

Internship Program : Soulvibe.Tech

PROJECT TITLE:

**“Analyzing Socioeconomic
and Demographic Influences
on Income”**

BATCH NAME : SVT/DAINT/2025/06/B09.

BY

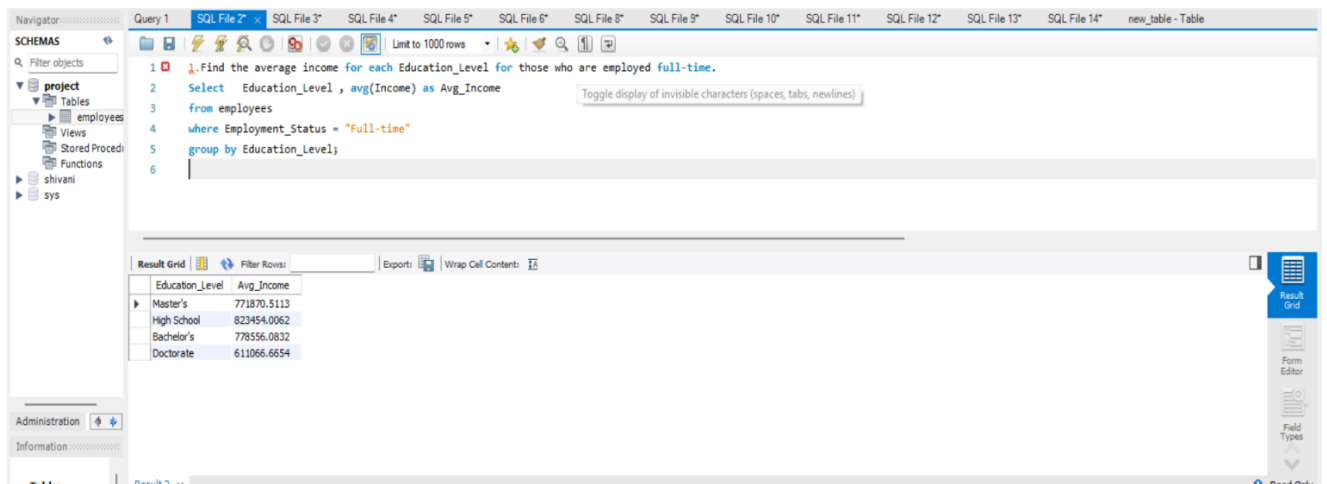
K.SHIVANI

Introduction

Overview of the main objective

I conducted a thorough analysis of socioeconomic and demographic data for this assignment using SQL. I gained important insights into the variables affecting income levels and living circumstances by using queries to filter, group, and aggregate data on income, household characteristics, employment, and education. A greater comprehension of the connections between individual demographics and economic results was made possible by this analytical method. The results offer a strong basis for strategic planning and well-informed decision-making.

1. Find the average income for each Education_Level for those who are employed full-time?



The screenshot shows a SQL IDE interface. The top pane displays a query in a file named 'SQL File 2*'. The query is as follows:

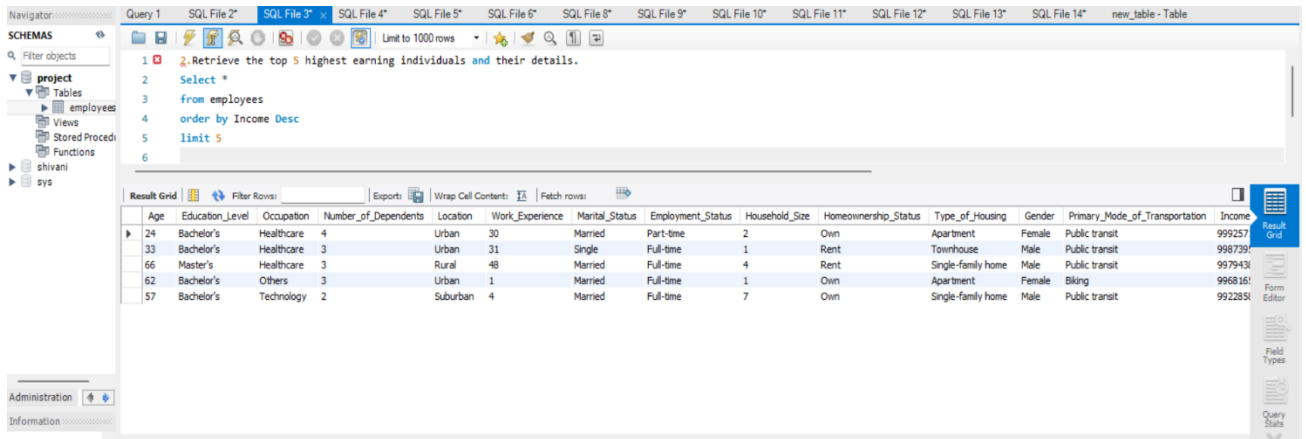
```
1. Find the average income for each Education_Level for those who are employed full-time.
2. Select Education_Level , avg(Income) as Avg_Income
3. from employees
4. where Employment_Status = "Full-time"
5. group by Education_Level;
```

The bottom pane shows the 'Result Grid' with the following data:

Education_Level	Avg_Income
Master's	771870.5113
High School	823454.0062
Bachelor's	778556.0832
Doctorate	611066.6654

In order to examine how education affects salaries in the full-time workforce, this query determines the average income for each level of education among those who are employed full-time.

2. Retrieve the top 5 highest earning individuals and their details?



The screenshot shows a SQL IDE interface with a query editor and a results grid. The query is as follows:

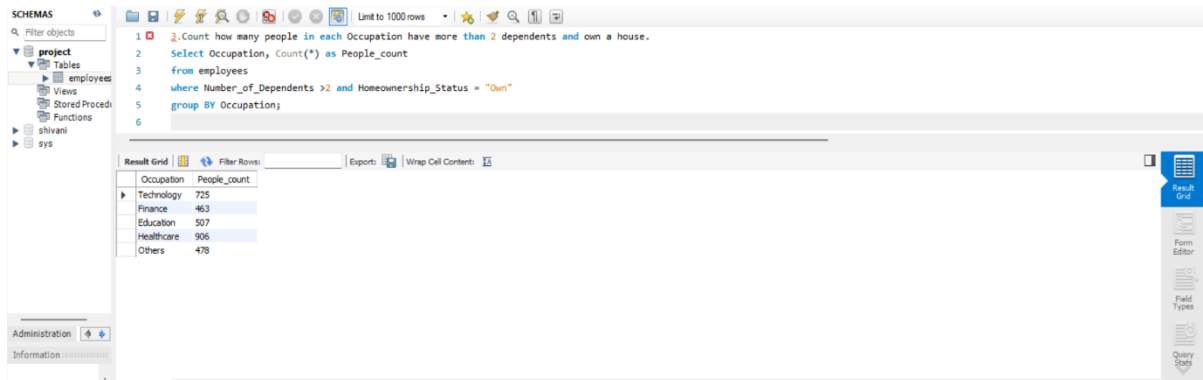
```
1 Retrieve the top 5 highest earning individuals and their details.
2 Select *
3 from employees
4 order by Income Desc
5 limit 5
6
```

The results grid displays the following data:

	Age	Education_Level	Occupation	Number_of_Dependents	Location	Work_Experience	Marital_Status	Employment_Status	Household_Size	Homeownership_Status	Type_of_Housing	Gender	Primary_Mode_of_Transportation	Income
24	24	Bachelor's	Healthcare	4	Urban	30	Married	Part-time	2	Own	Apartment	Female	Public transit	999257
33	33	Bachelor's	Healthcare	3	Urban	31	Single	Full-time	1	Rent	Townhouse	Male	Public transit	998739
66	66	Master's	Healthcare	3	Rural	48	Married	Full-time	4	Rent	Single-family home	Male	Public transit	997943
62	62	Bachelor's	Others	3	Urban	1	Married	Full-time	1	Own	Apartment	Female	Biking	996816
57	57	Bachelor's	Technology	2	Suburban	4	Married	Full-time	7	Own	Single-family home	Male	Public transit	992285

This query limits the output to the top five entries and sorts the dataset in descending order of income to return the top five highest earning persons along with all of their details. For more research or comparison, it assists in identifying high-income profiles.

3. Count how many people in each Occupation have more than 2 dependents and own a house?



The screenshot shows a SQL query editor with a query window and a results grid. The query is as follows:

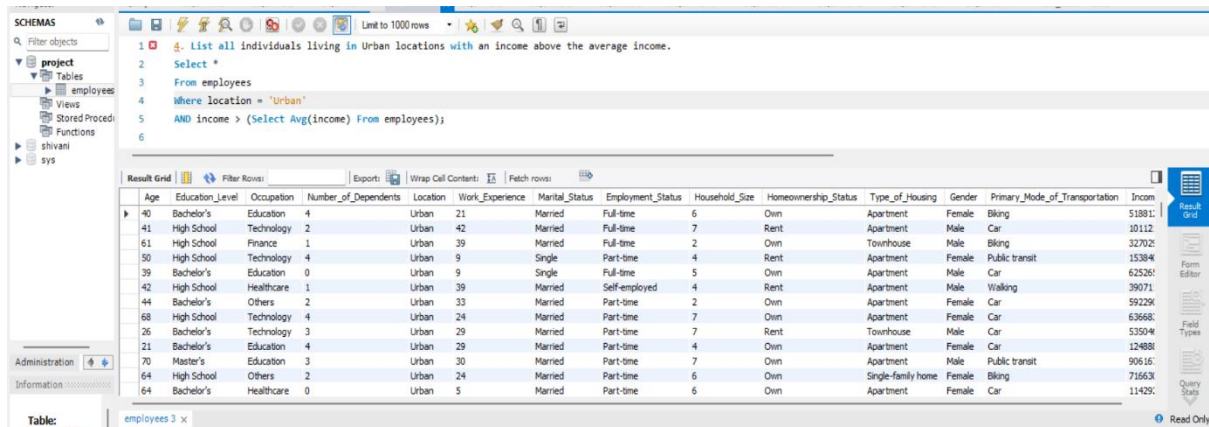
```
1 3. Count how many people in each Occupation have more than 2 dependents and own a house.
2 Select Occupation, Count(*) as People_count
3 from employees
4 where Number_of_Dependents > 2 and Homeownership_Status = "Own"
5 group BY Occupation;
6
```

The results grid displays the following data:

Occupation	People_count
Technology	725
Finance	463
Education	507
Healthcare	906
Others	478

The purpose of this query is to examine how homeownership, family size, and occupation are related. Those who own their homes and have more than two dependents are the target audience. A better insight of which occupational groupings are more likely to support larger families while keeping homeownership is provided by the query, which groups the data by occupation and counts the number of people who match both criteria. Assessing housing patterns across various work sectors and finding financially solid professions can both benefit from this knowledge.

4. List all individuals living in Urban locations with an income above the average income?



The screenshot shows a database query tool interface. The SQL query is as follows:

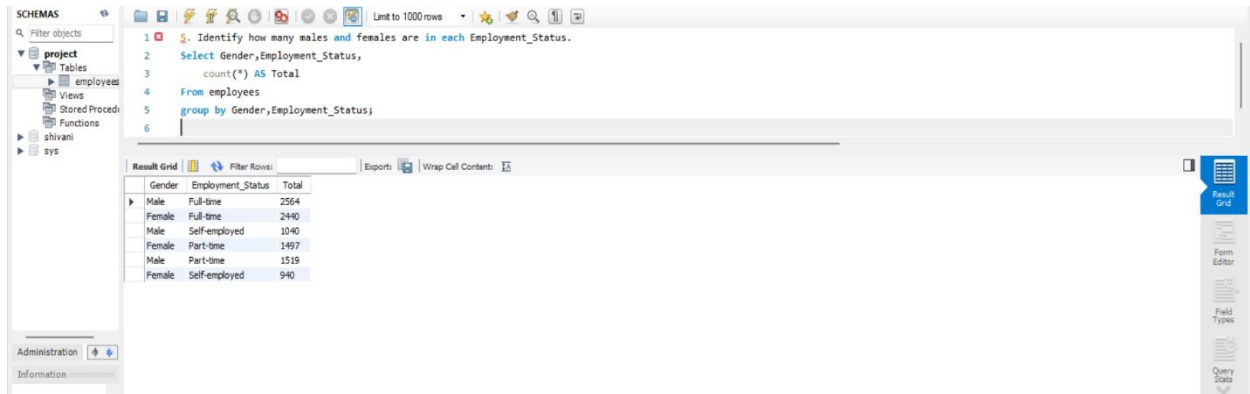
```
1 List all individuals living in Urban locations with an income above the average income.
2
3 Select *
4 From employees
5 Where location = 'Urban'
6 AND income > (Select Avg(income) From employees);
```

The results are displayed in a table with the following columns: Age, Education_Level, Occupation, Number_of_Dependents, Location, Work_Experience, Marital_Status, Employment_Status, Household_Size, Homeownership_Status, Type_of_Housing, Gender, Primary_Mode_of_Transportation, and Income. The table contains 14 rows of data, all of which are individuals living in Urban locations with an income above the average.

Age	Education_Level	Occupation	Number_of_Dependents	Location	Work_Experience	Marital_Status	Employment_Status	Household_Size	Homeownership_Status	Type_of_Housing	Gender	Primary_Mode_of_Transportation	Income
40	Bachelor's	Education	4	Urban	21	Married	Full-time	6	Own	Apartment	Female	Biking	51881
41	High School	Technology	2	Urban	42	Married	Full-time	7	Rent	Apartment	Male	Car	10112
61	High School	Finance	1	Urban	39	Married	Full-time	2	Own	Townhouse	Male	Biking	32702
50	High School	Technology	4	Urban	9	Single	Part-time	4	Rent	Apartment	Female	Public transit	15384
39	Bachelor's	Education	0	Urban	9	Single	Full-time	5	Own	Apartment	Male	Car	62526
42	High School	Healthcare	1	Urban	39	Married	Self-employed	4	Rent	Apartment	Male	Walking	39071
44	Bachelor's	Others	2	Urban	33	Married	Part-time	2	Own	Apartment	Female	Car	59229
68	High School	Technology	4	Urban	24	Married	Part-time	7	Own	Apartment	Female	Car	63668
26	Bachelor's	Technology	3	Urban	29	Married	Part-time	7	Rent	Townhouse	Male	Car	53504
21	Bachelor's	Education	4	Urban	29	Married	Part-time	4	Own	Apartment	Female	Car	12488
70	Master's	Education	3	Urban	30	Married	Part-time	7	Own	Apartment	Male	Public transit	90616
64	High School	Others	2	Urban	24	Married	Part-time	6	Own	Single-family home	Female	Biking	71663
64	Bachelor's	Healthcare	0	Urban	5	Married	Part-time	6	Own	Apartment	Female	Car	11429

This search finds people with higher-than-average incomes who live in cities. By highlighting high incomes in urban areas, it offers valuable information for economic research, resource planning, and targeted marketing.

5. Identify how many males and females are in each Employment_Status?



The screenshot shows a SQL query editor with a query window and a results grid. The query is as follows:

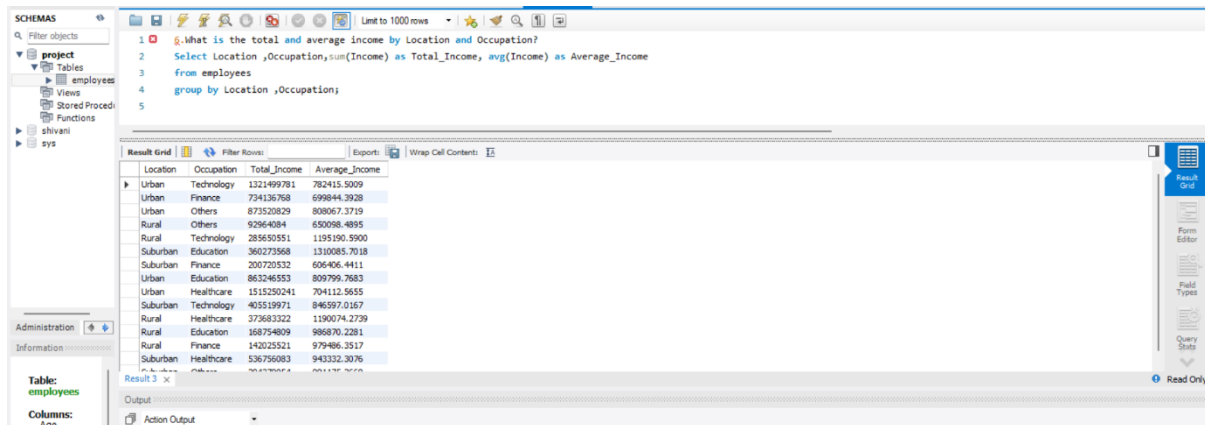
```
1 Identify how many males and females are in each Employment_Status.  
2 Select Gender, Employment_Status,  
3 count(*) AS Total  
4 From employees  
5 group by Gender, Employment_Status;  
6
```

The results grid displays the following data:

Gender	Employment_Status	Total
Male	Full-time	2564
Female	Full-time	2440
Male	Self-employed	1040
Female	Part-time	1497
Male	Part-time	1519
Female	Self-employed	940

This query counts the number of men and women in each Employment_Status category. Understanding the gender distribution across various employment types and identifying potential differences in gender are made easier by classifying the data by gender and employment status.

6. What is the total and average income by Location and Occupation?



The screenshot shows a SQL query editor with a query window and a result grid. The query is as follows:

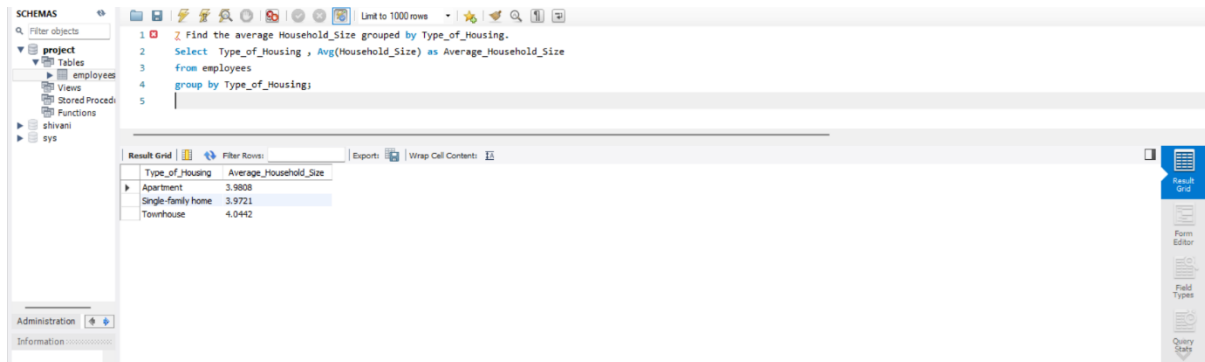
```
1 6. What is the total and average income by Location and Occupation?  
2 Select Location ,Occupation,sum(Income) as Total_Income, avg(Income) as Average_Income  
3 from employees  
4 group by Location ,Occupation;
```

The result grid displays the following data:

Location	Occupation	Total_Income	Average_Income
Urban	Technology	1321499781	782415.5009
Urban	Finance	734136768	698844.3928
Urban	Others	873520829	808067.3719
Rural	Others	92964084	650098.4895
Rural	Technology	285650551	1195190.5900
Suburban	Education	360273968	1310085.7018
Suburban	Finance	200720532	606406.4411
Urban	Education	863246553	809799.7683
Urban	Healthcare	151520241	704112.5655
Suburban	Technology	405519971	846597.0167
Rural	Healthcare	373683322	1190074.2739
Rural	Education	168754609	986870.2281
Rural	Finance	142025521	979486.3517
Suburban	Healthcare	536756083	943332.3076

This query determines the average and total income by occupation and location. It provides important insights into how incomes differ by profession and geography by combining income data from various employment roles and geographic locations. Understanding economic differences and adjusting company or policy initiatives accordingly are made easier with the use of that information.

7. Find the average Household_Size grouped by Type_of_Housing?



The screenshot shows a SQL query editor interface. The query is as follows:

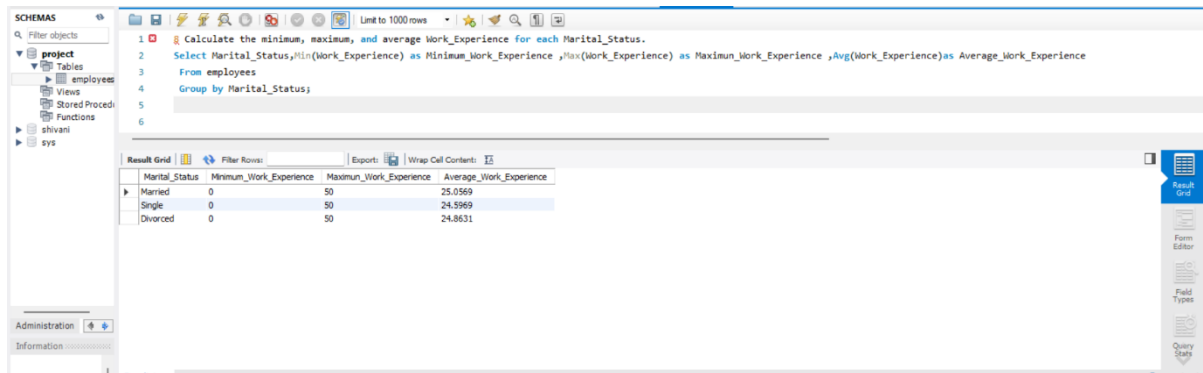
```
1 2 Find the average Household_Size grouped by Type_of_Housing.  
2 Select Type_of_Housing , Avg(Household_Size) as Average_Household_Size  
3 from employees  
4 group by Type_of_Housing;  
5
```

The results are displayed in a grid with the following data:

Type_of_Housing	Average_Household_Size
Apartment	3.9808
Single-family home	3.9721
Townhouse	4.0442

The average number of household members for each form of reflecting is calculated by this query. It provides information on family structures and living space efficiency by displaying the differences in household size across different housing categories.

8. Calculate the minimum, maximum, and average Work_Experience for each Marital_Status?



The screenshot shows a database query editor interface. The SQL query is as follows:

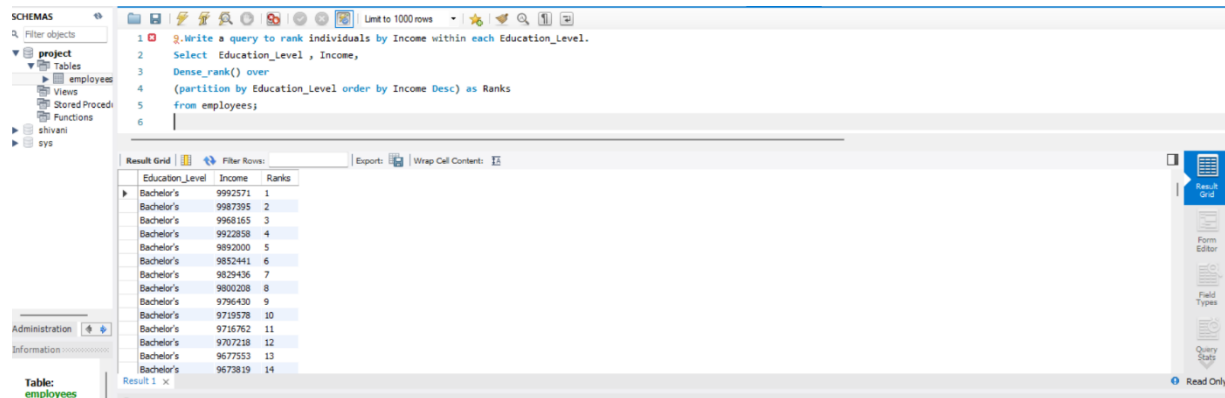
```
1 Calculate the minimum, maximum, and average Work_Experience for each Marital_Status.  
2 Select Marital_Status, Min(Work_Experience) as Minimum_Work_Experience, Max(Work_Experience) as Maximum_Work_Experience, Avg(Work_Experience) as Average_Work_Experience  
3 From employees  
4 Group by Marital_Status;
```

The results are displayed in a table with the following data:

Marital_Status	Minimum_Work_Experience	Maximum_Work_Experience	Average_Work_Experience
Married	0	50	25.0569
Single	0	50	24.5969
Divorced	0	50	24.8631

The minimum, maximum, and average work experience for each individual sorted by Marital_Status is determined by this query. It provides awareness on how job experience differs amongst various marital groups, which can help with research on demographic trends and career growth.

9. Write a query to rank individuals by Income within each Education_Level.



The screenshot shows a database query editor with a SQL query in the top pane and its results in the bottom pane. The query is designed to rank individuals by income within each education level using the DENSE_RANK function.

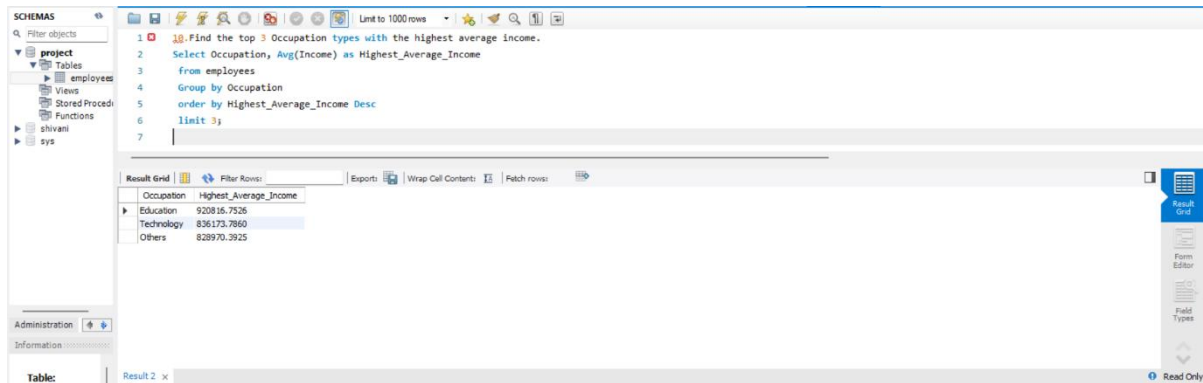
```
1 9. Write a query to rank individuals by Income within each Education_Level.  
2 Select Education_Level , Income,  
3 Dense_rank() over  
4 (partition by Education_Level order by Income Desc) as Ranks  
5 from employees;  
6
```

The results are displayed in a table with three columns: Education_Level, Income, and Ranks. The data shows 14 rows of results, all for the 'Bachelor's' education level, ranked by income in descending order.

Education_Level	Income	Ranks
Bachelor's	9992571	1
Bachelor's	9987395	2
Bachelor's	9968165	3
Bachelor's	9923858	4
Bachelor's	9892000	5
Bachelor's	9852441	6
Bachelor's	9829436	7
Bachelor's	9800208	8
Bachelor's	9796430	9
Bachelor's	9719578	10
Bachelor's	9716762	11
Bachelor's	9707218	12
Bachelor's	9677553	13
Bachelor's	9673819	14

People in each educational level are ranked by their income in this regard. By ranking each individual based on how much they earn as compared to those with the same educational background, it becomes easier to identify high earners and to understand how income is distributed among various educational groupings.

10. Find the top 3 Occupation types with the highest average income?



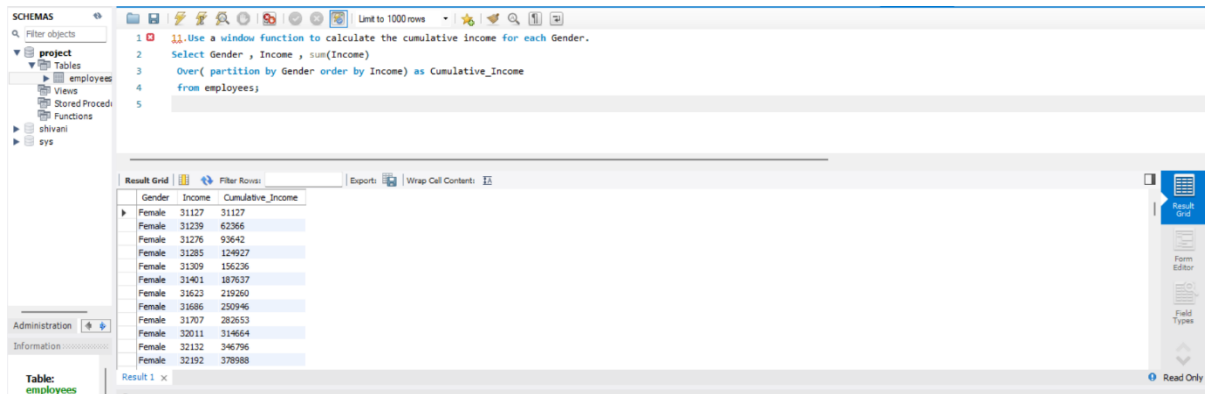
The screenshot shows a database query editor interface. On the left, a 'SCHEMAS' pane lists a 'project' database containing 'employees', 'Views', 'Stored Procedures', 'Functions', 'shivani', and 'sys'. The main editor contains a SQL query to find the top 3 occupations by average income. Below the query, the 'Result Grid' displays the results in a table format. The table has two columns: 'Occupation' and 'Highest_Average_Income'. The results are: Education (920816.7526), Technology (836173.7860), and Others (828970.3925). The interface also includes a toolbar with icons for query execution, a 'Filter Rows' button, and an 'Export' button. The status bar at the bottom indicates 'Table: Result 2' and 'Read Only'.

```
1 10. Find the top 3 Occupation types with the highest average income.  
2 Select Occupation, Avg(Income) as Highest_Average_Income  
3 from employees  
4 Group by Occupation  
5 order by Highest_Average_Income Desc  
6 limit 3;  
7
```

Occupation	Highest_Average_Income
Education	920816.7526
Technology	836173.7860
Others	828970.3925

By determining the mean salary for each occupation and ranking them appropriately, this query determines the top three occupations with the greatest average income. This research offers important insights into wage discrepancies across professions by highlighting the most expensive job categories within the dataset.

11. Use a window function to calculate the cumulative income for each Gender?

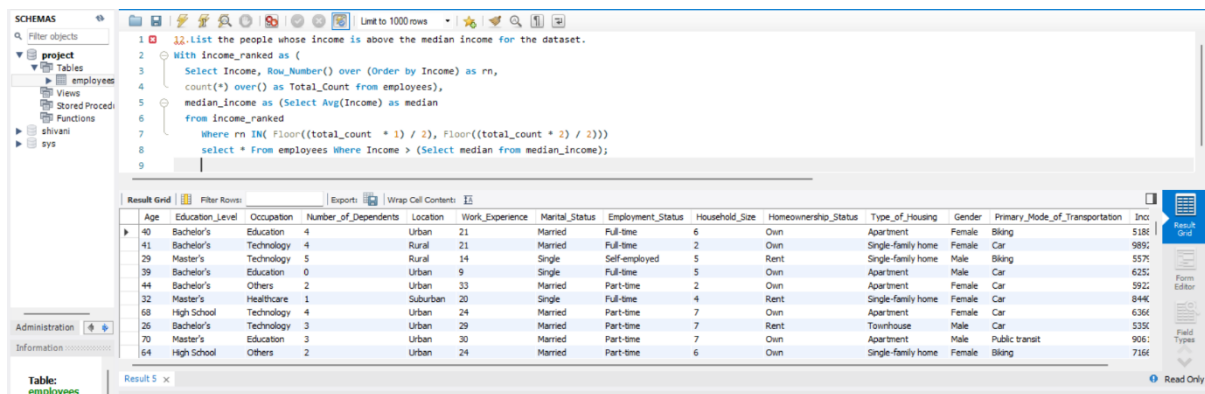


The screenshot shows a SQL IDE interface. On the left, a 'SCHEMAS' pane lists a database named 'project' with tables 'employees', 'views', 'stored_procedures', 'functions', 'shivani', and 'sys'. The main editor displays a SQL query:
1. 11. Use a window function to calculate the cumulative income for each Gender.
2. Select Gender, Income, sum(Income)
3. Over(partition by Gender order by Income) as Cumulative_Income
4. from employees;
5.
Below the query, a 'Result Grid' shows the output. It has columns 'Gender', 'Income', and 'Cumulative_Income'. The results are for 'Female' employees, ordered by income. The cumulative income is calculated as a running total of the individual incomes for each gender group.

Gender	Income	Cumulative_Income
Female	31127	31127
Female	31239	62366
Female	31276	93642
Female	31285	124927
Female	31209	156236
Female	31401	187637
Female	31623	219260
Female	31686	250946
Female	31707	282653
Female	32011	314664
Female	32132	346796
Female	32192	378988

This query determines each gender's cumulative income using a window function. It offers insight into the distribution and aggregation patterns of income between males and females by adding up the incomes in each gender group in a running total order.

12. List the people whose income is above the median income for the dataset?



```
1 12. List the people whose income is above the median income for the dataset.
2
3 With income_ranked as (
4   select Income, Row_Number() over (Order by Income) as rn,
5   count(*) over() as Total_Count from employees),
6   median_income as (Select Avg(Income) as median
7   from income_ranked
8   where rn IN( Floor((total_count * 1) / 2), Floor((total_count * 2) / 2)))
9   select * from employees Where Income > (Select median from median_income);
```

Age	Education_Level	Occupation	Number_of_Dependents	Location	Work_Experience	Marital_Status	Employment_Status	Household_Size	Homeownership_Status	Type_of_Housing	Gender	Primary_Mode_of_Transportation	Income
40	Bachelor's	Education	4	Urban	21	Married	Full-time	6	Own	Apartment	Female	Biking	5188
41	Bachelor's	Technology	4	Rural	21	Married	Full-time	2	Own	Single-family home	Female	Car	9890
29	Master's	Technology	5	Rural	14	Single	Self-employed	5	Rent	Single-family home	Male	Biking	5575
39	Bachelor's	Education	0	Urban	9	Single	Full-time	5	Own	Apartment	Male	Car	6252
44	Bachelor's	Others	2	Urban	33	Married	Part-time	2	Own	Apartment	Female	Car	3922
32	Master's	Healthcare	1	Suburban	20	Single	Full-time	4	Rent	Single-family home	Female	Car	8946
68	High School	Technology	4	Urban	24	Married	Part-time	7	Own	Apartment	Female	Car	6366
26	Bachelor's	Technology	3	Urban	29	Married	Part-time	7	Rent	Townhouse	Male	Car	5330
70	Master's	Education	3	Urban	30	Married	Part-time	7	Own	Apartment	Male	Public transit	9061
64	High School	Others	2	Urban	24	Married	Part-time	6	Own	Single-family home	Female	Biking	7166

The list of people whose income is higher than the dataset's median income is gathered by this query. It identifies individuals in the upper half of the income range and highlights the higher earners by comparing each person's income to the median value. This provides insights into income distribution.

CONCLUSION

- In order to find trends in income, employment, education, and living conditions, I Investigated important socioeconomic and demographic factors using a SQL-based analysis of the employee's table. I was able to learn more about lifestyle characteristics, household dynamics, and geographic patterns by composing questions.
- Important Lessons Learnt: SQL made structured and effective data exploration possible.
- I gained a deeper knowledge of data linkages through querying.
- It will assist us in understanding employee's behaviour.

THANKYOU