DWM Project

==========

Batch : B1

Submitted by: Khushi Mantri (04)
Netal Rathi(06)
Rashi Potey(07)
Shinosha Jain(09)

==========

Problem Statement:

Using AWS Cloud, perform basic queries to analyze and derive insights from datasets.

Introduction:

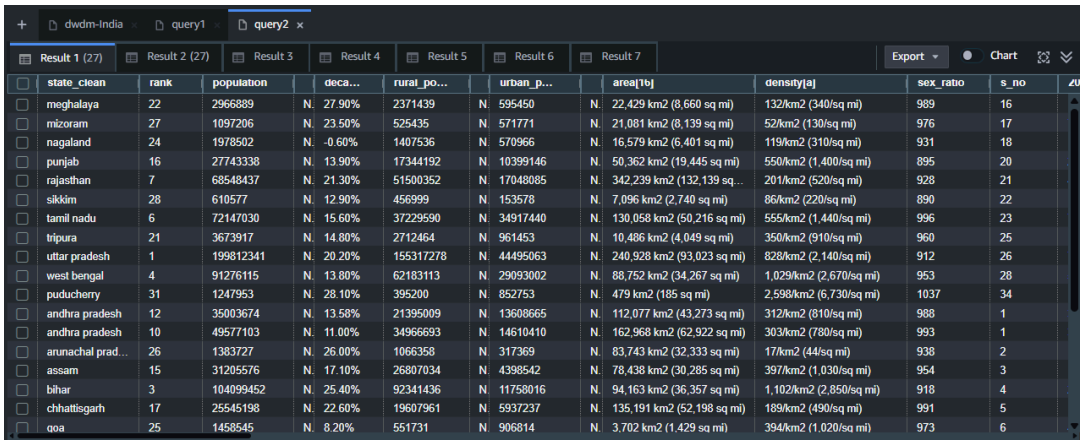
In this project, we leveraged AWS Cloud services to carry out a basic yet insightful economic and demographic analysis of Indian states using two datasets: the India Census and State-wise State Domestic Product (SDP). The primary objective was to identify and explore key patterns by analyzing indicators such as Per Capita Income and Sex Ratio. Utilizing AWS tools including Amazon S3 for storage, AWS Glue for ETL processing, and Amazon Redshift for querying, we executed multiple analyses to understand the economic conditions of different states, their demographic distributions, and the relationships between these variables.

SQL Script:

Query1:

```
SELECT * FROM spectrum_schema.output;
```

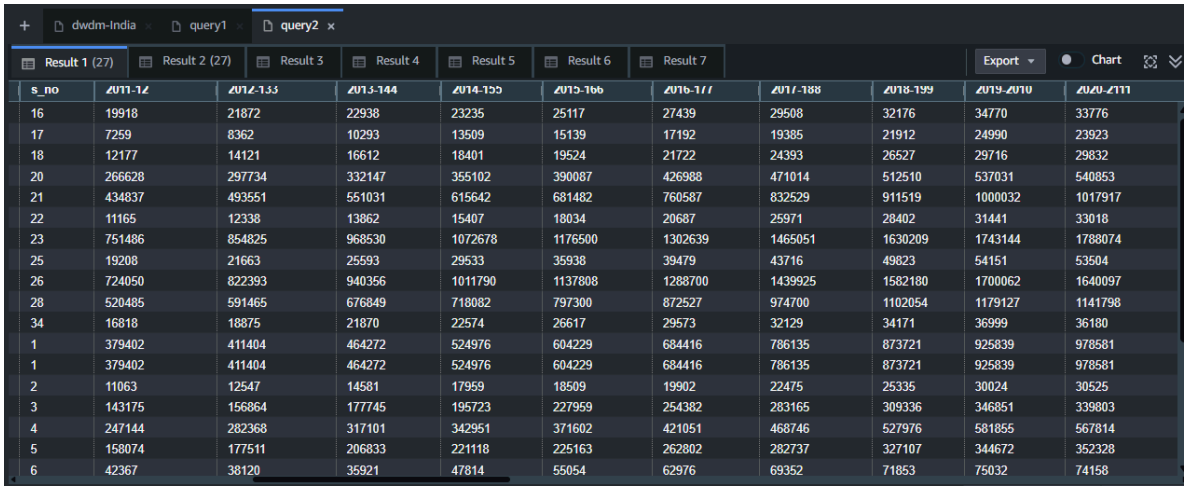
Result1:



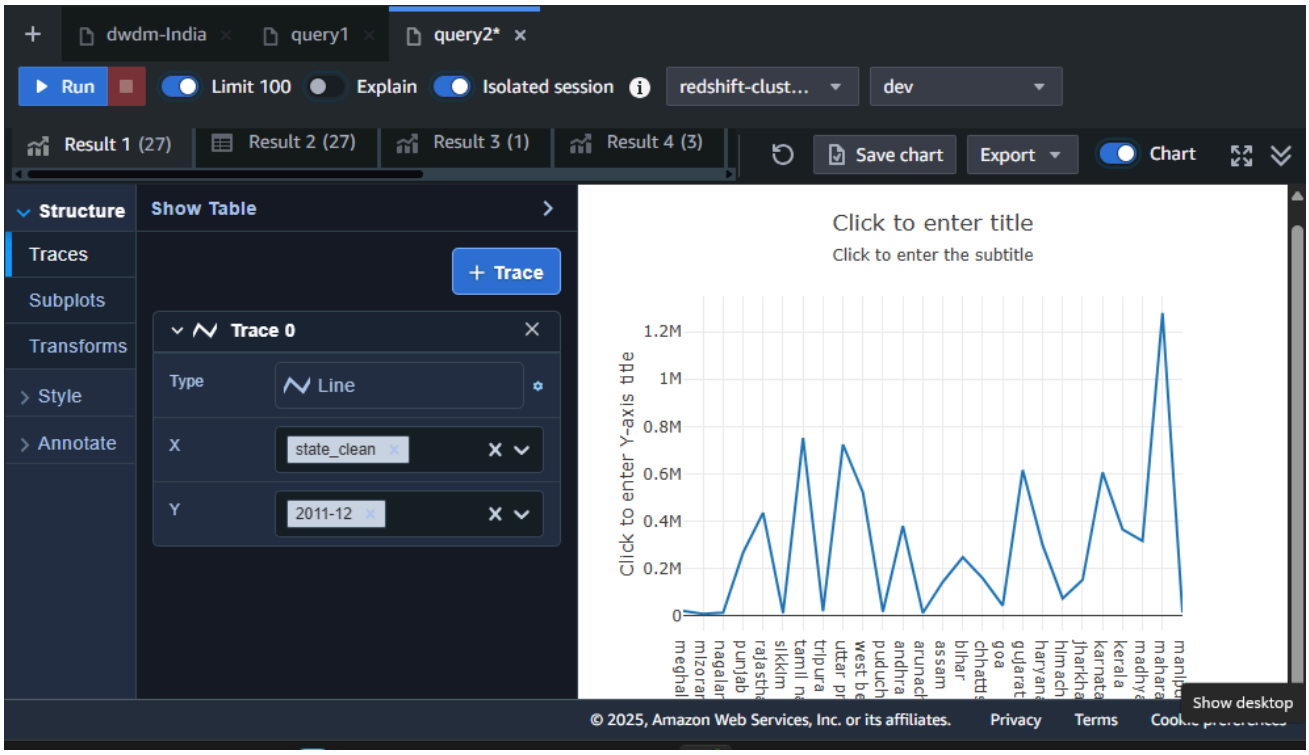
dwdm-India | query1 | query2

Result 1 (27) | Result 2 (27) | Result 3 | Result 4 | Result 5 | Result 6 | Result 7 | Export | Chart

state_clean	rank	population	deca...	rural_po...	urban_p...	area[16]	density[a]	sex_ratio	s_no
meghalaya	22	2966889	N. 27.90%	2371439	N. 595450	N. 22,429 km2 (8,660 sq mi)	132/km2 (340/sq mi)	989	16
mizoram	27	1097206	N. 23.50%	525435	N. 571771	N. 21,081 km2 (8,139 sq mi)	52/km2 (130/sq mi)	976	17
nagaland	24	1978502	N. -0.60%	1407536	N. 570966	N. 16,579 km2 (6,401 sq mi)	119/km2 (310/sq mi)	931	18
punjab	16	27743338	N. 13.90%	17344192	N. 10399146	N. 50,362 km2 (19,445 sq mi)	550/km2 (1,400/sq mi)	895	20
rajasthan	7	68548437	N. 21.30%	51500352	N. 17048085	N. 342,239 km2 (132,139 sq...	201/km2 (520/sq mi)	928	21
sikkim	28	610577	N. 12.90%	456999	N. 153578	N. 7,096 km2 (2,740 sq mi)	86/km2 (220/sq mi)	890	22
tamil nadu	6	72147030	N. 15.60%	37229590	N. 34917440	N. 130,058 km2 (50,216 sq mi)	555/km2 (1,440/sq mi)	996	23
tripura	21	3673917	N. 14.80%	2712464	N. 961453	N. 10,486 km2 (4,049 sq mi)	350/km2 (910/sq mi)	960	25
uttar pradesh	1	199812341	N. 20.20%	155317278	N. 44495063	N. 240,928 km2 (93,023 sq mi)	828/km2 (2,140/sq mi)	912	26
west bengal	4	91276115	N. 13.80%	62183113	N. 29093002	N. 88,752 km2 (34,267 sq mi)	1,029/km2 (2,670/sq mi)	953	28
puducherry	31	1247953	N. 28.10%	395200	N. 852753	N. 479 km2 (185 sq mi)	2,598/km2 (6,730/sq mi)	1037	34
andhra pradesh	12	35003674	N. 13.58%	21395009	N. 13608665	N. 112,077 km2 (43,273 sq mi)	312/km2 (810/sq mi)	988	1
andhra pradesh	10	49577103	N. 11.00%	34966693	N. 14610410	N. 162,968 km2 (62,922 sq mi)	303/km2 (780/sq mi)	993	1
arunachal prad...	26	1383727	N. 26.00%	1066358	N. 317369	N. 83,743 km2 (32,333 sq mi)	17/km2 (44/sq mi)	938	2
assam	15	31205576	N. 17.10%	26807034	N. 4398542	N. 78,438 km2 (30,285 sq mi)	397/km2 (1,030/sq mi)	954	3
bihar	3	104099452	N. 25.40%	92341436	N. 11758016	N. 94,163 km2 (36,357 sq mi)	1,102/km2 (2,850/sq mi)	918	4
chhattisgarh	17	25545198	N. 22.60%	19607961	N. 5937237	N. 135,191 km2 (52,198 sq mi)	189/km2 (490/sq mi)	991	5
goa	25	1458545	N. 8.20%	551731	N. 906814	N. 3,702 km2 (1,429 sq mi)	394/km2 (1,020/sq mi)	973	6



s_no	2011-12	2012-133	2013-144	2014-155	2015-166	2016-177	2017-188	2018-199	2019-2010	2020-2111
16	19918	21872	22938	23235	25117	27439	29508	32176	34770	33776
17	7259	8362	10293	13509	15139	17192	19385	21912	24990	23923
18	12177	14121	16612	18401	19524	21722	24393	26527	29716	29832
20	266628	297734	332147	355102	390087	426988	471014	512510	537031	540853
21	434837	493551	551031	615642	681482	760587	832529	911519	1000032	1017917
22	11165	12338	13862	15407	18034	20687	25971	28402	31441	33018
23	751486	854825	968530	1072678	1176500	1302639	1465051	1630209	1743144	1788074
25	19208	21663	25593	29533	35938	39479	43716	49823	54151	53504
26	724050	822393	940356	1011790	1137808	1288700	1439925	1582180	1700062	1640097
28	520485	591465	676849	718082	797300	872527	974700	1102054	1179127	1141798
34	16818	18875	21870	22574	26617	29573	32129	34171	36999	36180
1	379402	411404	464272	524976	604229	684416	786135	873721	925839	978581
1	379402	411404	464272	524976	604229	684416	786135	873721	925839	978581
2	11063	12547	14581	17959	18509	19902	22475	25335	30024	30525
3	143175	156864	177745	195723	227959	254382	283165	309336	346851	339803
4	247144	282368	317101	342951	371602	421051	468746	527976	581855	567814
5	158074	177511	206833	221118	225163	262802	282737	327107	344672	352328
6	42367	38120	35921	47814	55054	62976	69352	71853	75032	74158



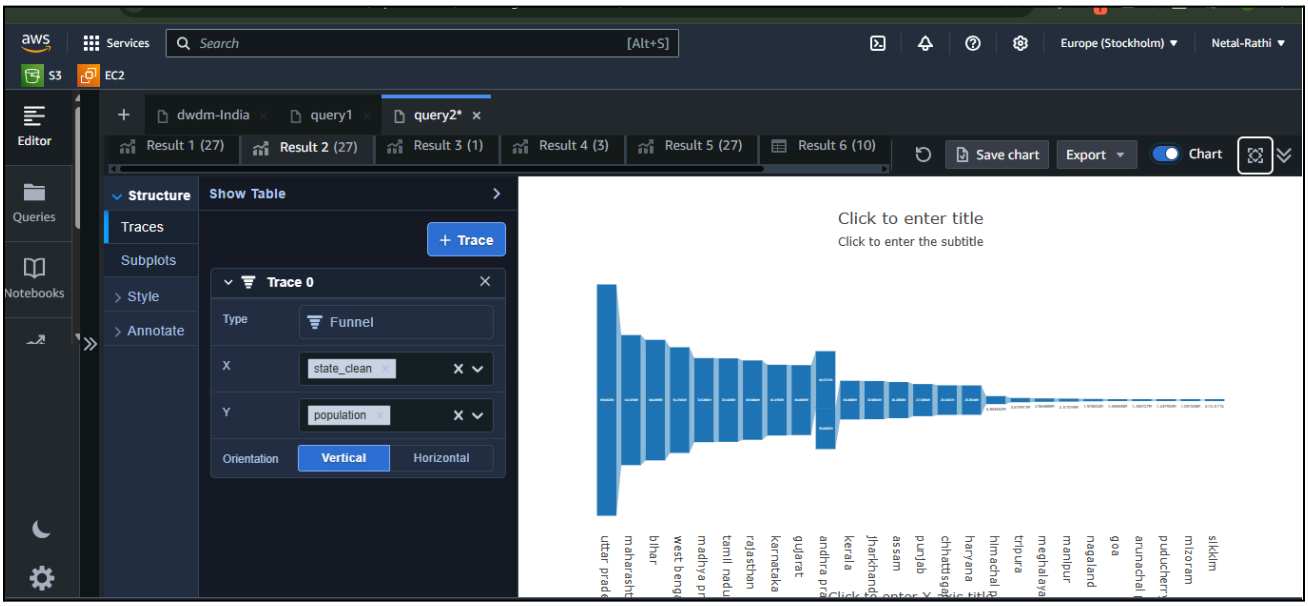
Query2:

```
--sorting in desc order according to population
SELECT state_clean, Population
FROM spectrum_schema.output
ORDER BY CAST(Population AS BIGINT) DESC;
```

Result2:

Result 1 (27) | Result 2 (27) | Result 3

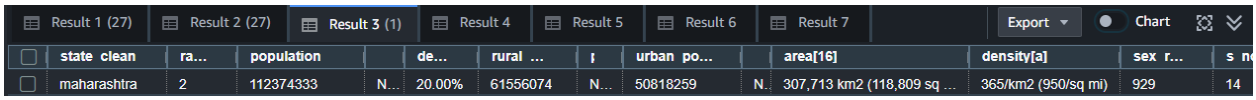
state_clean	population
uttar pradesh	199812341
maharashtra	112374333
bihar	104099452
west bengal	91276115
madhya pradesh	72626809
tamil nadu	72147030
rajasthan	68548437
karnataka	61095297
gujarat	60439692
andhra pradesh	49577103
andhra pradesh	35003674
kerala	33406061
jharkhand	32988134
assam	31205576
punjab	27743338
chhattisgarh	25545198
haryana	25351462
himachal pradesh	6864602
tripura	3673917
meghalaya	2966889
manipur	2570390
nagaland	1978502
goa	1458545
arunachal pradesh	1383727
puducherry	1247953
mizoram	1097206
sikkim	610577



Query3:

```sql
--maharashtra data
SELECT *
FROM spectrum_schema.output
WHERE LOWER(state_clean) = 'maharashtra';
```

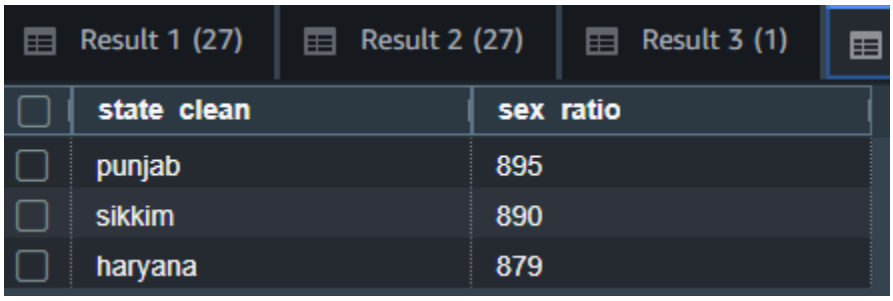
Result3:

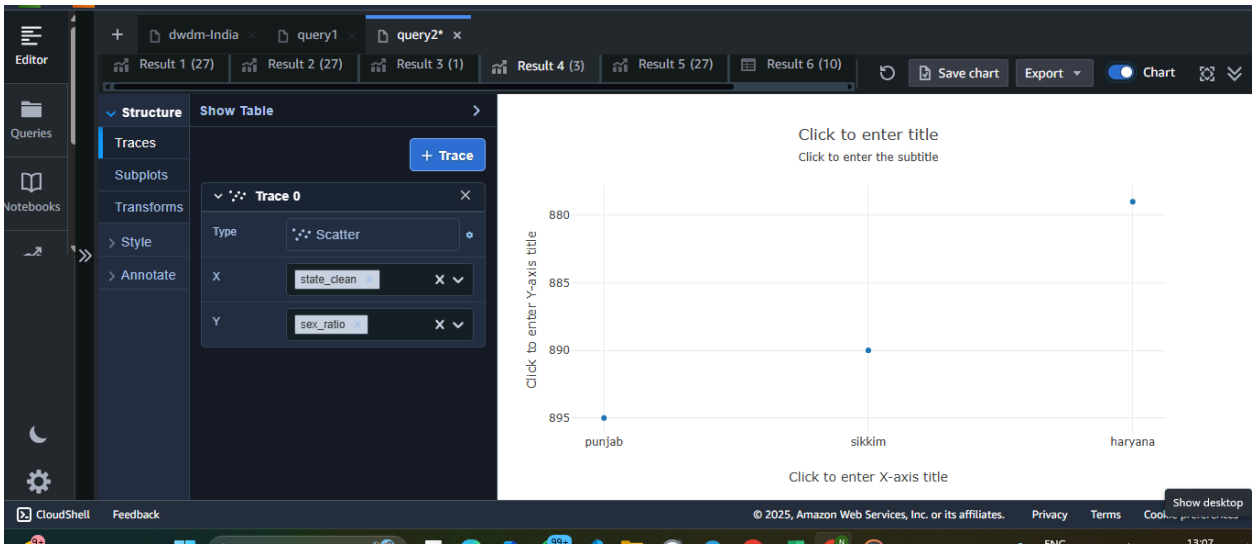


Query4:

```sql
--state and sex ratio of the states where sex ratio < 900
SELECT state_clean, Sex_ratio
FROM spectrum_schema.output
WHERE CAST(Sex_ratio AS INT) < 900;
```

Result4:





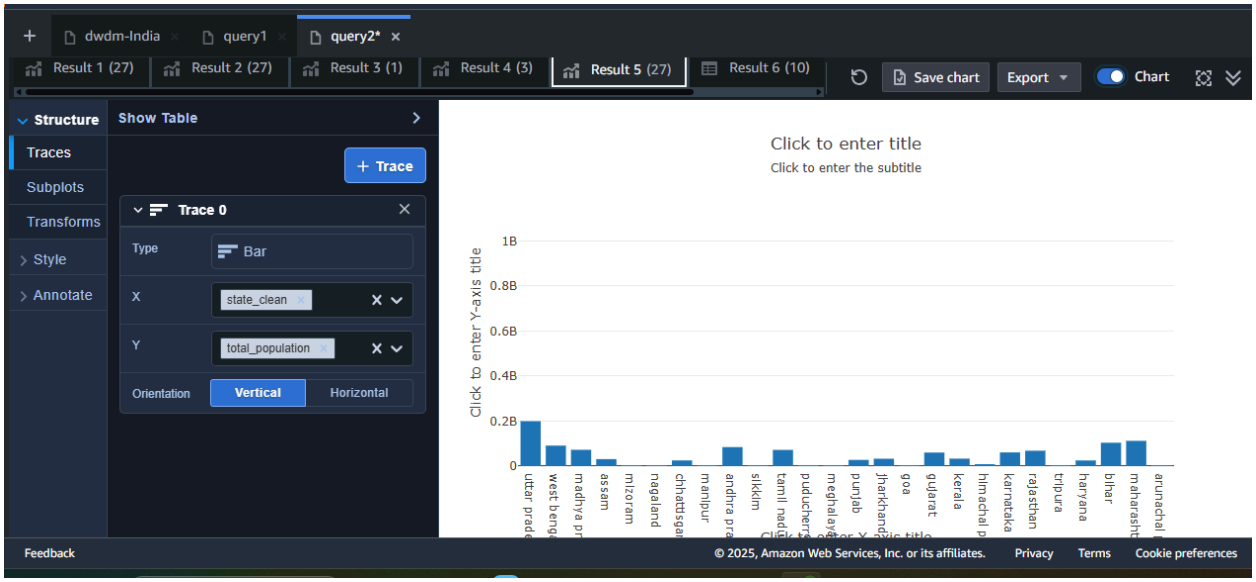
Query5:

```
--state wise total population
SELECT state_clean, SUM(CAST(Population AS BIGINT)) AS total_population
FROM spectrum_schema.output
GROUP BY ROLLUP(state_clean);
```

Result5:

Result 1 (27) | Result 2 (27) | Result 3 (1)

state_clean	total_population
sikkim	610577
tamil nadu	72147030
puducherry	1247953
meghalaya	2966889
punjab	27743338
jharkhand	32988134
goa	1458545
gujarat	60439692
kerala	33406061
himachal pradesh	6864602
karnataka	61095297
mizoram	1097206
nagaland	1978502
chhattisgarh	25545198
manipur	2570390
andhra pradesh	84580777
uttar pradesh	199812341
west bengal	91276115
madhya pradesh	72626809
assam	31205576
rajasthan	68548437
tripura	3673917
haryana	25351462
bihar	104099452
maharashtra	112374333
arunachal pradesh	1383727
NULL	1127092360



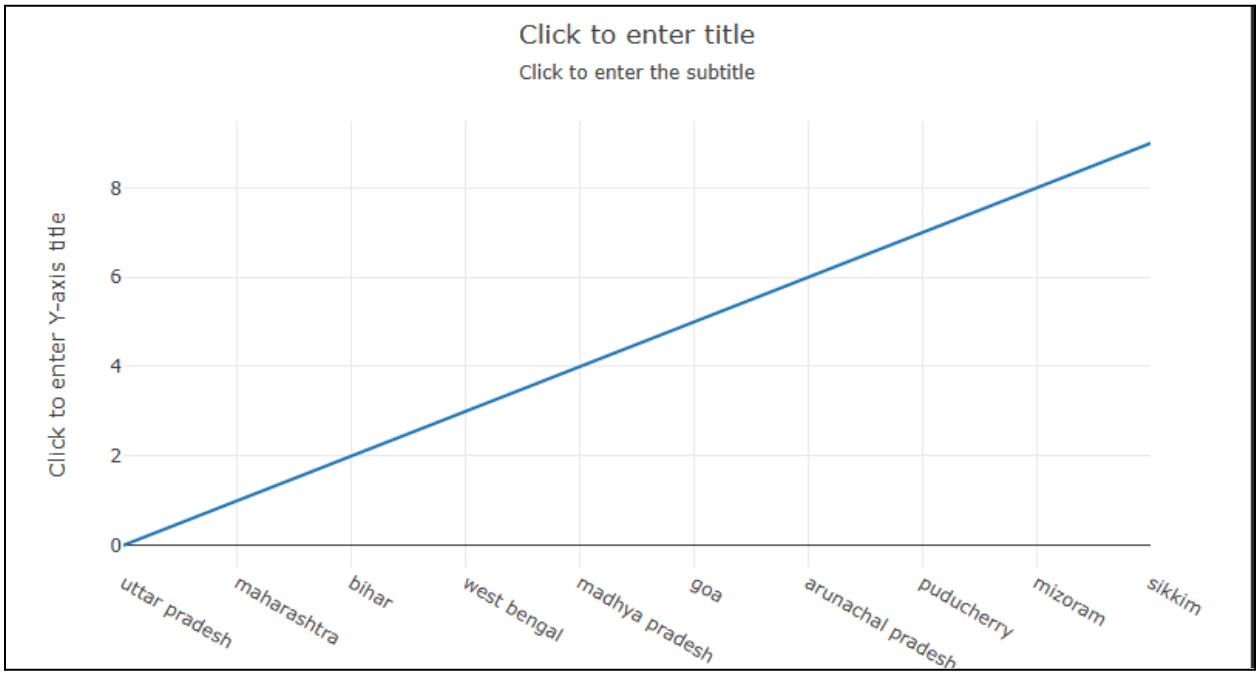
Query6:

```
(
  SELECT state_clean,
         SUM(2018-199) / SUM(population) AS per_capita_gsdp
  FROM spectrum_schema.output
  GROUP BY state_clean
  ORDER BY per_capita_gsdp ASC
  LIMIT 5
)
UNION
(
  SELECT state_clean,
         SUM(2018-199) / SUM(population) AS per_capita_gsdp
  FROM spectrum_schema.output
  GROUP BY state_clean
  ORDER BY per_capita_gsdp DESC
  LIMIT 5
)
ORDER BY per_capita_gsdp ASC;
```

Result6:

Result 1 (27) | Result 2 (27) | Result 3 (1)

state_clean	per_capita_gsdp
uttar pradesh	0.0000091035418077605
maharashtra	0.0000161869703822847
bihar	0.0000174736750775594
west bengal	0.0000199285431900777
madhya pradesh	0.000025045847739228
goa	0.0012471332732277715
arunachal pradesh	0.0013145656621573475
puducherry	0.0014575869443801168
mizoram	0.0016578472957676133
sikkim	0.0029791492309733252



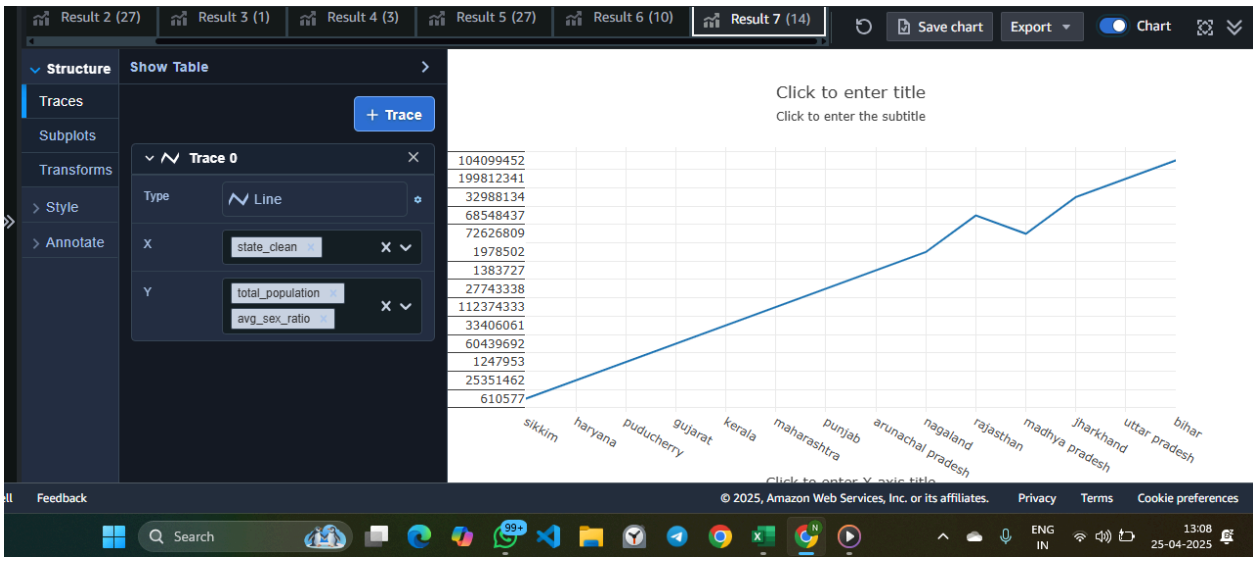
Query7:

```
SELECT state_clean,
       SUM(population) AS total_population,
       AVG(sex_ratio) AS avg_sex_ratio,
       SUM("2018-199") / SUM(population) AS per_capita_gsdp
FROM spectrum_schema.output
WHERE sex_ratio < 950
GROUP BY state_clean
ORDER BY per_capita_gsdp DESC
```

Result7:

Result 1 (27) | Result 2 (27) | Result 3 (1) | Result 4 (3) | Result 5 (27) | Result 6 (10)

state_clean	total_population	avg_sex_ratio	per_capita_gsdp
sikkim	610577	890	0.0465
haryana	25351462	879	0.0275
puducherry	1247953	1037	0.0273
gujarat	60439692	919	0.0246
kerala	33406061	1084	0.0235
maharashtra	112374333	929	0.0225
punjab	27743338	895	0.0184
arunachal pradesh	1383727	938	0.0183
nagaland	1978502	931	0.0134
rajasthan	68548437	928	0.0132
madhya pradesh	72626809	931	0.0114
jharkhand	32988134	948	0.0092
uttar pradesh	199812341	912	0.0079
bihar	104099452	918	0.005



Sources Of dataset:

1) https://www.education.gov.in/statistics-new/statistics-new
2) https://mospi.gov.in/data