisons, and inform policy decisions. Additionally, using joins and aggregate functions allows for a comprehensive understanding of relationships within the data, ultimately supporting socio-economic planning and development initiatives.

Dataset: https://www.census2011.co.in/district.php#google_vignette

How Data Looks?

Select * from dataset1;

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

	District	State	Growth	Sex_Ratio	Literacy
▶	South Andaman	Andaman And Nicobar Islands	0.14	871	89.13
	North And Middle Andaman	Andaman And Nicobar Islands	0	925	83.91
	Nicobars	Andaman And Nicobar Islands	-0.12	777	78.06
	Rangareddy	Andhra Pradesh	0.48	961	75.87
	East Godavari	Andhra Pradesh	0.05	1006	70.99
	Guntur	Andhra Pradesh	0.09	1003	67.4
	Krishna	Andhra Pradesh	0.08	992	73.74

Select * from dataset2;

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
District	State	Area_km2	Population
Adilabad	Andhra Pradesh	16105	2741239
Agra	Uttar Pradesh	4041	4418797
Ahmadabad	Gujarat	8107	7214225
Ahmadnagar	Maharashtra	17048	4543159
Aizawl	Mizoram	3576	400309
Ajmer	Rajasthan	8481	2583052
Akola	Maharashtra	5676	1813906

Number of rows in our dataset?

Select count(*) as "Number of rows" from dataset1;




	Number of rows
▶	640

Select count(*) as "Number of rows" from dataset2;

	Number of rows
▶	640

Dataset for jharkhand and bihar?

```
SELECT
    dataset1.District,
    dataset1.state,
    dataset1.Growth,
    dataset1.sex_ratio,
    dataset1.literacy,
    dataset2.area_km2,
    dataset2.population
FROM
    dataset1
INNER JOIN
    dataset2
ON
    dataset1.District = dataset2.District AND
    dataset1.state = dataset2.state
WHERE
    dataset1.state IN ('Jharkhand', 'Bihar');
```

Result Grid  Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 							
	District	state	Growth	sex_ratio	literacy	area_km2	population
▶	Araria	Bihar	0.3	921	53.53	2830	2811569
	Arwal	Bihar	0.19	928	67.43	638	700843
	Banka	Bihar	0.26	907	58.17	3020	2034763
	Begusarai	Bihar	0.26	895	63.87	1918	2970541
	Bhagalpur	Bihar	0.25	880	63.14	2569	3037766
	Bhojpur	Bihar	0.22	907	70.47	2395	2728407
	Bokaro	Jharkhand	0.16	922	72.01	2883	2062330





Population in India?

```
SELECT SUM(population) AS Population_of_India  
FROM dataset2;
```

Result Grid			Filter Rows: <input type="text"/>	Export: 	Wrap Cell Content: 
	Population_of_India				
	1210854977				

Average Growth per State?

```
SELECT
    state,
    ROUND(AVG(growth), 2) * 100 AS
average_growth
FROM
    dataset1
GROUP BY
    state;
```

Result Grid   Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 		
	state	average_growth
▶	Andaman And Nicobar Islands	1
	Andhra Pradesh	11
	Arunachal Pradesh	28.000000000000004
	Assam	16
	Bihar	25
	Chandigarh	17
	Chhattisgarh	20

Average Sex Ratio per State?

```
SELECT
  state,
  ROUND(AVG(sex_ratio), 0) AS
average_sex_ratio
FROM
  dataset1
GROUP BY
  state
ORDER BY
  average_sex_ratio DESC;
```

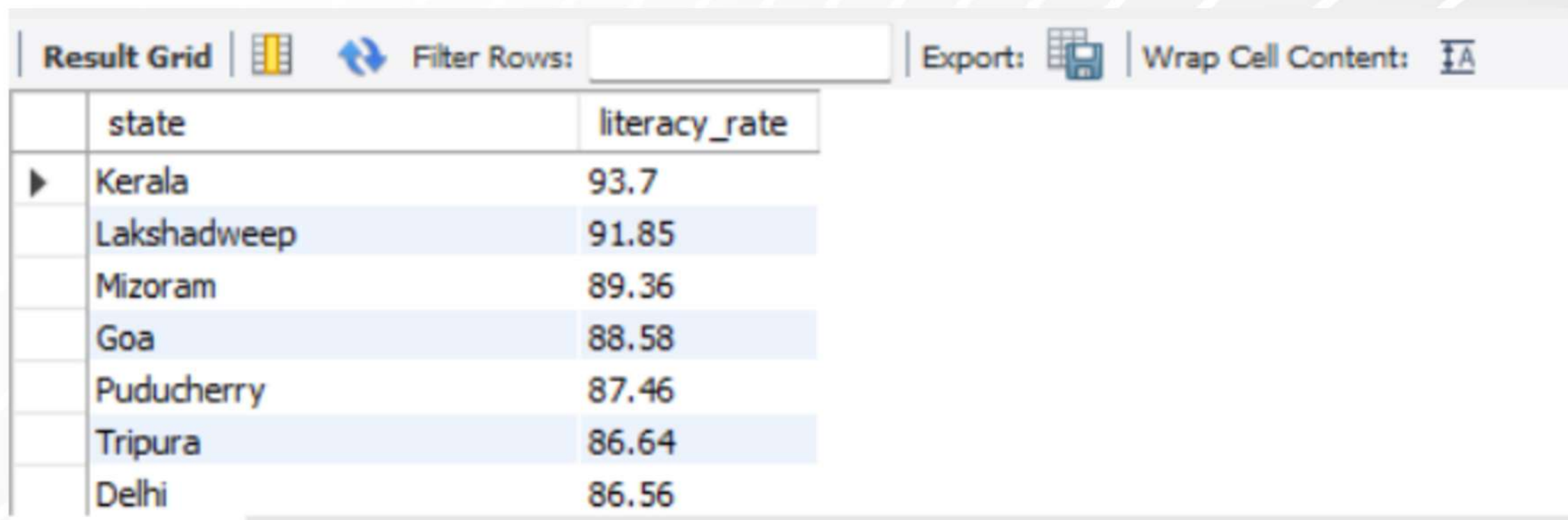


The screenshot shows a database query result grid. At the top, there are controls: 'Result Grid' with a table icon, 'Filter Rows:' with a search box, 'Export:' with a download icon, and 'Wrap Cell Content:' with a text wrap icon. The table below has two columns: 'state' and 'average_sex_ratio'. The data is sorted in descending order of the average sex ratio. The states listed are Kerala (1080), Puducherry (1075), Uttarakhand (1010), Tamil Nadu (999), Andhra Pradesh (995), Chhattisgarh (995), and Karnataka (984). Each row is highlighted with a light blue background.

state	average_sex_ratio
Kerala	1080
Puducherry	1075
Uttarakhand	1010
Tamil Nadu	999
Andhra Pradesh	995
Chhattisgarh	995
Karnataka	984

Average Literacy rate per State?

```
SELECT
  state,
  ROUND(AVG(literacy), 2) AS literacy_rate
FROM
  dataset1
GROUP BY
  state
ORDER BY
  literacy_rate DESC;
```



The screenshot shows a database query result grid with a toolbar at the top. The toolbar includes a 'Result Grid' tab, a grid icon, a refresh icon, a 'Filter Rows:' input field, an 'Export:' button with a grid icon, and a 'Wrap Cell Content:' button with a text icon. The table below displays the results of the query, ordered by literacy rate in descending order. The columns are 'state' and 'literacy_rate'.

	state	literacy_rate
▶	Kerala	93.7
	Lakshadweep	91.85
	Mizoram	89.36
	Goa	88.58
	Puducherry	87.46
	Tripura	86.64
	Delhi	86.56

Top and Bottom States in Literacy Rate

DELIMITER //

CREATE PROCEDURE prod()

BEGIN

SELECT * FROM (

SELECT state AS Bottomstates, ROUND(AVG(literacy), 2) AS literacy_rate

FROM dataset1

GROUP BY state

ORDER BY 2 ASC

LIMIT 3

) al

ORDER BY 2 DESC;

SELECT state AS Topstates, ROUND(AVG(literacy), 2) AS literacy_rate

FROM dataset1

GROUP BY state

ORDER BY 2 DESC

LIMIT 3;

END //

DELIMITER ;

-- Calling the procedure

CALL prod();

Result Grid	Filter Rows:
Topstates	literacy_rate
Kerala	93.7
Lakshadweep	91.85
Mizoram	89.36

Result Grid	Filter Rows:
Bottomstates	literacy_rate
Rajasthan	64.6
Arunachal Pradesh	63.86
Bihar	61.76

Name the States starting with 'a'

```
SELECT DISTINCT(state)
FROM dataset1
WHERE state LIKE 'a%';
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	state			
▶	Andaman And Nicobar Islands			
	Andhra Pradesh			
	Arunachal Pradesh			
	Assam			

Total males and Females per state

```
-- males=population/sex_ratio+1
```

```
SELECT
```

```
state,
```

```
SUM(males) AS "Total males",
```

```
SUM(females) AS "Total females"
```

```
FROM
```

```
(
```

```
SELECT
```

```
d1.state,
```

```
d1.district,
```

```
ROUND(d2.population / ((d1.sex_ratio / 1000) + 1), 0) AS "males",
```

```
ROUND((d2.population * (d1.sex_ratio / 1000)) / ((d1.sex_ratio / 1000) + 1), 0)
```

```
AS "females"
```

```
FROM
```

```
dataset2 d2
```

```
INNER JOIN
```

```
dataset1 d1
```

```
ON
```

```
d1.district = d2.district
```

```
AND d1.state = d2.state
```

```
) AS tl
```




```
GROUP BY
```

```
state;
```

Result Grid	Filter Rows:	Export:
state	Total males	Total females
Andaman And Nicobar Islands	148014	126970
Andhra Pradesh	42439617	42141160
Arunachal Pradesh	624752	582402
Assam	15939857	15265719
Bihar	50277809	46182199
Chandigarh	580556	474894





Literate people per state

```
-- literate_people = population * literacy_ratio
SELECT
  d1.state,
  SUM(ROUND(d2.population * (d1.literacy / 100), 0)) AS "Total literate people"
FROM
  dataset2 d2
INNER JOIN
  dataset1 d1
ON
  d1.district = d2.district
  AND d1.state = d2.state
GROUP BY
  d1.state;
```

Result Grid   Filter Rows: <input type="text"/>			Export: 	Wrap Cell Content: 
	state	Total literate people		
▶	Andaman And Nicobar Islands	241015		
	Andhra Pradesh	56671677		
	Arunachal Pradesh	761557		
	Assam	22484409		
	Bihar	59605350		
	Chandigarh	908215		





Population in previous Census

```
-- prev + growth * prev = population
SELECT
  d1.state,
  SUM(ROUND(d2.population / (d1.growth + 1), 0)) AS "Previous census population",
  SUM(d2.population) AS "Current census population"
FROM
  dataset2 d2
INNER JOIN
  dataset1 d1
ON
  d1.district = d2.district
  AND d1.state = d2.state
GROUP BY
  d1.state;
```

Result Grid   Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 			
	state	Previous census population	Current census population
▶	Andaman And Nicobar Islands	250762	274984
	Andhra Pradesh	76254233	84580777
	Arunachal Pradesh	974615	1207154
	Assam	26666850	31205576
	Bihar	77039398	96460008
	Chandigarh	902094	1055450

Area given to per person in each state

```
SELECT
  d1.state,
  ROUND((SUM(d2.area_km2) * 1000) / (SUM(ROUND(d2.population / (d1.growth +
1), 0))), 2) AS "Previous census area m2 per person",
  ROUND((SUM(d2.area_km2) * 1000) / (SUM(d2.population)), 2) AS "Current census
area m2 per person"
FROM
  dataset2 d2
INNER JOIN
  dataset1 d1
ON
  d1.district = d2.district
  AND d1.state = d2.state
GROUP BY
  d1.state;
```

Result Grid   Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 			
	state	Previous census area m2 per person	Current census area m2 per person
▶	Andaman And Nicobar Islands	18	16.41
	Andhra Pradesh	3.61	3.25
	Arunachal Pradesh	82.37	66.50
	Assam	2.94	2.51
	Bihar	1.13	0.90
	Chandigarh	0.13	0.11

Top 3 district with high literacy rate per State

```
SELECT
  state,
  district,
  literacy
FROM
  (
    SELECT
      state,
      district,
      literacy,
      DENSE_RANK() OVER (PARTITION BY state ORDER BY literacy DESC) AS
        literate
    FROM
      dataset1
  ) AS ll
WHERE
  literate <= 3;
```

Result Grid	Filter Rows:	Export:	Wr
state	district	literacy	
Andaman And Nicobar Islands	South Andaman	89.13	
Andaman And Nicobar Islands	North And Middle Andaman	83.91	
Andaman And Nicobar Islands	Nicobars	78.06	
Andhra Pradesh	Hyderabad	83.25	
Andhra Pradesh	Rangareddy	75.87	
Andhra Pradesh	West Godavari	74.63	
Arunachal Pradesh	Papumpare	79.95	

RESULTS

- 1. Total Population:** The analysis provides the total population of India and breaks it down by state, including the counts of males and females.
- 2. Growth and Literacy Rates:** It identifies the average growth rates and literacy rates for each state, highlighting the top three states with the highest growth and the lowest literacy rates.
- 3. Sex Ratio Insights:** The average sex ratios are calculated for each state, revealing which states have the highest and lowest ratios.
- 4. District-Level Analysis:** The top three districts in each state with the highest literacy rates are identified, offering insights into regional educational performance.
- 5. Historical Population Estimates:** The analysis estimates previous census populations based on current growth rates, providing context for demographic changes over time.