

PROJECT-1

TRIMESTER-I

- "SQL ASSIGNMENT"

INTRODUCTION

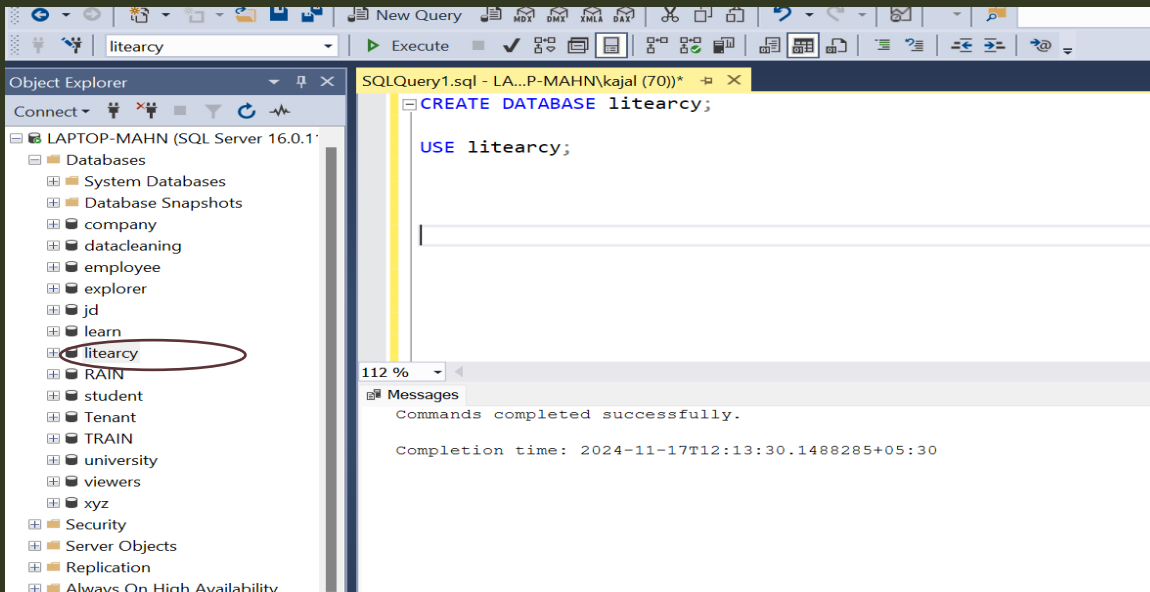
- **Literacy:** Contains information such as serial number (sr no), district, state, and literacy rate (lit_rate).
- **Population:** Contains data about states and their population.

Using SQL and Python, we performed various operations like data retrieval, analysis, and visualization to explore and analyze the data efficiently. This project highlights the power of combining SQL for database management and Python for advanced analysis in a seamless environment.

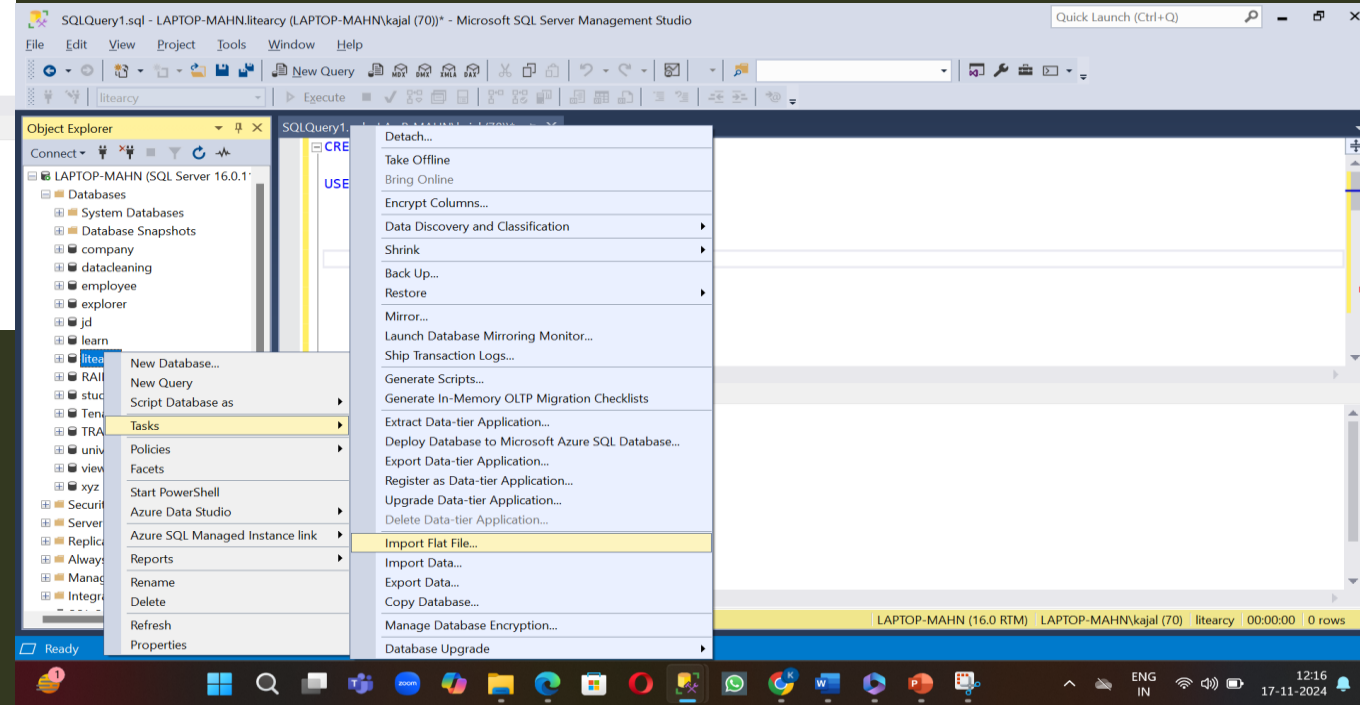


SQL+PYTHON

CREATING DATABASE AND IMPORTING EXCEL FILES.

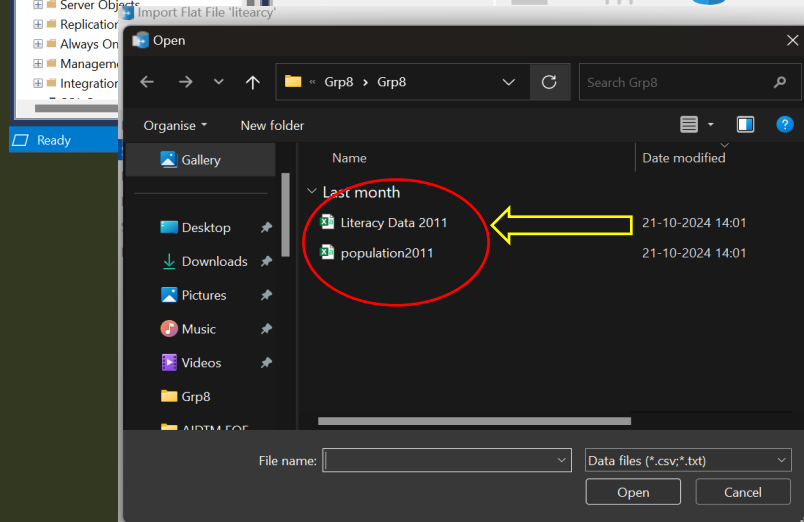
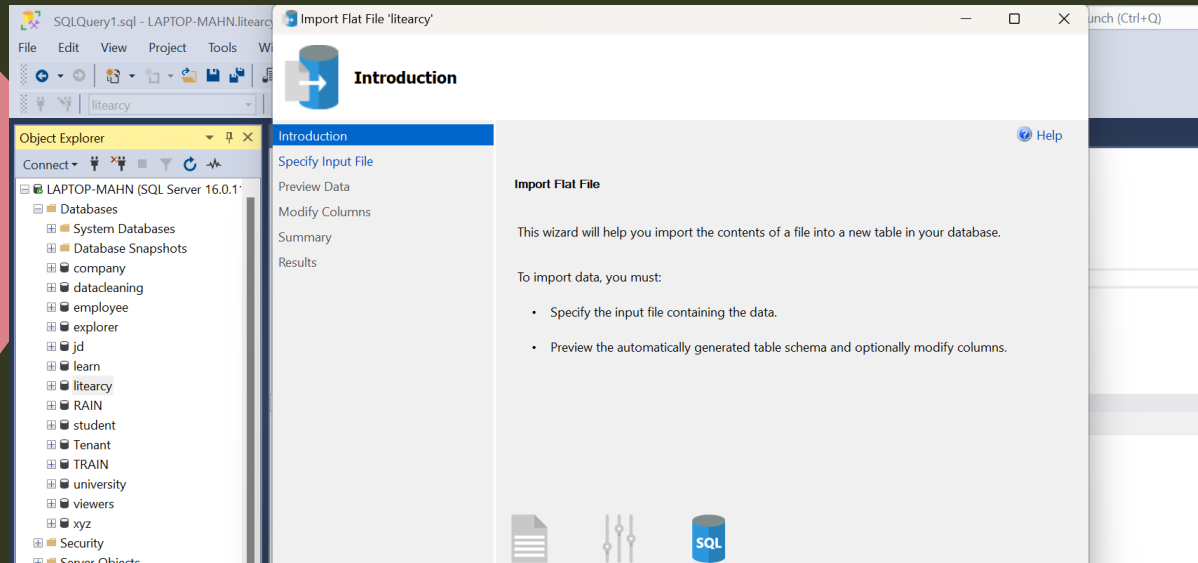


- To import the excel files click on task option and then on import flat files as shown in the image.



- After creating a database refresh the object explorer.
- You will find the created database inside the sys.databases
- Right click on the created database to look for task option.

- After importing flat file you will find the import file window click ok/next.
- Then browse the file and upload it.
- And at last you will find the preview. data window click on next .



Preview Data

Introduction

Specify Input File

Preview Data

Modify Columns

Summary

Results

Preview Data

This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.

sr_no	District	State	Literacy
1	Thane	Maharas...	84.53
2	North Tw...	West Ben...	84.06
3	Bangalore	Karnataka	87.67
4	Pune	Maharas...	86.15
5	Mumbai ...	Maharas...	89.91
6	South Tw...	West Ben...	77.51
7	Barddha...	West Ben...	76.21
8	Ahmada...	Gujarat	85.31
9	Murshida...	West Ben...	66.59
10	Jaipur	Rajasthan	75.51
11	Nashik	Maharas...	82.31
12	Surat	Gujarat	85.53
13	Allahabad	Uttar Pra...	72.32

Column names changed due to invalid characters, duplication, etc. Column names can be edited in Modify Columns page.

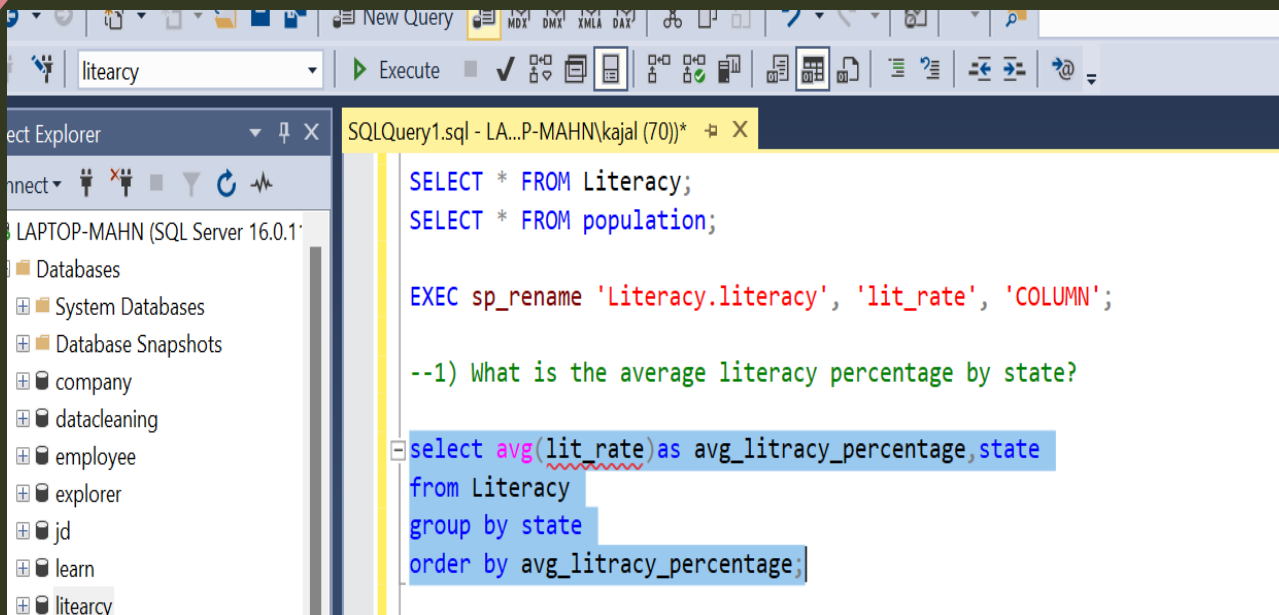
☒ Use Rich Data Type Detection - may provide a closer type fit. However, cells with anomalous values may be dropped.

< Previous

Next >

Cancel

Query 1) What is the average literacy percentage by state?



```

SELECT * FROM Literacy;
SELECT * FROM population;

EXEC sp_rename 'Literacy.literacy', 'lit_rate', 'COLUMN';

--1) What is the average literacy percentage by state?

select avg(lit_rate) as avg_litracy_percentage, state
from Literacy
group by state
order by avg_litracy_percentage;
    
```

INTERPRETATION AND ANALYSIS

The results show how well-educated people are in each state on average (based on literacy rates).

A **higher average literacy rate** indicates a more literate population in that state, while a **lower average** suggests areas for improvement.

from Literacy

112 %

Results Messages

	avg_litracy_percentage	state
1	60.9927271062678	Bihar
2	67.3914814701787	Uttar Pradesh
3	67.9626668294271	Andhra Pradesh
4	70.7233352661133	Rajasthan
5	71.0899963378906	Orissa
6	76.5499995095389	West Bengal
7	77.3099975585938	Chhattisgarh
8	80.5750007629395	Karnataka
9	80.9414280482701	Tamil Nadu
10	81.9581812078303	Maharashtra
11	82.1999969482422	Punjab

Query executed successfully.

output

Query) Which state has the highest literacy percentage overall?

QUERY

```
SQLQuery1.sql - LA...P-MAHN\kajal (70))*
group by state
order by avg_litracy_percentage;

--2) Which state has the highest literacy percentage overall?

select top 1 avg(lit_rate) as avg_litracy_percentage, state
from Literacy
group by state
order by avg_litracy_percentage desc;
```

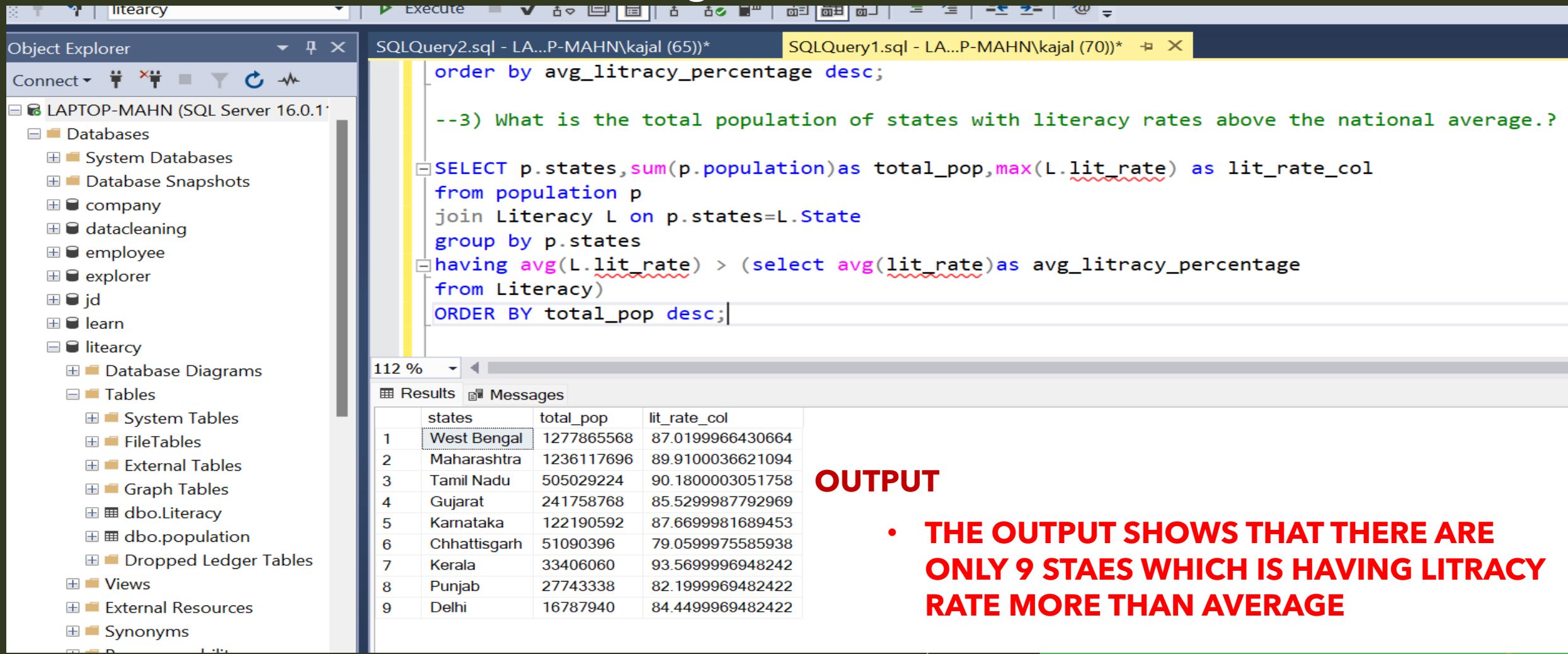
112 %

Results Messages

	avg_litracy_percentage	state
1	93.5699996948242	Kerala

OUTPUT

QUERY 3) What is the total population of states with literacy rates above the national average.?



The screenshot displays the SQL Server Enterprise Manager interface. On the left, the Object Explorer shows the database structure for 'LAPTOP-MAHN (SQL Server 16.0.1)'. The main window shows a SQL query in 'SQLQuery2.sql' with the following text:

```
order by avg_litracy_percentage desc;

--3) What is the total population of states with literacy rates above the national average.?
```

```
SELECT p.states, sum(p.population) as total_pop, max(L.lit_rate) as lit_rate_col
from population p
join Literacy L on p.states=L.State
group by p.states
having avg(L.lit_rate) > (select avg(lit_rate) as avg_litracy_percentage
from Literacy)
ORDER BY total_pop desc;
```

Below the query editor, the 'Results' tab shows the output of the query. The results are displayed in a table with the following data:

	states	total_pop	lit_rate_col
1	West Bengal	1277865568	87.0199966430664
2	Maharashtra	1236117696	89.9100036621094
3	Tamil Nadu	505029224	90.1800003051758
4	Gujarat	241758768	85.5299987792969
5	Karnataka	122190592	87.6699981689453
6	Chhattisgarh	51090396	79.0599975585938
7	Kerala	33406060	93.5699996948242
8	Punjab	27743338	82.1999969482422
9	Delhi	16787940	84.4499969482422

OUTPUT

- THE OUTPUT SHOWS THAT THERE ARE ONLY 9 STATES WHICH ARE HAVING LITERACY RATE MORE THAN AVERAGE

Query 4) What is the average literacy percentage OF states with a population over 10 million?

The screenshot displays the SQL Server Enterprise Manager interface. On the left, the 'Databases' folder is expanded, showing 'LAPTOP-MAHN (SQL Server 16.0.1)'. The 'Tables' folder is also expanded, listing various system and user tables. The main pane shows a SQL query in the 'Query Editor' window. The query is as follows:

```
ORDER BY total_pop desc;

--4) What is the average literacy percentage OF states with a population over 10 million?

select avg(L.lit_rate) as avg_lit_rate,p.states
from Literacy L
join population p on p.states=L.state
where p.population > 1000000
Group by states
order by avg_lit_rate desc;
```

The query is highlighted in blue. To the right of the query, the word 'query' is written in red. Below the query, the 'Results' tab is selected, showing the output of the query. The output is a table with two columns: 'avg_lit_rate' and 'states'. The data is sorted in descending order of 'avg_lit_rate'.

	avg_lit_rate	states
1	93.5699996948242	Kerala
2	84.4499996948242	Delhi
3	82.6799983978271	Gujarat
4	82.1999969482422	Punjab
5	81.9581812078303	Maharashtra
6	80.9414280482701	Tamil Nadu
7	80.5750007629395	Karnataka
8	77.3099975585938	Chhattisgarh
9	76.5499995095389	West Bengal
10	71.0899963378906	Orissa
11	70.7233352661133	Rajasthan

The output is highlighted in blue. To the left of the output, the word 'output' is written in red. Below the output, a status bar indicates 'Query executed successfully.'

- Average of literacy rate is been shown in descending order to maintain an orderly format otherwise without that also we can run this query.

Query 5) Which districts have a literacy percentage below 60%?

```
--5)Which districts have a literacy percentage below 60%?
```

```
select district,state,lit_rate  
from Literacy  
where lit_rate <60.00;
```

QUERY

112 %

Results Messages

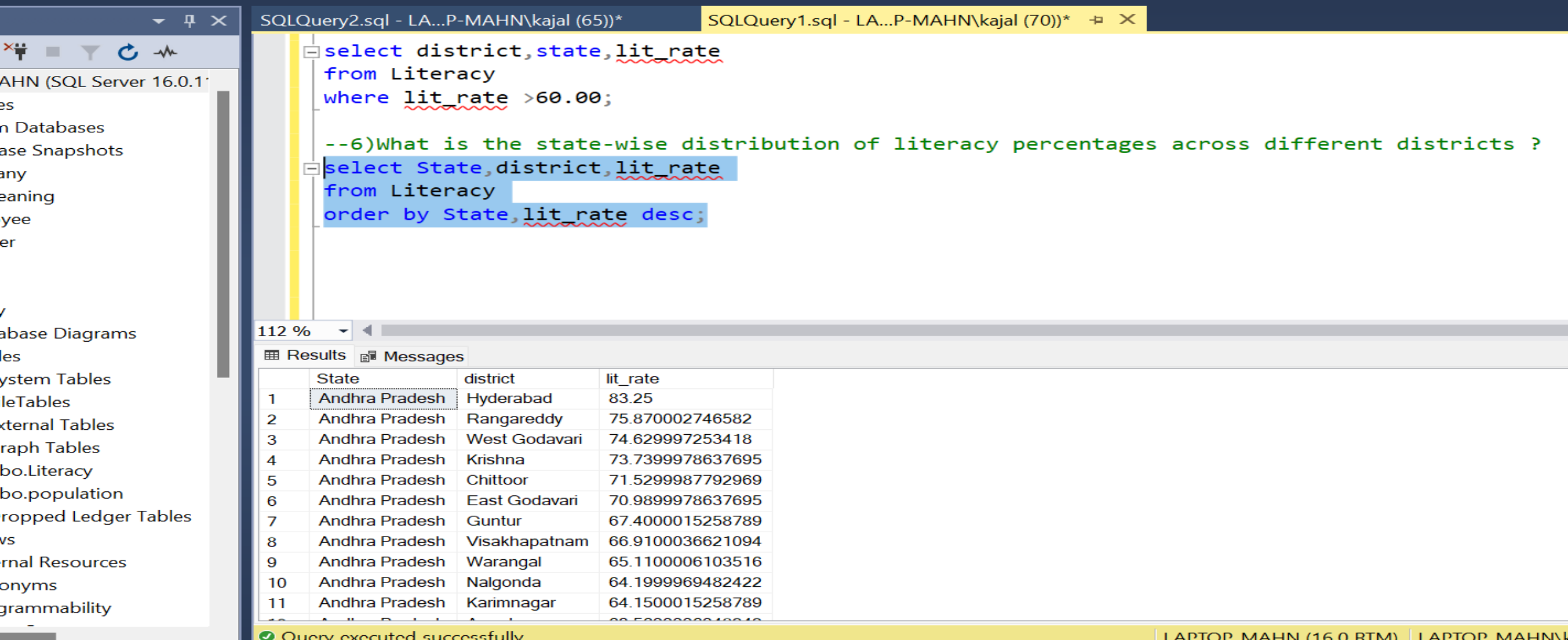
	district	state	lit_rate
1	Purbi Champaran	Bihar	55.7900009155273
2	Moradabad	Uttar Pradesh	56.7700004577637
3	Madhubani	Bihar	58.6199989318848
4	Bareilly	Uttar Pradesh	58.4900016784608
5	Kurnool	Andhra Pradesh	59.9700012207031
6	Mahbubnagar	Andhra Pradesh	55.0400009155273
7	Darbhanga	Bihar	56.560001373291
8	Pashchim Champaran	Bihar	55.7000007629395
9	Budaun	Uttar Pradesh	51.2900009155273
10	Bahraich	Uttar Pradesh	49.3600006103516
11	Gonda	Uttar Pradesh	58.7099990844727

OUTPUT

Query executed successfully.

LAPTOP-MAHN (16.0 RTM) | LAPTOP-MAHN\kajal (52) | litearcy | 00:00:00 | 12 rows

Query 6) What is the state-wise distribution of literacy percentages across different districts ?



The screenshot shows the SQL Server Enterprise Manager interface. The left pane displays the database structure for 'LAPTOP MAHN (16.0.10586.1)'. The central pane shows the execution of two SQL queries. The first query filters for literacy rates greater than 60.00. The second query, which is highlighted, sorts the data by state and then by literacy rate in descending order. The bottom pane displays the results of the second query as a table with 11 rows, all from Andhra Pradesh, sorted by literacy rate from highest to lowest.

SQLQuery2.sql - LA...P-MAHN\kajal (65))*

```
select district, state, lit_rate
from Literacy
where lit_rate > 60.00;
```

--6)What is the state-wise distribution of literacy percentages across different districts ?

```
select State, district, lit_rate
from Literacy
order by State, lit_rate desc;
```

112 %

Results Messages

	State	district	lit_rate
1	Andhra Pradesh	Hyderabad	83.25
2	Andhra Pradesh	Rangareddy	75.870002746582
3	Andhra Pradesh	West Godavari	74.629997253418
4	Andhra Pradesh	Krishna	73.7399978637695
5	Andhra Pradesh	Chittoor	71.5299987792969
6	Andhra Pradesh	East Godavari	70.9899978637695
7	Andhra Pradesh	Guntur	67.4000015258789
8	Andhra Pradesh	Visakhapatnam	66.9100036621094
9	Andhra Pradesh	Warangal	65.1100006103516
10	Andhra Pradesh	Nalgonda	64.1999969482422
11	Andhra Pradesh	Karimnagar	64.1500015258789

Query executed successfully

LAPTOP MAHN (16.0.10586.1) | LAPTOP MAHN

Quey 7) Find the state with the lowest literacy percentage ?

```
group by state
order by minimum_lit_rate;

select top 1 state, avg(lit_rate) as average_lit_rate
from Literacy
group by state
order by average_lit_rate;
```

112 %

Results Messages

	state	average_lit_rate
1	Bihar	60.9927271062678

Here we used (avg) to analyse the overall minimum literacy percentage.

✓ Query executed successfully.

LAPTOP-MAHN (16.0 RTM) | LAPTOP-MAHN\kajal (70) | litearcy | 00:00:00

OTHER INTERPRITATION AND ANALYSIS

- **STATE WISE DISTRIBUTION ANALYSIS** -: PERCENTAGE CONTRIBUTION OF EACH DISTRICT'S LITERACY RATE TO THE OVERALL LITERACY RATE OF ITS STATE.
- **LITERACY GAPS** -: ANALYSIS OF THE GAP BETWEEN THE HIGHEST AND LOWEST LITERACY RATES.
- **LITERACY CLUSTERING** -: CLUSTERING DISTRICT INTO LITERACY CATEGORIES SUCH AS
'LOW'
'MODERATE'
'HIGH'

THANKYOU...!