

Internship program: Soulvibe.Tech



“Employee Data Analysis using SQL”

Batch SVT/DAINT/2025/06/B09

➤ INTRODUCTION ◀

Overview of objective



In this task, I was asked to analyze Income of employee data using SQL. The objective was to derive meaningful business insights by writing queries to Age, Education Level, Occupation, Income, Work experience etc., This helps in understanding identify complex relationships, and uncover meaningful insights. I used SQL to explore meaningful insights and complex relationships specific questions about the dataset.

1 : Find the average income for each Education Level for those who are employed full-time.

```
SELECT
    Education_Level,
    AVG(Income) AS Average_Income, Employment_Status
FROM
    project
WHERE
    Employment_Status = 'Full-Time'
GROUP BY
    Education_Level;
```

100% 1:24

Result Grid



Filter Rows:

Search

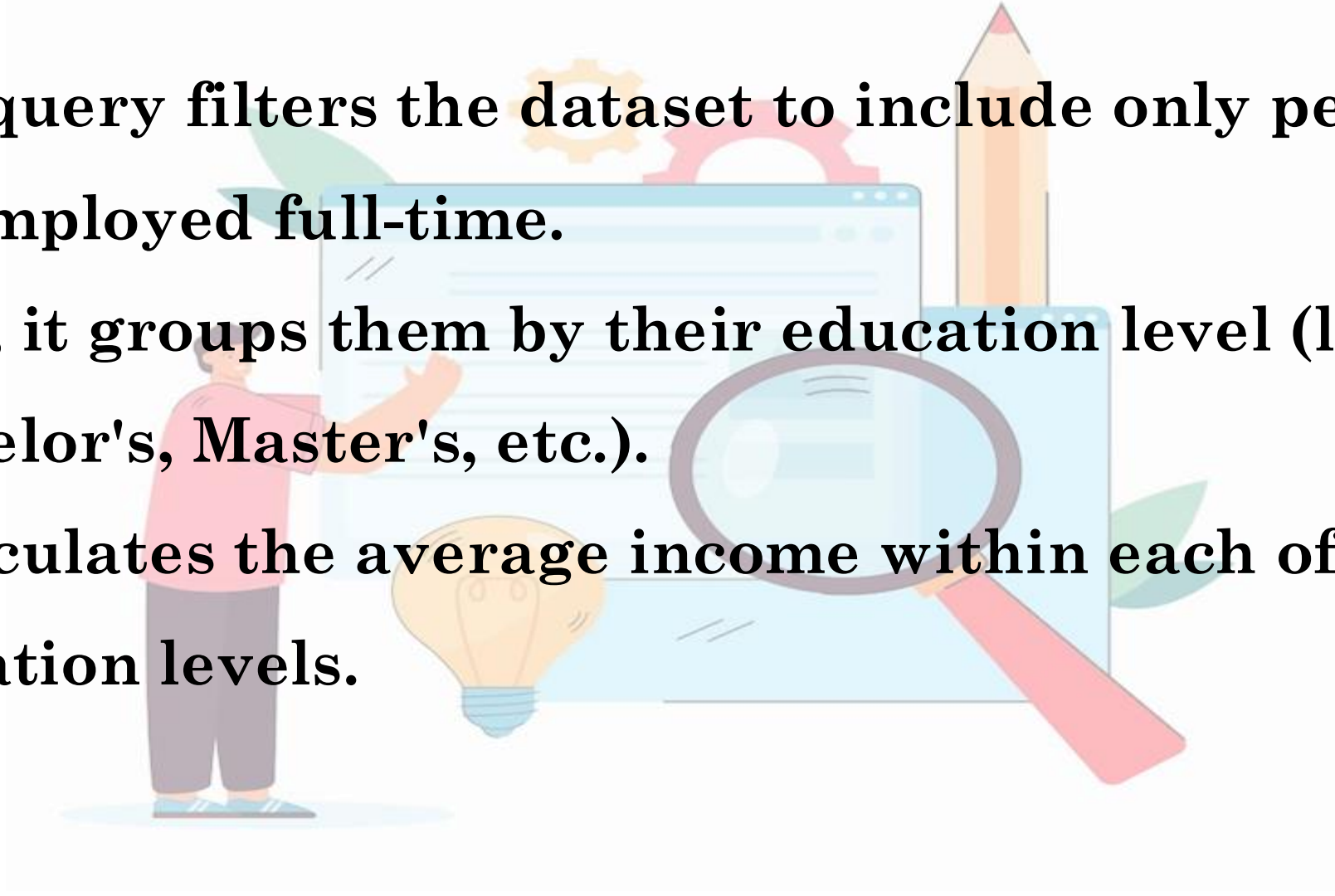
Export:



Education_Level	Average_Income	Employment_Stat...	
Master's	771870.5113	Full-time	
High School	823454.0062	Full-time	
Bachelor's	778556.0832	Full-time	
Doctorate	611066.6654	Full-time	

What this query does ?

- This query filters the dataset to include only people who are employed full-time.
- Then, it groups them by their education level (like Bachelor's, Master's, etc.).
- It calculates the average income within each of those education levels.



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



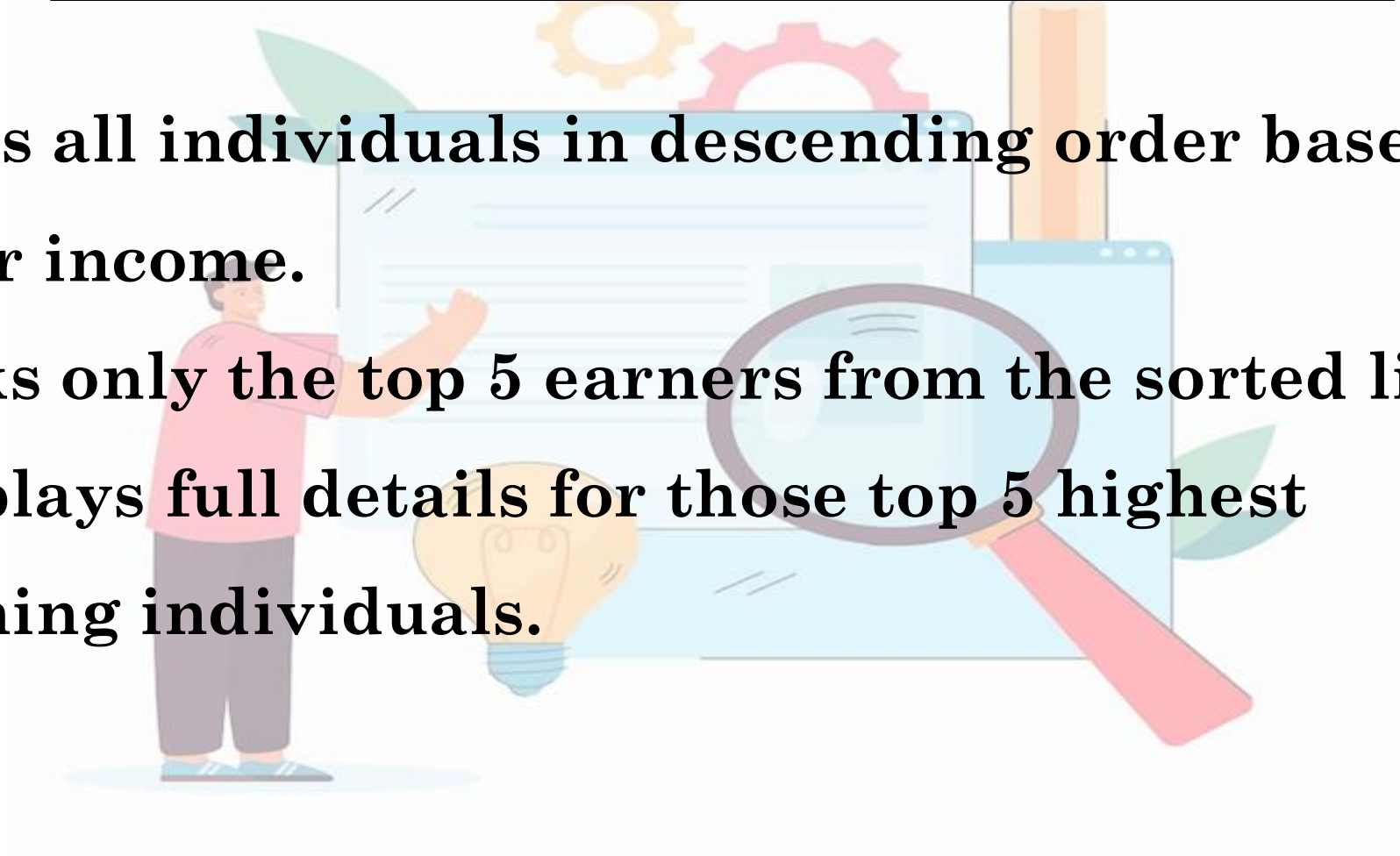
```
16
17 • select * FROM PROJECT
18 ORDER BY INCOME DESC
19 LIMIT 5;
20
```

Age	Education_Level	Occupation	Number_of_Dependen...	Location	Work_Experience	Marital_Stat...	Employment_Stat...	Household_Size	Homeownership_Stat...
24	Bachelor's	Healthcare	4	Urban	30	Married	Part-time	2	Own
33	Bachelor's	Healthcare	3	Urban	31	Single	Full-time	1	Rent
66	Master's	Healthcare	3	Rural	48	Married	Full-time	4	Rent
62	Bachelor's	Others	3	Urban	1	Married	Full-time	1	Own
57	Bachelor's	Technology	2	Suburban	4	Married	Full-time	7	Own

Type_of_Housing	Gender	Primary_Mode_of_Transportat...	Income
Apartment	Female	Public transit	9992571
Townhouse	Male	Public transit	9987395
Single-family home	Male	Public transit	9979438
Apartment	Female	Biking	9968165
Single-family home	Male	Public transit	9922858

What this query does ?

- . **Sorts all individuals in descending order based on their income.**
- . **Picks only the top 5 earners from the sorted list.**
- . **Displays full details for those top 5 highest earning individuals.**



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

```
34 SELECT Occupation, COUNT(*) AS EMP_COUNT, Homeownership_Status as house_type
35 FROM project
36 where Number_of_Dependents > 2
37 and Homeownership_Status = 'Own'
38 group by Occupation;
```

22631.2861.231204-0538.23H2_NI_RELEASE_
SVC_REFRESH_CLIENTCONSUMER_RET_A64F
RE_en-us.iso

00%

1:42

Result Grid



Filter Rows:

Search

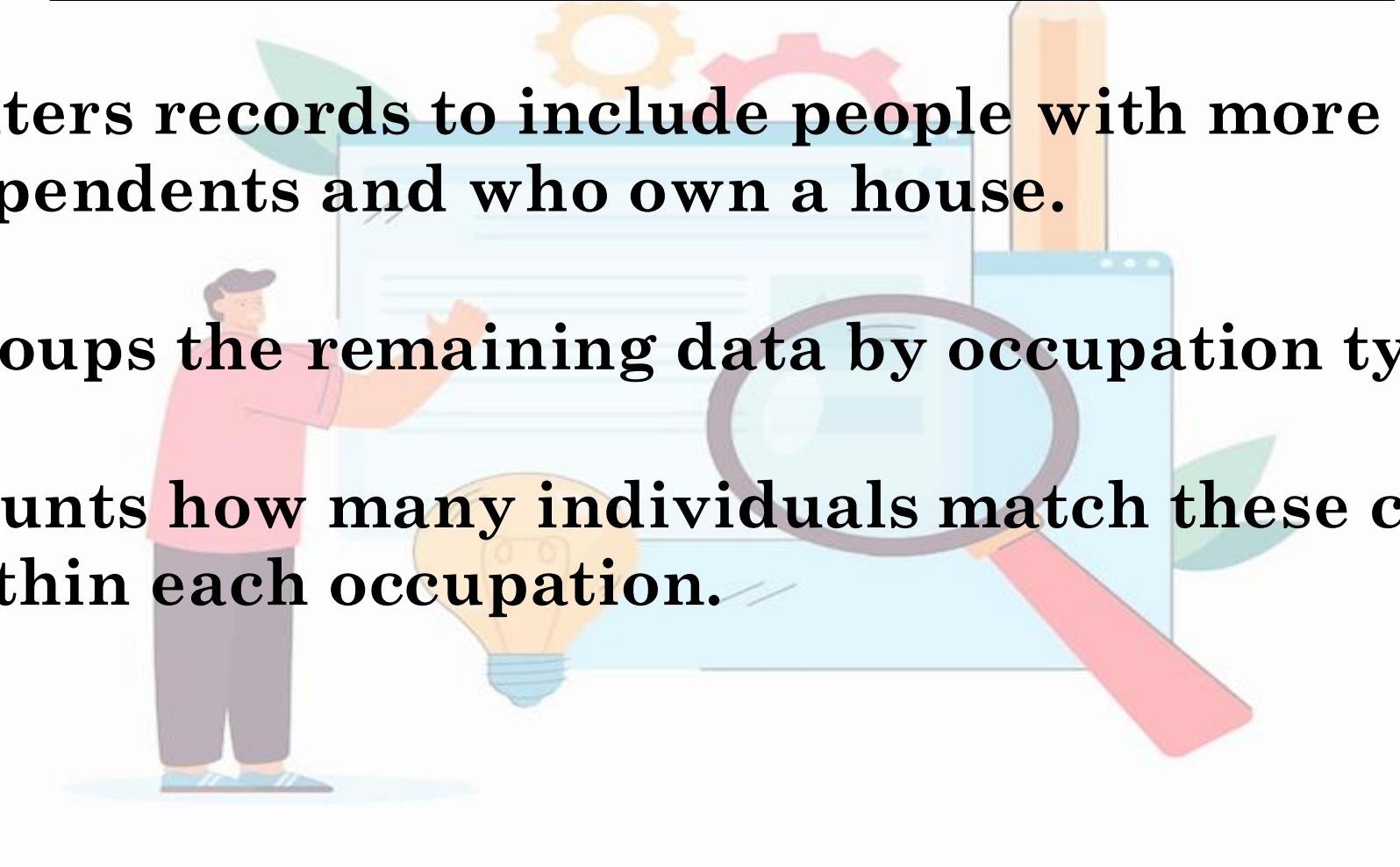
Export:



Occupation	EMP_COUNT	house_type	
Technology	725	Own	
Finance	463	Own	
Education	507	Own	
Healthcare	906	Own	
Others	478	Own	

What this query does ?

- **Filters records to include people with more than 2 dependents and who own a house.**
- **Groups the remaining data by occupation type.**
- **Counts how many individuals match these conditions within each occupation.**



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

```
47 SELECT Location, Income, Occupation, Gender
48 FROM project
49 WHERE Location = 'Urban'
50 AND Income > (
51     SELECT AVG(Income)
52     FROM project
53     WHERE Location = 'Urban'
54 )
55 ORDER BY Income Desc;
56
```

00%

4:54

Result Grid



Filter Rows:

Search

Export:



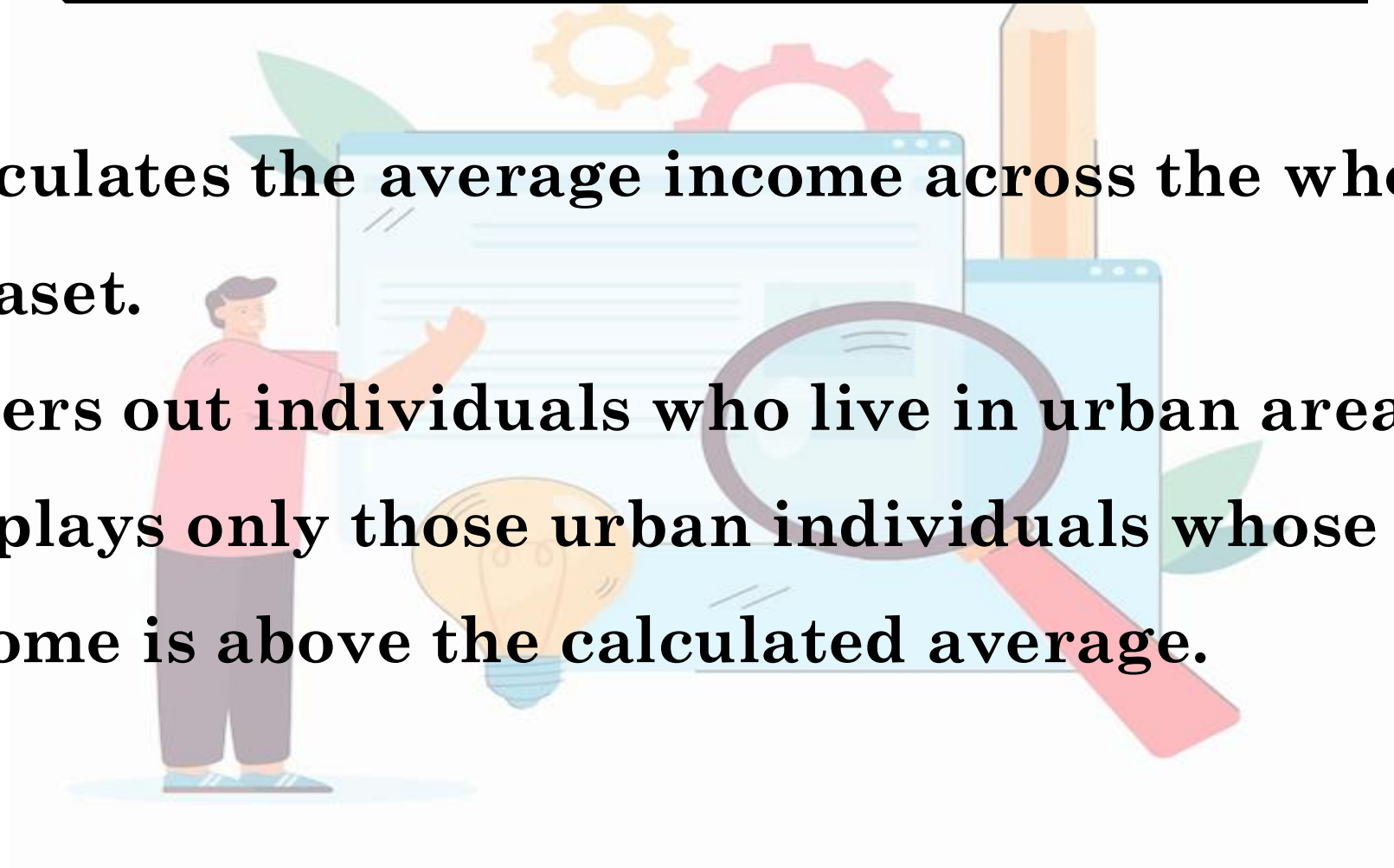
Fetch rows:



	Location	Income	Occupation	Gender	
	Urban	9992571	Healthcare	Female	
	Urban	9987395	Healthcare	Male	
	Urban	9968165	Others	Female	
	Urban	9914893	Healthcare	Female	
	Urban	9871131	Education	Female	
	Urban	9859518	Healthcare	Male	
	Urban	9852441	Healthcare	Male	
	Urban	9829544	Healthcare	Male	
	Urban	9829436	Finance	Female	
	Urban	9800208	Technology	Female	
	Urban	9796430	Finance	Female	
	Urban	9795312	Technology	Female	
	Urban	9732313	Education	Female	
	Urban	9716762	Healthcare	Male	
	Urban	9686527	Finance	Female	
	Urban	9677553	Others	Male	

What this query does ?

- **Calculates the average income across the whole dataset.**
- **Filters out individuals who live in urban areas.**
- **Displays only those urban individuals whose income is above the calculated average.**



5 : Identify how many males and females are in each Employment Status.

```
66
67 #Identify how many males and females are in each Employment Status.
68 * SELECT Employment_Status,Gender,count(*) as "Gender_count"
69 FROM project
70 group by Employment_Status,Gender
71 order by Employment_Status,Gender Desc;
```

00% 40:71

Result Grid



Filter Rows:



Search

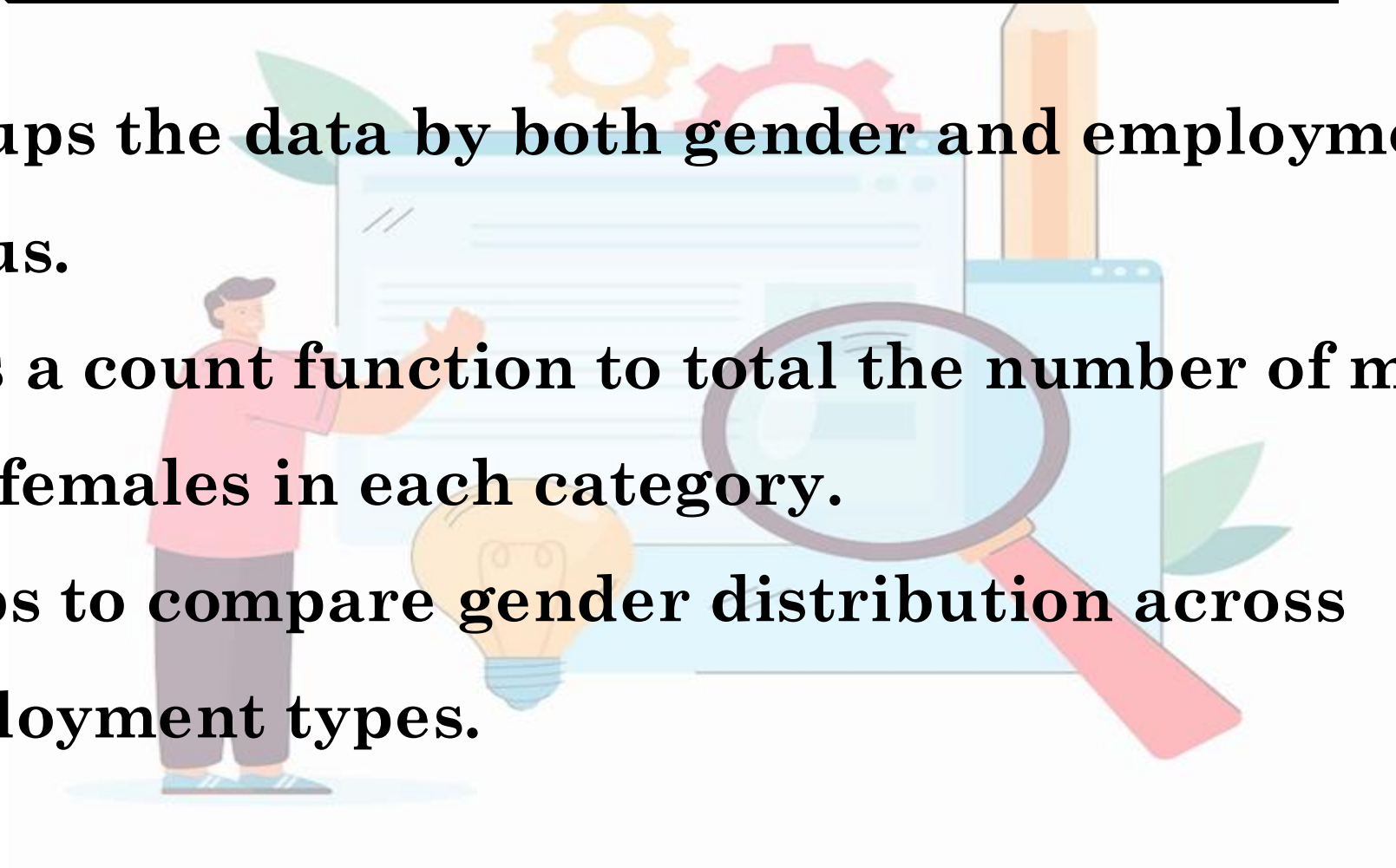
Export:



Employment_Stat...	Gender	Gender_count
Full-time	Male	2564
Full-time	Female	2440
Part-time	Male	1519
Part-time	Female	1497
Self-employed	Male	1040
Self-employed	Female	940

What this query does ?

- **Groups the data by both gender and employment status.**
- **Uses a count function to total the number of males and females in each category.**
- **Helps to compare gender distribution across employment types.**



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

```
80
81 #What is the total and average income by Location and Occupation?
82 SELECT SUM(Income) AS "Total income",AVG(Income) AS "Avg income",Location,
83 Occupation
84 FROM project
85 GROUP BY Location, Occupation
86 ORDER BY SUM(Income);
87
```

100% 22:86

Result Grid



Filter Rows:



Search

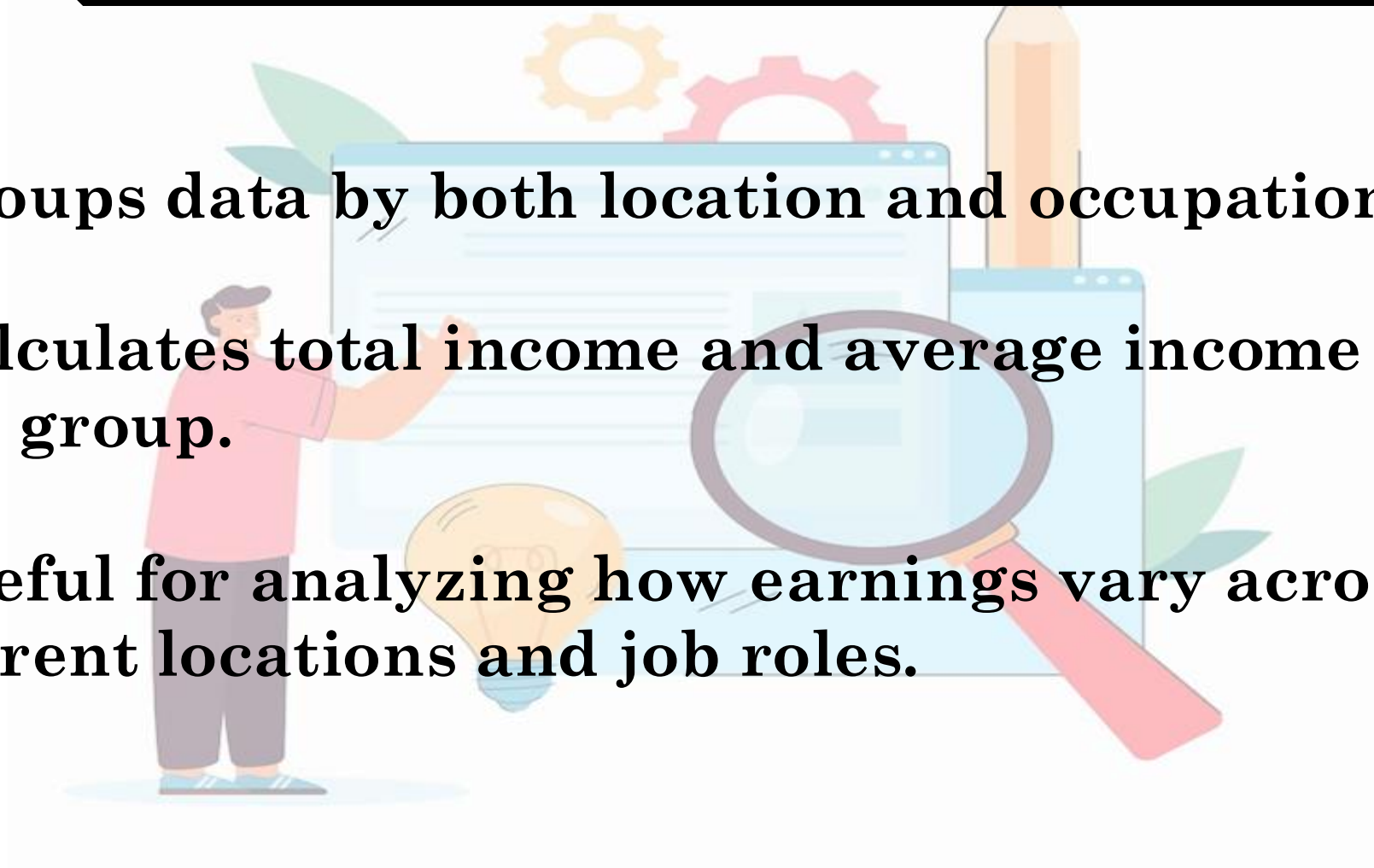
Export:



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

What this query does ?

- **Groups data by both location and occupation.**
- **Calculates total income and average income for each group.**
- **Useful for analyzing how earnings vary across different locations and job roles.**



7: Find the average Household Size grouped by Type_of_Housing.

```
92 #Find the average Household_Size grouped by Type_of_Housing.  
93 • select AVG(Household_Size),Type_of_Housing  
94 FROM project  
95 group by Type_of_Housing;  
96  
97  
98  
99  
100
```

[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)

100% 43:93

Result Grid



Filter Rows:



Search

Export:



Madanu Akhil_SQL_INCOME

AVG(Household_Siz...	Type_of_Housing	
3.9808	Apartment	
3.9721	Single-family home	
4.0442	Townhouse	

SYMMETRICAL

CENTERED
WING/GABLE

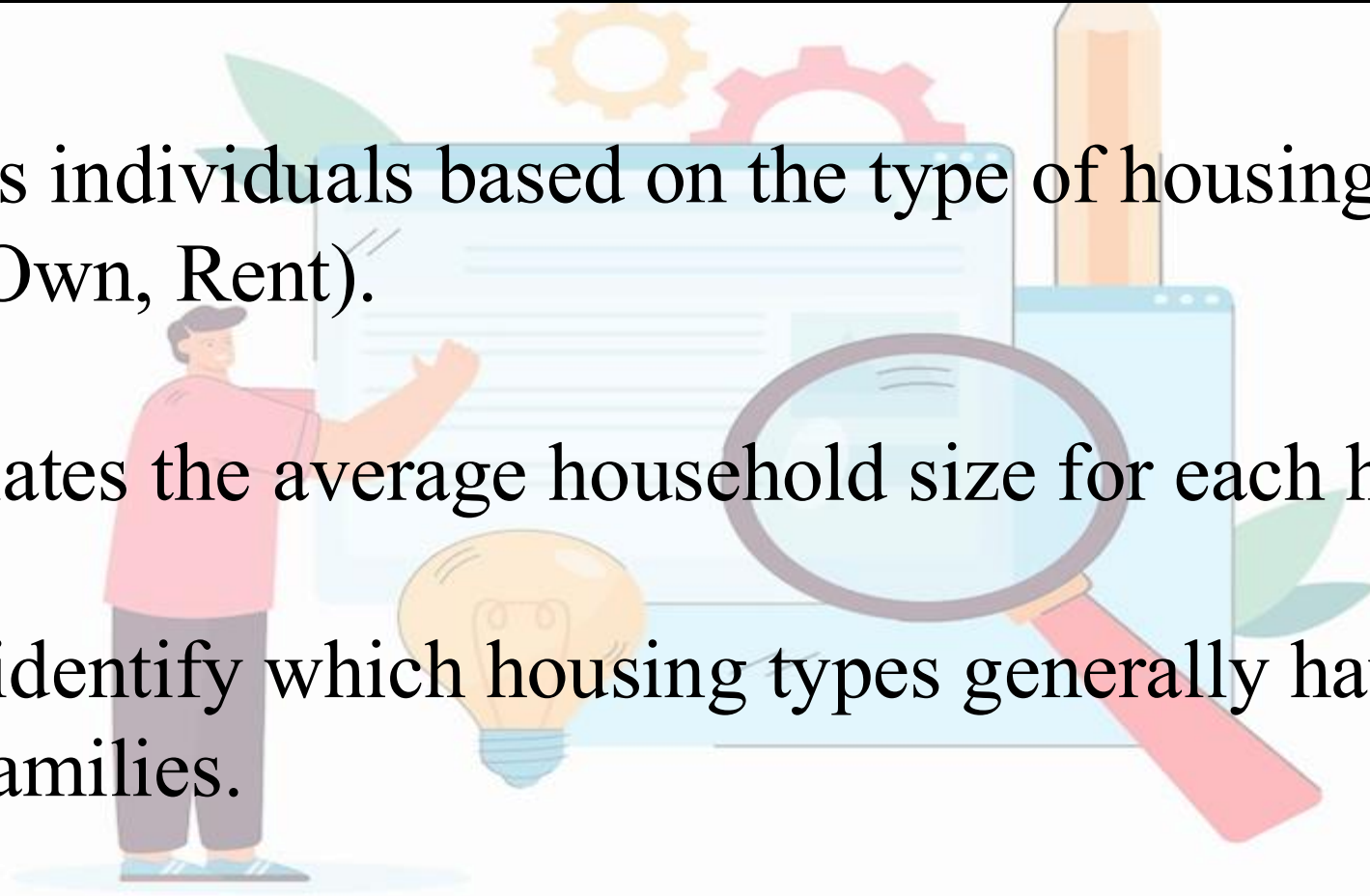
ASYMMETRICAL

TOWN HOUSE

TOWERED

What this query does ?

- Groups individuals based on the type of housing they live in (e.g., Own, Rent).
- Calculates the average household size for each housing type.
- Helps identify which housing types generally have larger or smaller families.



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

```
105 #Calculate the minimum, maximum, and average Work_Experience for each Marital_Status.
106 select min(Work_Experience),max(Work_Experience),Avg(Work_Experience),Marital_Status
107 from project
108 group by Marital_Status
```

100% 24:108

Result Grid



Filter Rows:



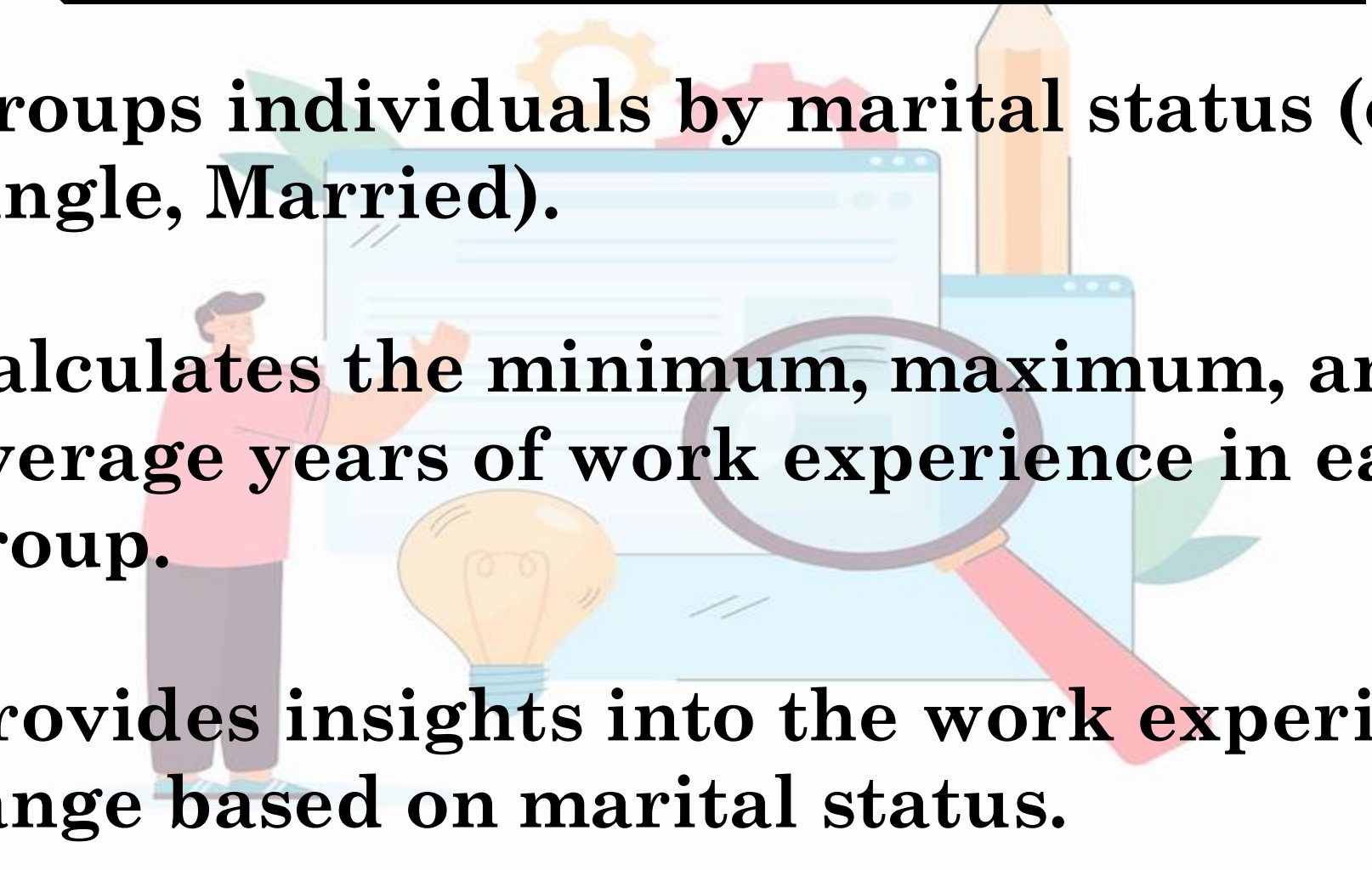
Search

Export:



min(Work_Experien...	max(Work_Experien...	Avg(Work_Experien...	Marital_Stat...	
0	50	25.0569	Married	
0	50	24.5969	Single	
0	50	24.8631	Divorced	

What this query does ?

- **Groups individuals by marital status (e.g., Single, Married).**
 - **Calculates the minimum, maximum, and average years of work experience in each group.**
 - **Provides insights into the work experience range based on marital status.**
- 
- A background illustration featuring a person in a pink shirt and grey pants pointing at a large blue screen. A magnifying glass with a red handle is positioned over the screen. Above the screen are two interlocking gears, one yellow and one pink. To the left of the screen is a glowing yellow lightbulb. To the right is a yellow pencil. The entire scene is set against a light blue background with green leaves.

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

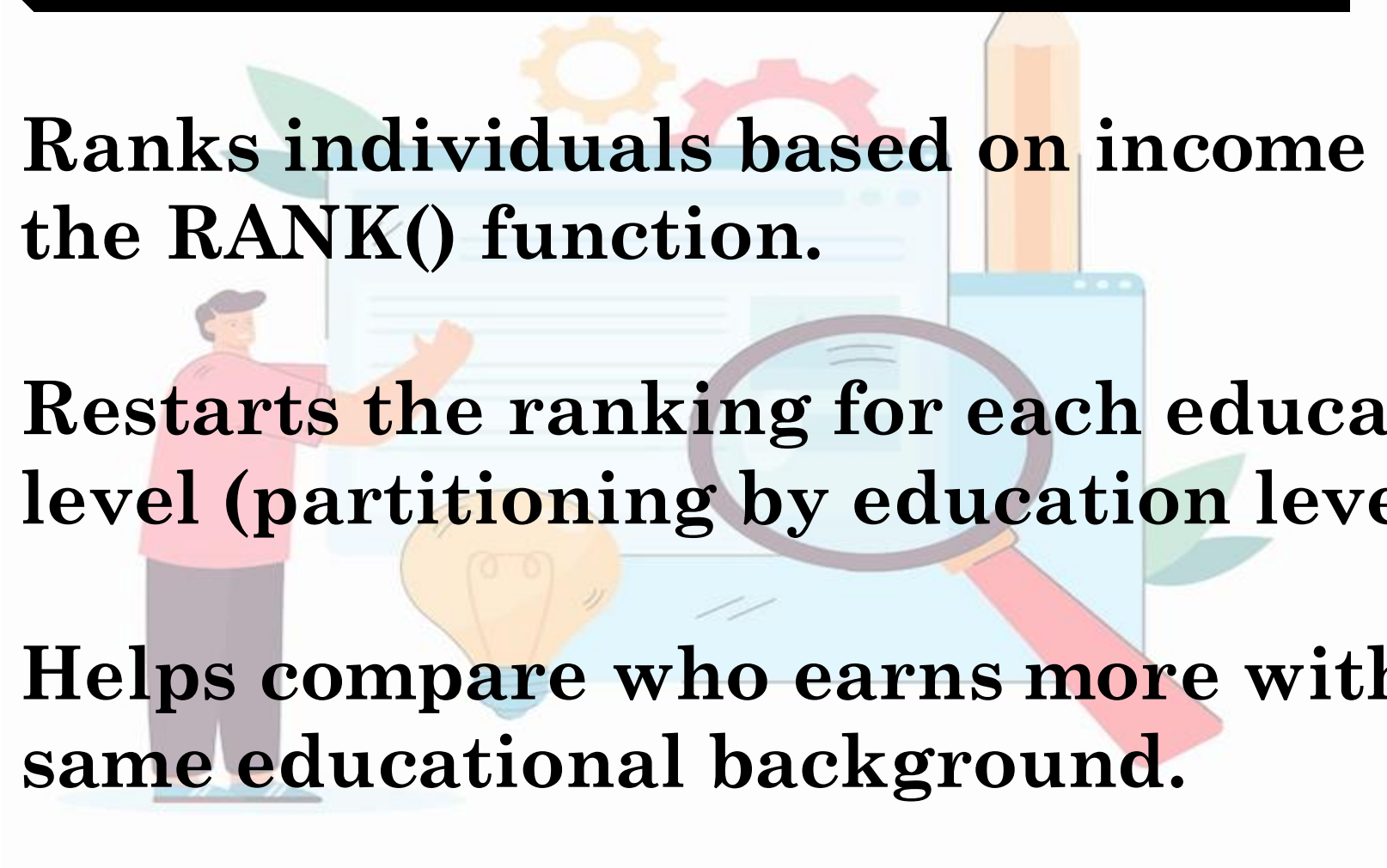
```
116
117 #Write a query to rank individuals by Income within each Education_Level.
118 • SELECT Education_Level,Income,
119     RANK() OVER (
120     PARTITION BY Education_Level
121     ORDER BY Income DESC
122     ) AS Income_Rank
123 FROM project;
```

Education_Level	Income	Income_Rank
Bachelor's	9992571	1
Bachelor's	9987395	2
Bachelor's	9968165	3
Bachelor's	9922858	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
Bachelor's	9653848	15
Bachelor's	9639102	16

Education_Level	Income	Income_Rank
Doctorate	6675817	13
Doctorate	6357275	14
Doctorate	6329045	15
Doctorate	5875804	16
Doctorate	5726383	17
Doctorate	5081235	18
Doctorate	5031158	19
Doctorate	4772933	20
Doctorate	4684104	21
Doctorate	4641934	22
Doctorate	4453380	23
Doctorate	4324067	24
Doctorate	4277787	25
Doctorate	4271221	26
Doctorate	4178989	27
Doctorate	4178540	28





Education_Level	Income	Income_Rank
Master's	9979438	1
Master's	9914893	2
Master's	9686527	3
Master's	9556337	4
Master's	9550359	5
Master's	9519316	6
Master's	9504728	7
Master's	9471717	8
Master's	9431162	9
Master's	9392668	10
Master's	9370134	11
Master's	9350351	12
Master's	9212955	13
Master's	9138349	14
Master's	9127155	15
Master's	9092242	16

What this query does ?

- Ranks individuals based on income using the RANK() function.
 - Restarts the ranking for each education level (partitioning by education level).
 - Helps compare who earns more within the same educational background.
- 
- A background illustration featuring a person in a pink shirt and grey pants pointing at a large blue screen. A magnifying glass with a red handle is positioned over the screen. Above the screen are two interlocking gears, one yellow and one pink. Below the screen is a glowing yellow lightbulb. The entire scene is set against a light blue background with green leaves.

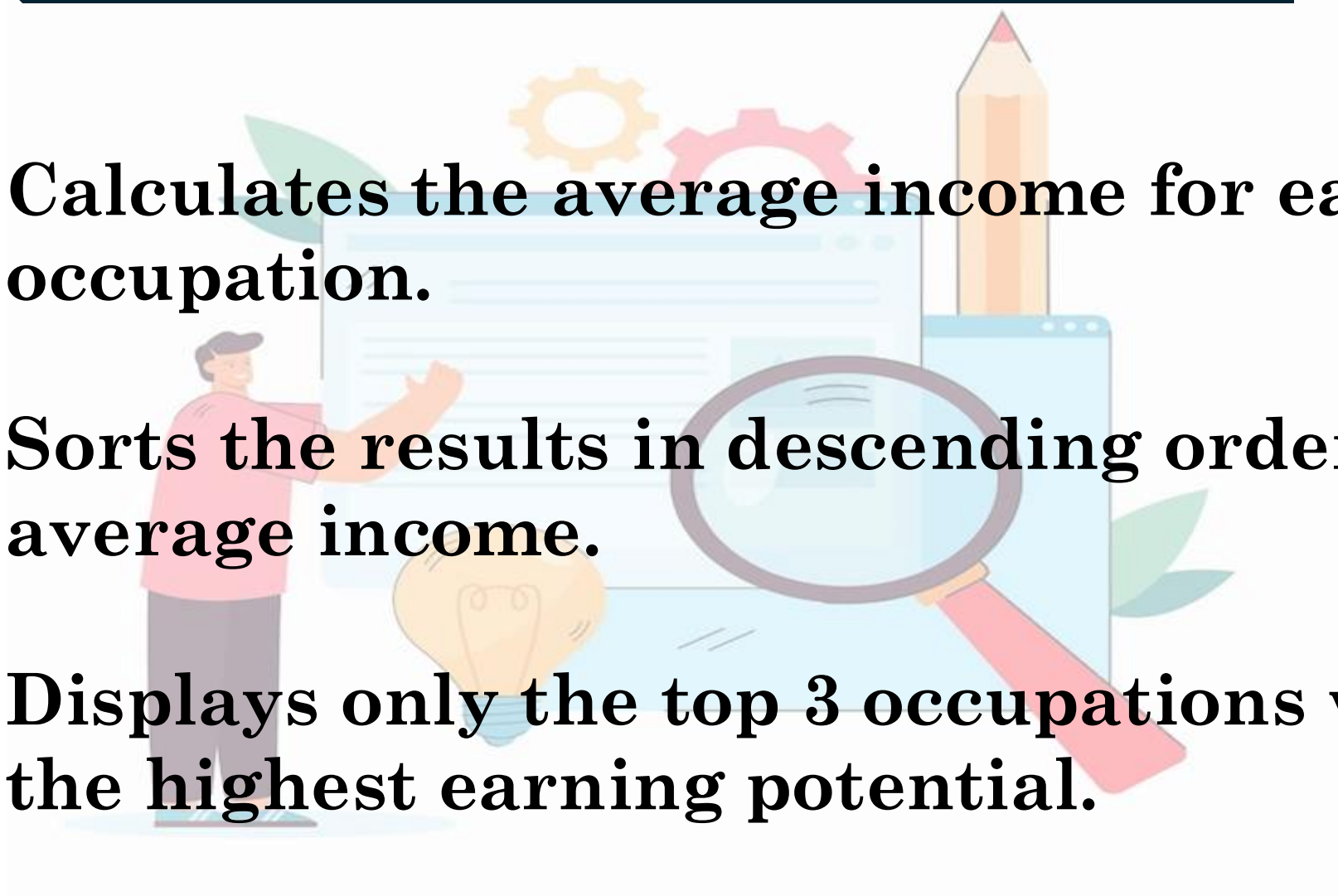
10: Find the top 3 Occupation types with the highest average income.

```
33 #Find the top 3 Occupation types with the highest average income.
34 SELECT Occupation, AVG(Income) AS Highest_Avg_Income
35 FROM project
36 GROUP BY Occupation
37 ORDER BY Highest_Avg_Income DESC
38 LIMIT 3;
```

Result Grid   Filter Rows: Export:  Fetch rows: 

Occupation	Highest_Avg_Inco...	
Education	920816.7526	
Technology	836173.7860	
Others	828970.3925	

What this query does ?

- **Calculates the average income for each occupation.**
 - **Sorts the results in descending order of average income.**
 - **Displays only the top 3 occupations with the highest earning potential.**
- 
- A background illustration featuring a person in a pink shirt and grey pants pointing at a large blue screen. A magnifying glass with a red handle is positioned over the screen. Above the screen are two gears, one yellow and one pink. To the right of the screen is a large yellow pencil. Below the screen is a glowing yellow lightbulb. The entire scene is set against a light blue background with green leaves.

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

```
48 #Use a window function to calculate the cumulative income for each Gender
49 * SELECT Gender,Income,SUM(Income) OVER (
50 PARTITION BY Gender
51 ORDER BY Income
52 ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW
53 ) AS Cumulative_Income
54 FROM project;
```

Gender	Income	Cumulative_Inco...
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

Male	33064	828642
Male	33104	861746
Male	33186	894932
Male	33326	928258
Male	33338	961596
Male	33528	995124
Male	33536	1028660
Male	33732	1062392
Male	33827	1096219
Male	33920	1130139
Male	33945	1164084
Male	33945	1198029
Male	34068	1232097
Male	34174	1266271
Male	34310	1300581
Male	34426	1335007
Male	34564	1369571

What this query does ?

- Uses a window function to compute a running total of income for each gender.
- Groups and orders the data so the cumulative income grows line-by-line.
- Shows how income adds up within male and female categories.

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

```
59 WITH IncomeRanks AS (  
60     SELECT *, PERCENT_RANK() OVER (ORDER BY Income) AS IncomePercentile  
61     FROM project  
62 )  
63 SELECT *  
64 FROM IncomeRanks  
65 WHERE IncomePercentile > 0.5  
66 order by IncomePercentile Desc  
67 limit 10 ;  
68
```

0% 11:167

Result Grid



Filter Rows:



Search

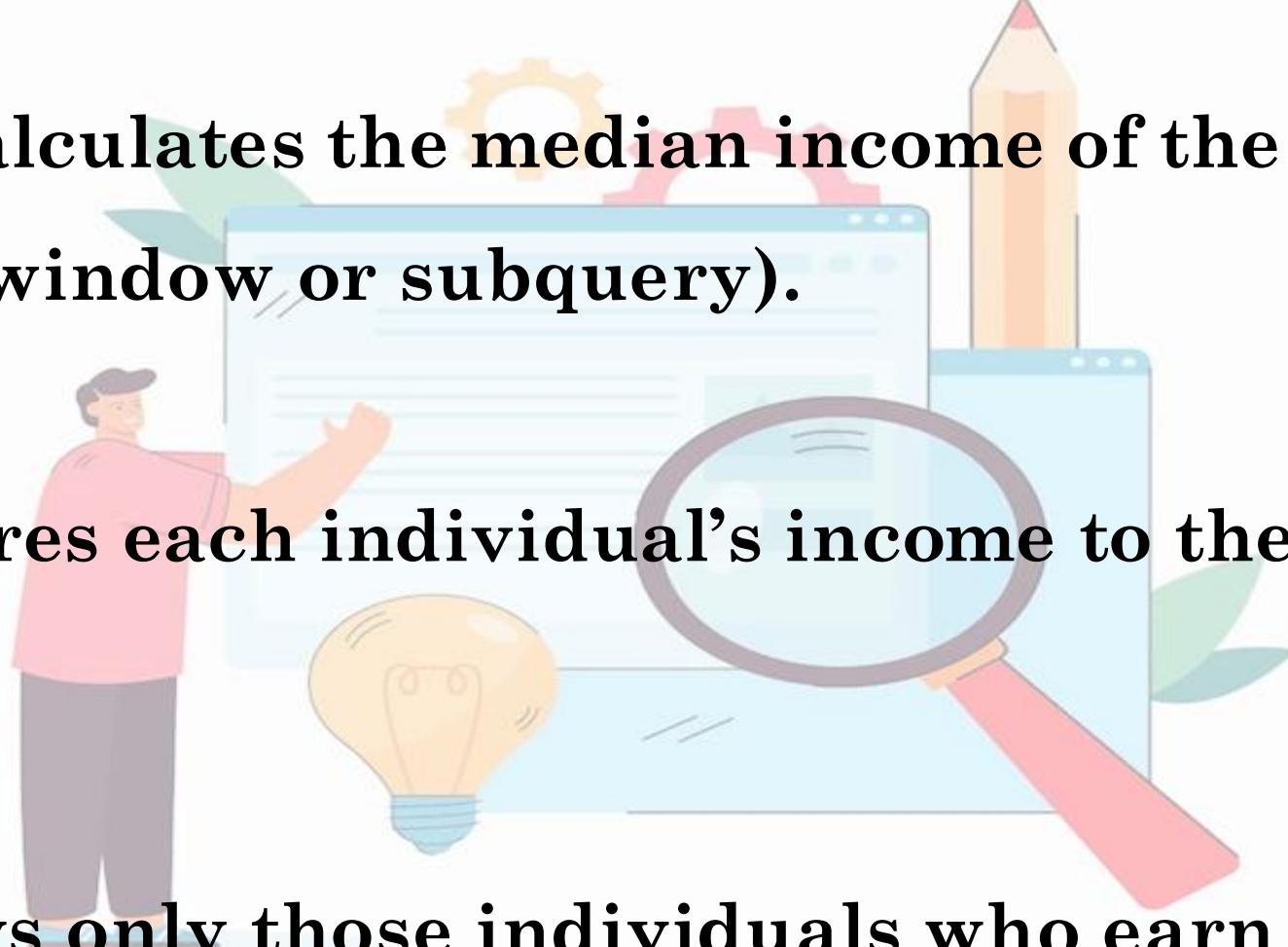
Export:



Marital_Stat...	Employment_Stat...	Household_Size	Homeownership_Stat...	Type_of_Housing	Gender	Primary_Mode_of_Transportat...	Income	IncomePercentile
Married	Part-time	2	Own	Apartment	Female	Public transit	9992571	1
Single	Full-time	1	Rent	Townhouse	Male	Public transit	9987395	0.9998999899989999
Married	Full-time	4	Rent	Single-family home	Male	Public transit	9979438	0.9997999799979999
Married	Full-time	1	Own	Apartment	Female	Biking	9968165	0.9996999699969997
Married	Full-time	7	Own	Single-family home	Male	Public transit	9922858	0.9995999599959996
Married	Full-time	2	Rent	Single-family home	Female	Car	9914893	0.9994999499949995
Married	Part-time	3	Own	Apartment	Male	Biking	9904254	0.9993999399939995
Married	Full-time	2	Own	Single-family home	Female	Car	9892000	0.9992999299929993
Single	Full-time	4	Own	Single-family home	Female	Public transit	9871131	0.9991999199919992
Single	Part-time	1	Rent	Single-family home	Male	Biking	9859518	0.9990999099909991

What this query does ?

- First calculates the median income of the dataset (using window or subquery).
- Compares each individual's income to the median value.
- Displays only those individuals who earn more than the median.



Conclusion

Through this SQL-based exploration of the Income data table, I gained hands-on experience in extracting, filtering, grouping, and summarizing data effectively. By writing and analyzing 12 different queries, I was able to:

- Understand the average income for each Education Level for those who are employed full-time.
- Analyze income based on criteria such as age, education, gender, employment status etc.,
- Category and Experience Analysis.
- House type and hold size based on Income.

Key Takeaways: Form deep data analysis with precision.

- Writing queries helped me understand the structure and relationships within the dataset.
- "Exploring data with SQL is a crucial first step that supports deeper analysis and effective visualization in tools like Power BI and Tableau."

The background is a dark blue, futuristic digital interface. It features a central glowing orb held by a stylized hand. The orb is composed of concentric circles with a purple and blue gradient. Surrounding the orb are various digital elements: binary code (0s and 1s) on the left and right, and several circular patterns resembling orbits or data paths. The overall aesthetic is high-tech and digital.

Thank you