

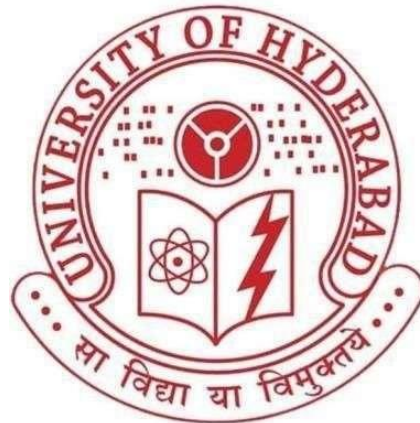
AN INTERNSHIP REPORT ON

“SOCIO ECONOMIC ANALYSIS(SQL)”

SUBMITTED BY: GUJJUKA MALLAMMA

**MASTER OF BUSINESS ADMINISTRATION
(BUSINESS ANALYTICS)**

ROLL NO. 23MBMB04



SCHOOL OF MANAGEMENT STUDIES

UNIVERSITY OF HYDERABAD

ACKNOWLEDGEMENT

I am happy to express my gratitude for completing the task-1 of the internship under the guidance of Soulville .The passion, goal orientation and timely advice motivated me to continue working with relentless energy.

I am grateful for the strong academic skills instilled in me by successfully completing my work with passion and dedication, and strict discipline.

I would like to thank **you for Soulville.Tech** who gave me the opportunity to do a good internship using Mysql workbench. I think this is an important time for my career development. I will try to use the knowledge and skills I have gained in the best way and will continue to strive to improve them to achieve my professional goals.

Name-Gujjuka Mallamma

1.INTRODUCTION

The dataset comprises **10,000 records** and **14 attributes**, providing detailed demographic and socioeconomic information about individuals. The key variables include age, education, occupation, dependents, income, and lifestyle factors like housing and transportation. The purpose of this analysis is to understand the general structure and key patterns within the data.

- **Age**

The average age is approximately **44 years**, with values ranging from **18 to 70 years**. This indicates a diverse working population covering young adults to near-retirement individuals.

- **Education Level**

Includes four categories: High School, Bachelor's, Master's, and Doctorate. The most common qualification is a Bachelor's degree, suggesting a moderately educated population.

- **Occupation**

Comprises five primary groups: **Technology, Healthcare, Finance, Others, and Education**. **Healthcare** and **Technology** dominate, implying a skill-oriented or service-based workforce.

- **Number of Dependents**

The average is around **2.5**, with a maximum of **5**. This shows most individuals have moderate family responsibilities.

- **Location**

Consists of **Urban, Suburban, and Rural** areas. **Urban residents** make up a significant portion (~70%), pointing to urban-centric data.

- **Work Experience**

Ranges from **0 to 50 years**, with an average around **25 years**. Indicates a balanced mix of early-career and experienced individuals.

- **Marital Status**

Categories: **Married**, **Single**, and **Divorced**. A majority are **Married**, often correlating with higher dependents and income stability.

- **Employment Status**

Includes **Full-time**, **Part-time**, and **Self-employed**. Over 50% are in **Full-time employment**, suggesting financial stability.

- **Household Size**

Ranges from **1 to 7**, with an average of around **4 members per household**. This aligns with traditional family setups.

- **Homeownership Status**

Divided into **Own** and **Rent**. More than 60% **own their homes**, indicating long-term residential stability.

- **Type of Housing**

Includes **Apartment**, **Townhouse**, and **Single-family home**. **Single-family homes** are slightly more prevalent, reflecting suburban or affluent urban demographics.

- **Gender**

Roughly evenly split between **Male** and **Female**. No gender bias observed in data distribution.

- **Primary Mode of Transportation**

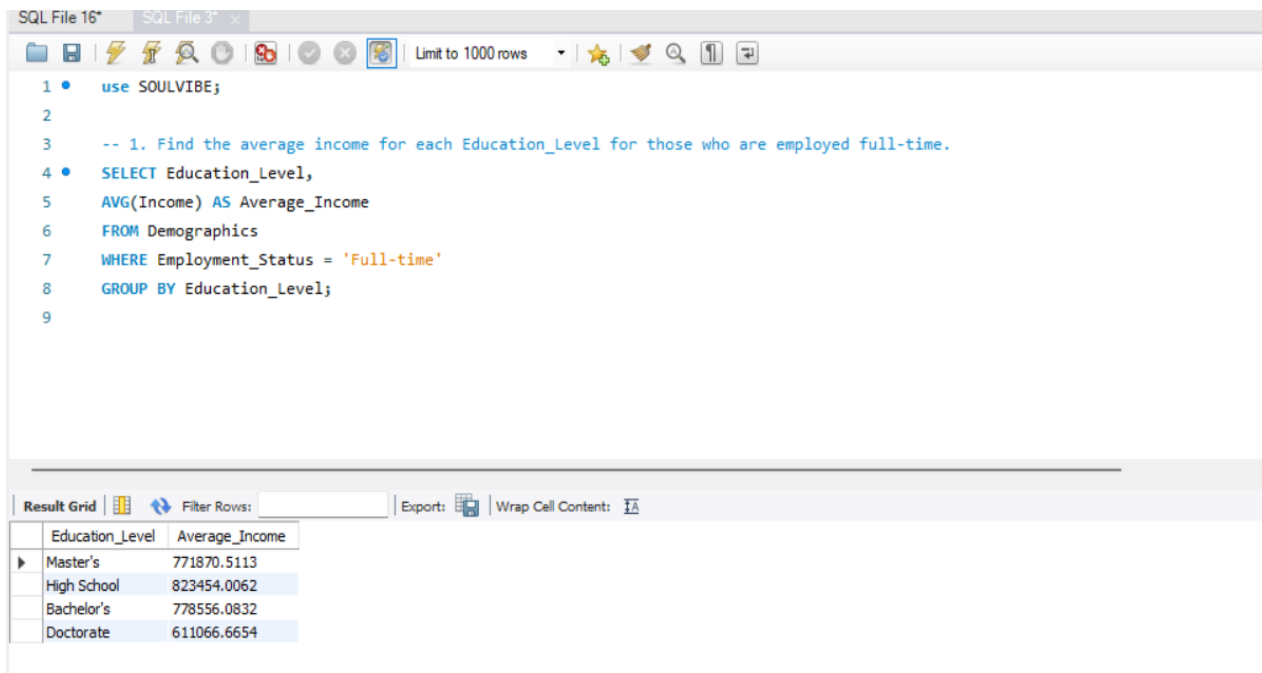
Categories: **Car**, **Public Transit**, **Walking**, and **Biking**. **Public Transit** is the most common, indicating urban dependence on shared transport.

- **Income**

Varies widely from ₹31,044 to ₹99,92,571 with an average of around ₹8.16 lakhs. High standard deviation implies income inequality; a few high earners skew the average.

TASK 1: SQL QUERIES

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



The screenshot shows a SQL IDE with two tabs: 'SQL File 16*' and 'SQL File 3*'. The query editor contains the following SQL code:

```
1 • use SOULVIBE;
2
3 -- 1. Find the average income for each Education_Level for those who are employed full-time.
4 • SELECT Education_Level,
5       AVG(Income) AS Average_Income
6   FROM Demographics
7  WHERE Employment_Status = 'Full-time'
8  GROUP BY Education_Level;
9
```

The results pane at the bottom shows a table with two columns: 'Education_Level' and 'Average_Income'. The data is as follows:

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

Explanation :

This query calculates the average income for individuals who are employed full-time, grouped by their education level.

It helps understand how education level influences earning potential in full-time roles.

Each education level is listed with its corresponding average income, showing trends across education categories.

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

```
11 -- 2. Retrieve the top 5 highest earning individuals and their details.
12 • SELECT *
13 FROM Demographics
14 ORDER BY Income DESC
15 LIMIT 5;
16
```

	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
▶	24	Bachelor's	Healthcare	4	Urban	30	Married	Part-time	2	Own	Apartment	Female	Public transit
	33	Bachelor's	Healthcare	3	Urban	31	Single	Full-time	1	Rent	Townhouse	Male	Public transit
	66	Master's	Healthcare	3	Rural	48	Married	Full-time	4	Rent	Single-family home	Male	Public transit
	62	Bachelor's	Others	3	Urban	1	Married	Full-time	1	Own	Apartment	Female	Biking
	57	Bachelor's	Technology	2	Suburban	4	Married	Full-time	7	Own	Single-family home	Male	Public transit

Explanation :

This query retrieves the details of the top 5 earners based on income.

Useful to identify highest earning profiles and their attributes such as education, occupation, or gender.

Displays key demographic and employment details of the top 5 income earners.

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

```
17 -- 3.Count how many people in each Occupation have more than 2 dependents and own a house.
18 • SELECT Occupation, COUNT(*) AS People_Count
19 FROM Demographics
20 WHERE Number_of_Dependents > 2 AND Homeownership_Status = 'Own'
21 GROUP BY Occupation;
22
23
```

	Occupation	People_Count
▶	Technology	725
	Finance	463
	Education	507
	Healthcare	906
	Others	478

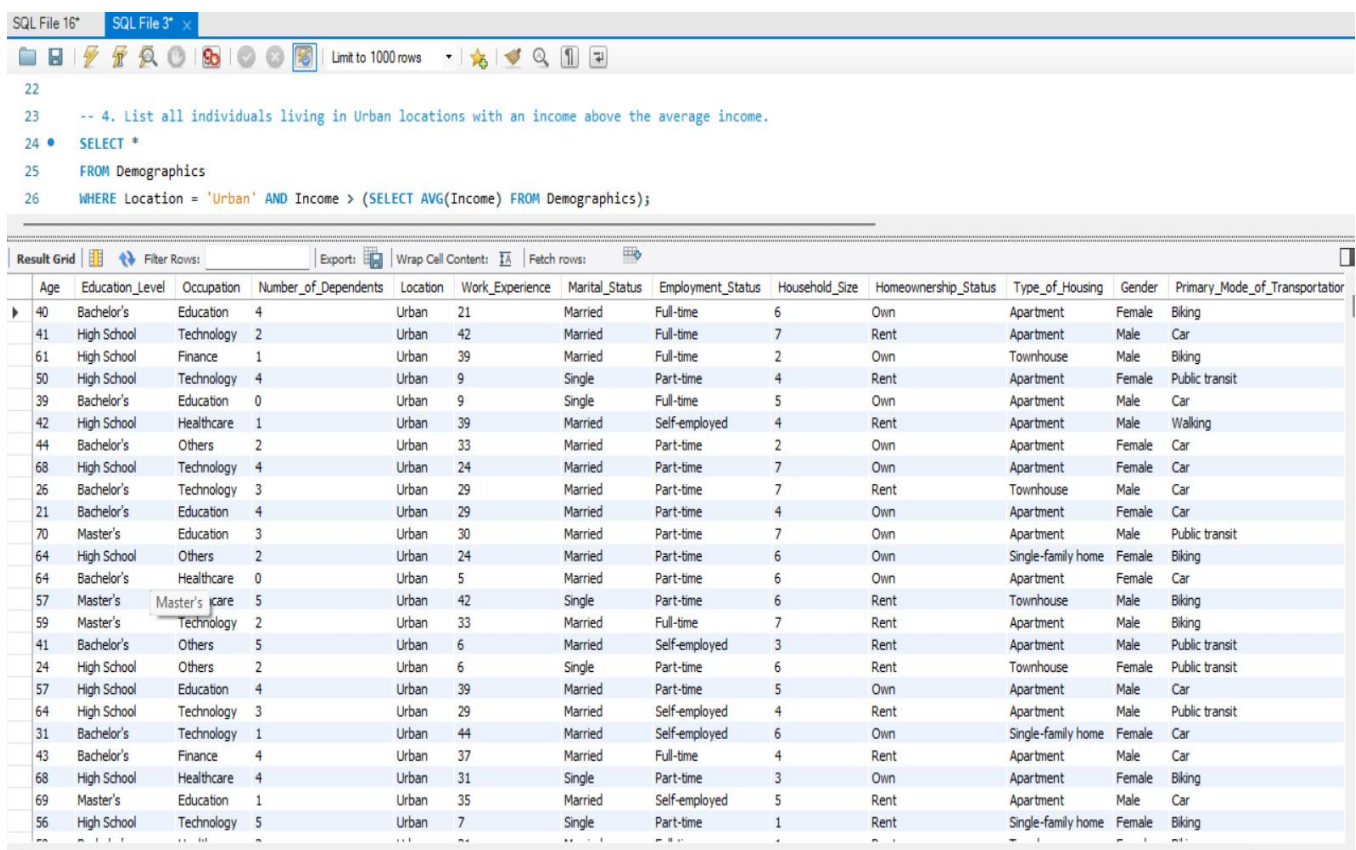
Explanation:

The above query counts how many individuals in each occupation have more than 2 dependents and own a house.

It is useful for understanding family and housing responsibilities in various professions.

The output table shows occupation-wise count, highlighting jobs with more household responsibilities.

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



The screenshot shows a SQL IDE with a query editor and a results grid. The query editor contains the following SQL code:

```
22
23 -- 4. List all individuals living in Urban locations with an income above the average income.
24 • SELECT *
25 FROM Demographics
26 WHERE Location = 'Urban' AND Income > (SELECT AVG(Income) FROM Demographics);
```




The results grid displays a table with 13 columns: Age, Education_Level, Occupation, Number_of_Dependents, Location, Work_Experience, Marital_Status, Employment_Status, Household_Size, Homeownership_Status, Type_of_Housing, Gender, and Primary_Mode_of_Transportation. The table contains 30 rows of data, all filtered for 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
▶	40	Bachelor's	Education	4	Urban	21	Married	Full-time	6	Own	Apartment	Female	Biking
	41	High School	Technology	2	Urban	42	Married	Full-time	7	Rent	Apartment	Male	Car
	61	High School	Finance	1	Urban	39	Married	Full-time	2	Own	Townhouse	Male	Biking
	50	High School	Technology	4	Urban	9	Single	Part-time	4	Rent	Apartment	Female	Public transit
	39	Bachelor's	Education	0	Urban	9	Single	Full-time	5	Own	Apartment	Male	Car
	42	High School	Healthcare	1	Urban	39	Married	Self-employed	4	Rent	Apartment	Male	Walking
	44	Bachelor's	Others	2	Urban	33	Married	Part-time	2	Own	Apartment	Female	Car
	68	High School	Technology	4	Urban	24	Married	Part-time	7	Own	Apartment	Female	Car
	26	Bachelor's	Technology	3	Urban	29	Married	Part-time	7	Rent	Townhouse	Male	Car
	21	Bachelor's	Education	4	Urban	29	Married	Part-time	4	Own	Apartment	Female	Car
	70	Master's	Education	3	Urban	30	Married	Part-time	7	Own	Apartment	Male	Public transit
	64	High School	Others	2	Urban	24	Married	Part-time	6	Own	Single-family home	Female	Biking
	64	Bachelor's	Healthcare	0	Urban	5	Married	Part-time	6	Own	Apartment	Female	Car
	57	Master's	Healthcare	5	Urban	42	Single	Part-time	6	Rent	Townhouse	Male	Biking
	59	Master's	Technology	2	Urban	33	Married	Full-time	7	Rent	Apartment	Male	Biking
	41	Bachelor's	Others	5	Urban	6	Married	Self-employed	3	Rent	Apartment	Male	Public transit
	24	High School	Others	2	Urban	6	Single	Part-time	6	Rent	Townhouse	Female	Public transit
	57	High School	Education	4	Urban	39	Married	Part-time	5	Own	Apartment	Male	Car
	64	High School	Technology	3	Urban	29	Married	Self-employed	4	Rent	Apartment	Male	Public transit
	31	Bachelor's	Technology	1	Urban	44	Married	Self-employed	6	Own	Single-family home	Female	Car
	43	Bachelor's	Finance	4	Urban	37	Married	Full-time	4	Rent	Apartment	Male	Car
	68	High School	Healthcare	4	Urban	31	Single	Part-time	3	Own	Apartment	Female	Biking
	69	Master's	Education	1	Urban	35	Married	Self-employed	5	Rent	Apartment	Male	Car
	56	High School	Technology	5	Urban	7	Single	Part-time	1	Rent	Single-family home	Female	Biking

The above query displays individual details filtered by location and income condition.

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

```
28      -- 5. Identify how many males and females are in each Employment_Status.
29 •    SELECT Employment_Status, Gender, COUNT(*) AS Count
30      FROM Demographics
31      GROUP BY Employment_Status, Gender;
32
33
34
```

Result Grid  Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 			
	Employment_Status	Gender	Count
▶	Full-time	Male	2564
	Full-time	Female	2440
	Self-employed	Male	1040
	Part-time	Female	1497
	Part-time	Male	1519
	Self-employed	Female	940

The above query Identifies the count of males and females in each employment status category.

The insights are useful to analyze gender representation across work statuses.

It shows the Gender distribution for categories like full-time, part-time, unemployed, etc.

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

The query shows both total and average income grouped by each location and occupation.





It Identifies which job roles and areas are most lucrative.

Show the output as the Combined analysis of geography and profession with respect to income.


```

33      -- 6. What is the total and average income by Location and Occupation?
34  •    SELECT Location, Occupation,
35          SUM(Income) AS Total_Income,
36          AVG(Income) AS Average_Income
37  FROM Demographics
38  GROUP BY Location, Occupation;
39
40

```

Result Grid |   Filter Rows: | Export:  | Wrap Cell Content: 





	Location	Occupation	Total_Income	Average_Income
▶	Urban	Technology	1321499781	782415.5009
	Urban	Finance	734136768	699844.3928
	Urban	Others	873520829	808067.3719
	Rural	Others	92964084	650098.4895
	Rural	Technology	285650551	1195190.5900
	Suburban	Education	360273568	1310085.7018
	Suburban	Finance	200720532	606406.4411
	Urban	Education	863246553	809799.7683
	Urban	Healthcare	1515250241	704112.5655
	Suburban	Technology	405519971	846597.0167
	Rural	Healthcare	373683322	1190074.2739
	Rural	Education	168754809	986870.2281
	Rural	Finance	142025521	979486.3517
	Suburban	Healthcare	536756083	943332.3076
	Suburban	Others	294379054	991175.2660

7. Find the average Household_Size grouped by Type_of_Housing.

```

40      -- 7. Find the average Household_Size grouped by Type_of_Housing.
41  •    SELECT Type_of_Housing, AVG(Household_Size) AS Avg_Household_Size
42  FROM Demographics
43  GROUP BY Type_of_Housing;
44
45

```

Result Grid |   Filter Rows: | Export:  | Wrap Cell Content: 

	Type_of_Housing	Avg_Household_Size
▶	Apartment	3.9808
	Single-family home	3.9721
	Townhouse	4.0442




The above query calculates the average household size for each type of housing.

It is used to give an idea of occupancy trends across different housing setups.

Provides the output as the housing types like apartments or independent homes are listed with average sizes.

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

```
45  -- 8. Calculate the minimum, maximum, and average Work_Experience for each Marital_Status.
46  •  SELECT Marital_Status,
47         MIN(Work_Experience) AS Min_Experience,
48         MAX(Work_Experience) AS Max_Experience,
49         AVG(Work_Experience) AS Avg_Experience
50  FROM Demographics
51  GROUP BY Marital_Status;
52
```

Result Grid  Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 				
	Marital_Status	Min_Experience	Max_Experience	Avg_Experience
▶	Married	0	50	25.0569
	Single	0	50	24.5969
	Divorced	0	50	24.8631

Finds minimum, maximum, and average work experience grouped by marital status.

It Correlates marital life stages with career experience.

Output Shows work experience range and average for each marital category.

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

```
53 -- 9. Write a query to rank individuals by Income within each Education_Level.
54 • SELECT *,
55     RANK() OVER (PARTITION BY Education_Level ORDER BY Income DESC) AS Income_Rank
56 FROM Demographics;
57
58
```

Result Grid														Filter Rows:		Export:	Wrap Cell Content:
	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_Transportat				
▶	24	Bachelor's	Healthcare	4	Urban	30	Married	Part-time	2	Own	Apartment	Female	Public transit				
	33	Bachelor's	Healthcare	3	Urban	31	Single	Full-time	1	Rent	Townhouse	Male	Public transit				
	62	Bachelor's	Others	3	Urban	1	Married	Full-time	1	Own	Apartment	Female	Biking				
	57	Bachelor's	Technology	2	Suburban	4	Married	Full-time	7	Own	Single-family home	Male	Public transit				
	41	Bachelor's	Technology	4	Rural	21	Married	Full-time	2	Own	Single-family home	Female	Car				
	37	Bachelor's	Healthcare	3	Urban	29	Divorced	Full-time	3	Rent	Townhouse	Male	Car				
	54	Bachelor's	Finance	4	Urban	50	Single	Full-time	6	Rent	Apartment	Female	Car				
	19	Bachelor's	Technology	5	Urban	20	Married	Part-time	3	Own	Apartment	Female	Biking				
	64	Bachelor's	Finance	2	Urban	4	Married	Self-employed	2	Rent	Townhouse	Female	Public transit				
	62	Bachelor's	Technology	3	Suburban	5	Married	Full-time	2	Own	Single-family home	Female	Public transit				
	38	Bachelor's	Healthcare	0	Urban	31	Married	Full-time	7	Rent	Single-family home	Male	Biking				
	69	Bachelor's	Education	0	Suburban	49	Single	Part-time	6	Own	Apartment	Female	Biking				
	65	Bachelor's	Others	5	Urban	14	Married	Full-time	7	Own	Single-family home	Male	Public transit				
	24	Bachelor's	Technology	0	Urban	1	Single	Full-time	5	Own	Single-family home	Male	Public transit				
	67	Bachelor's	Healthcare	1	Urban	14	Married	Part-time	2	Own	Apartment	Female	Public transit				
	57	Bachelor's	Healthcare	0	Urban	31	Single	Full-time	5	Rent	Single-family home	Female	Public transit				
	26	Bachelor's	Healthcare	3	Suburban	27	Single	Self-employed	7	Rent	Single-family home	Male	Car				
	45	Bachelor's	Healthcare	0	Rural	35	Single	Part-time	5	Own	Single-family home	Female	Public transit				
	55	Bachelor's	Education	2	Rural	44	Single	Full-time	7	Own	Single-family home	Male	Public transit				
	50	Bachelor's	Healthcare	5	Suburban	29	Married	Full-time	2	Own	Apartment	Female	Public transit				
	47	Bachelor's	Healthcare	1	Urban	37	Single	Part-time	3	Rent	Townhouse	Female	Car				

Ranks individuals based on income within their respective education levels using a window function.

It Highlights top earners per education category.

Output shows as Individuals are assigned a rank within their education group based on income.

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

```
58 -- 10 Find the top 3 Occupation types with the highest average income.
59 • SELECT Occupation, AVG(Income) AS Avg_Income
60 FROM Demographics
61 GROUP BY Occupation
62 ORDER BY Avg_Income DESC
63 LIMIT 3;
64
```

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

Fetch rows:

	Occupation	Avg_Income
▶	Education	920816.7526
	Technology	836173.7860
	Others	828970.3925

Retrieves the top 3 occupation types having the highest average income.

It Identifies the most financially rewarding job roles.

The out Displays occupations sorted by descending average income, limited to top 3.

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

```
65 -- 11. Use a window function to calculate the cumulative income for each Gender.
66 • SELECT Gender, Income,
67        SUM(Income) OVER (PARTITION BY Gender ORDER BY Income) AS Cumulative_Income
68 FROM Demographics;
69
70
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Gender	Income	Cumulative_Income	
Female	31127	31127	
Female	31239	62366	
Female	31276	93642	
Female	31285	124927	
Female	31309	156236	
Female	31401	187637	
Female	31623	219260	
Female	31686	250946	
Female	31707	282653	
Female	32011	314664	
Female	32132	346796	
Female	32192	378988	
Female	32230	411218	
Female	32450	443668	
Female	32845	476513	
Female	32959	509472	
Female	33074	542546	

Uses a window function to compute the running total of income for each gender.

Insight: Shows cumulative earnings progression within gender groups.

Output: Displays income aggregation per gender sorted appropriately.

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

SQL File 16*

SQL File 3*

Limit to 1000 rows

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

71 • WITH IncomeRanked AS (

72 SELECT Income,

73 PERCENT_RANK() OVER (ORDER BY Income) AS pr

74 FROM Demographics

75),

76 Median AS (

77 SELECT MAX(Income) AS Median_Income

78 FROM IncomeRanked

79 WHERE pr <= 0.5

80)

81 SELECT *

82 FROM Demographics

83 WHERE Income > (SELECT Median_Income FROM Median);

84

85

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

Fetch rows:

	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_Transport
▶	69	High School	Finance	0	Urban	4	Single	Full-time	7	Own	Apartment	Male	Biking
	32	High School	Others	2	Urban	Urban	Married	Full-time	1	Own	Apartment	Female	Car
	60	Bachelor's	Finance	3	Urban	15	Married	Self-employed	4	Own	Townhouse	Male	Walking
	25	High School	Others	1	Rural	6	Single	Part-time	1	Rent	Townhouse	Female	Car
	38	Master's	Technology	3	Rural	20	Married	Full-time	4	Rent	Single-family home	Male	Public transit
	40	Bachelor's	Education	4	Urban	21	Married	Full-time	6	Own	Apartment	Female	Biking
	41	Bachelor's	Technology	4	Rural	21	Married	Full-time	2	Own	Single-family home	Female	Car
	53	Master's	Technology	0	Suburban	32	Married	Full-time	3	Rent	Single-family home	Female	Public transit
	57	Bachelor's	Healthcare	4	Rural	38	Married	Self-employed	6	Rent	Apartment	Male	Car
	41	High School	Technology	2	Urban	42	Married	Full-time	7	Rent	Apartment	Male	Car
	39	High School	Technology	2	Rural	17	Single	Self-employed	3	Own	Apartment	Female	Public transit

Query Goal: Lists all people whose income exceeds the median income of the entire dataset.

Insight: Highlights higher-income individuals and their characteristics.

Output: Filters out the top half earners based on calculated median income.

CONCLUSION

This dataset reflects a balanced cross-section of urban and suburban working adults with varied education levels and family responsibilities. The data can be valuable for **predictive modeling, segmentation, policy making, or targeted marketing** based on demographic and economic factors.

13