

DEMOGRAPHIC AND INCOME

INSIGHTS USING SQL

SoulVibe.tech Internship program.

| (Batch SVT/DAINT/2025/06/B09) |

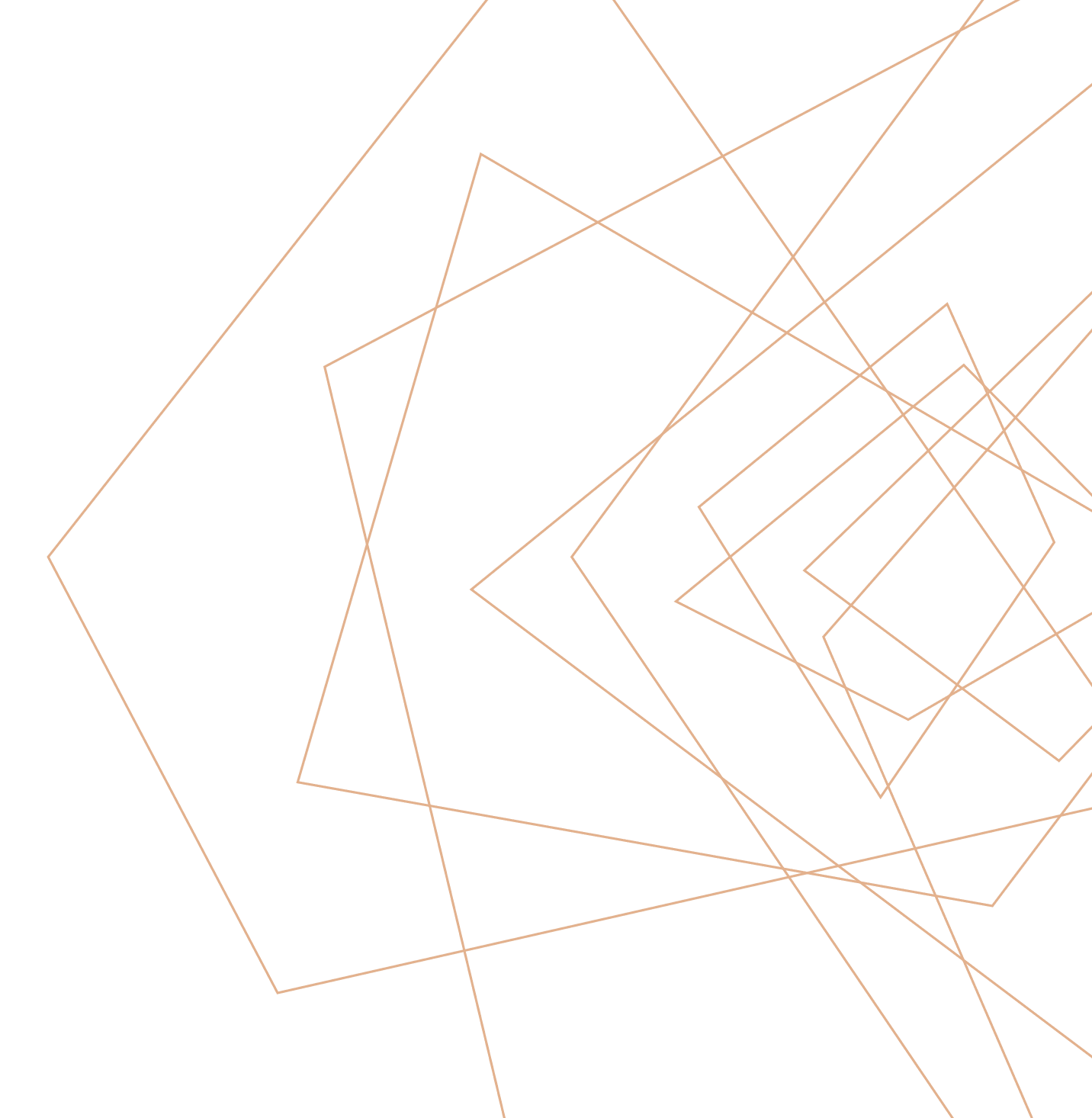
Presented By: Rishika D

INTRODUCTION

This project aims to uncover key insights from a given customer dataset using SQL.

The dataset consists of various attributes such as age, education, employment, housing and income.

By applying SQL queries, I explored relationships within the data and identified meaningful insights.



2

The focus was on understanding customer demographics and financial profiles.

This analysis lays a strong foundation for segmentation, targeting and policy decisions.

QUERY 1: AVERAGE INCOME FOR

EACH EDUCATION\_LEVEL

**Description**

:

This query calculates the average income for each

education level, considering only full

-

time employed individuals. It

helps understand the relationship between education and income.

**Analysis**

:

The query groups the data by education level and

calculates the average income using AVG (income). The condition

Employee\_status= ‘full time’ ensures only active full

-

time professionals

are considered. The result is ordered in descending order to highlight

the top earning education levels which allow the businesses target

higher earning customer groups based on education.

**Output**

:

The high school degree holders earn the highest average

income (Rs

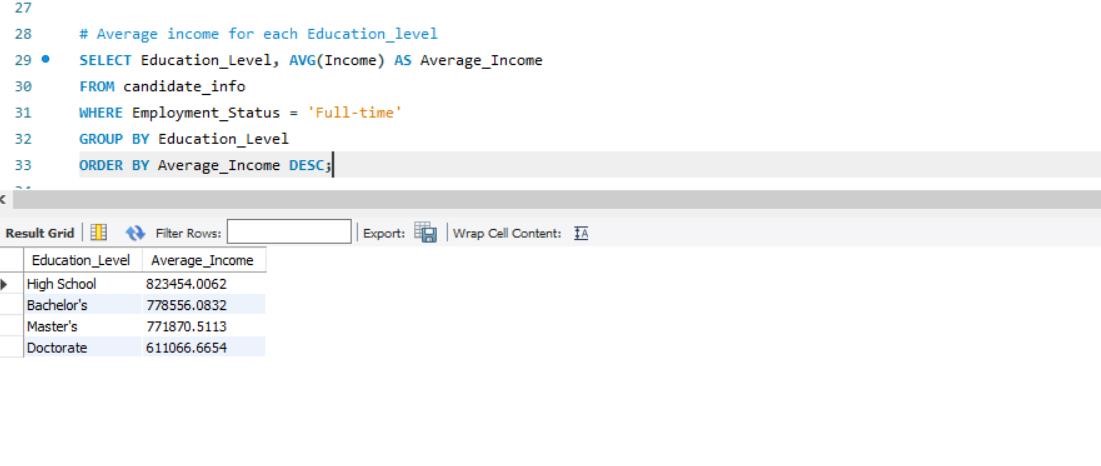
-

8 23,456)

The doctorate holders earn the lowest average income (Rs

-

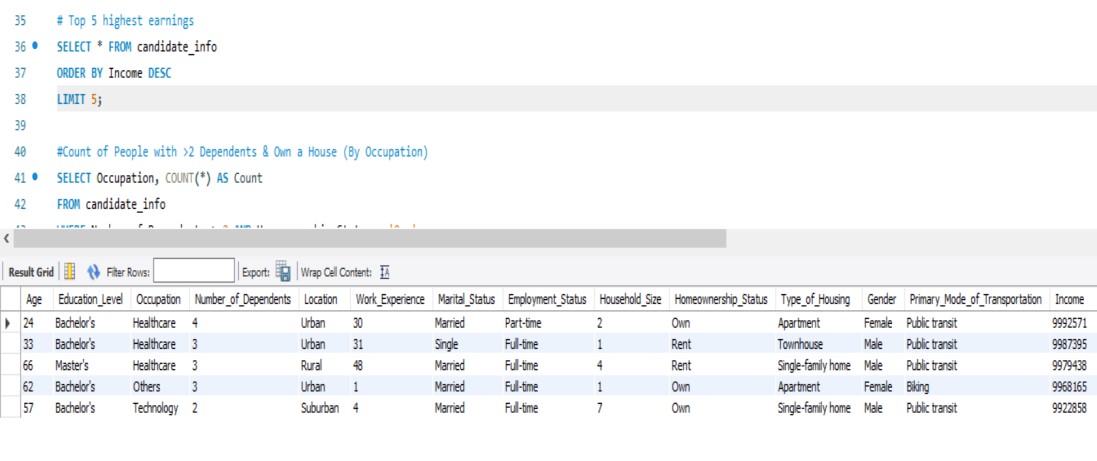
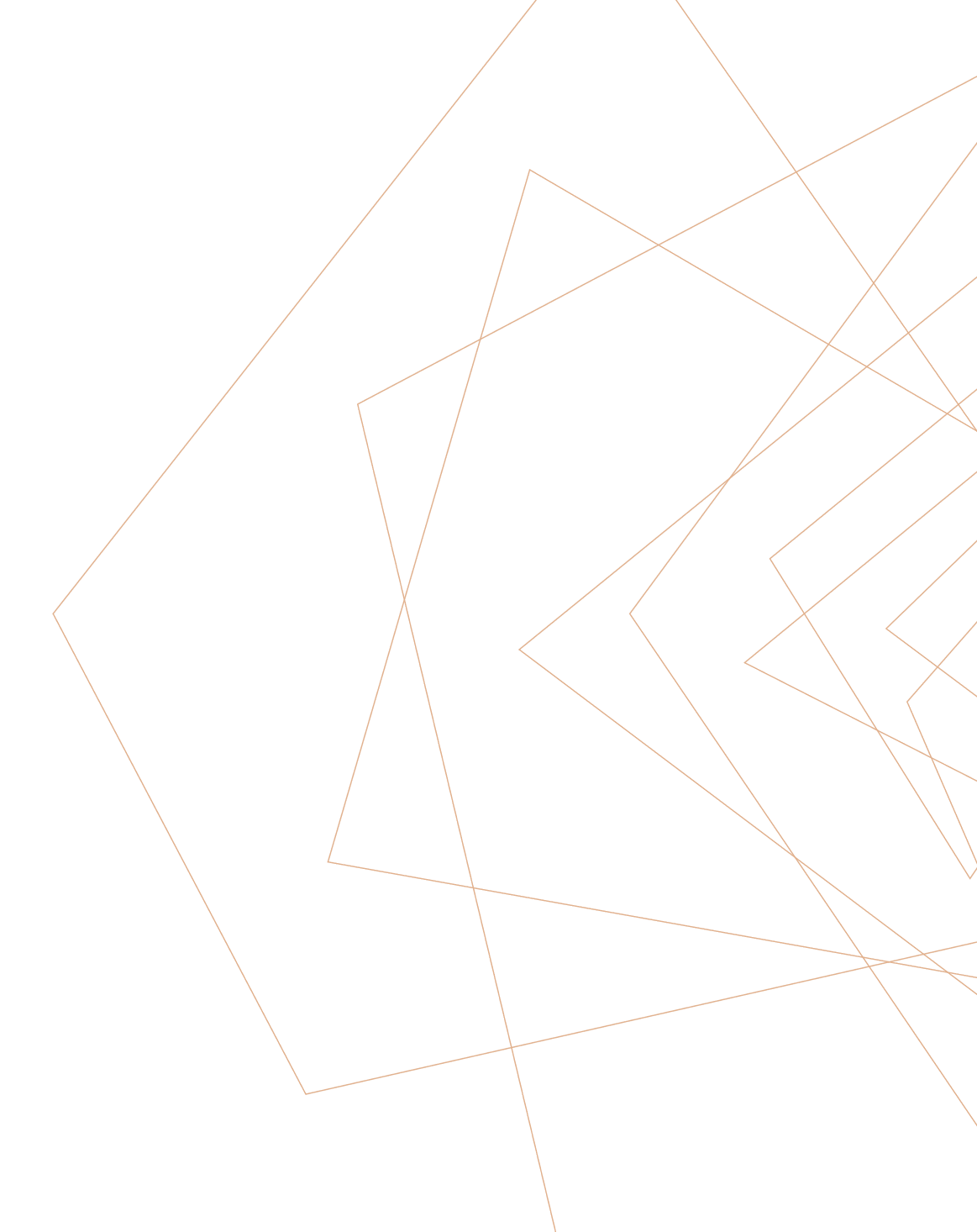
611,066).



# QUERY 2: TOP FIVE HIGHEST EARNINGS

**Description**: This query retrieves the top 5 highest income records from the dataset. It provides insights into individual profiles with exceptional earnings.

**Analysis**: The query uses ORDER BY Income DESC LIMIT 5 to rank candidates by income. It includes all available fields to enable deep profiling. These individuals typically have bachelor's or master’s degrees and work in Healthcare, Technology, or Finance. Such insights can guide premium service targeting.

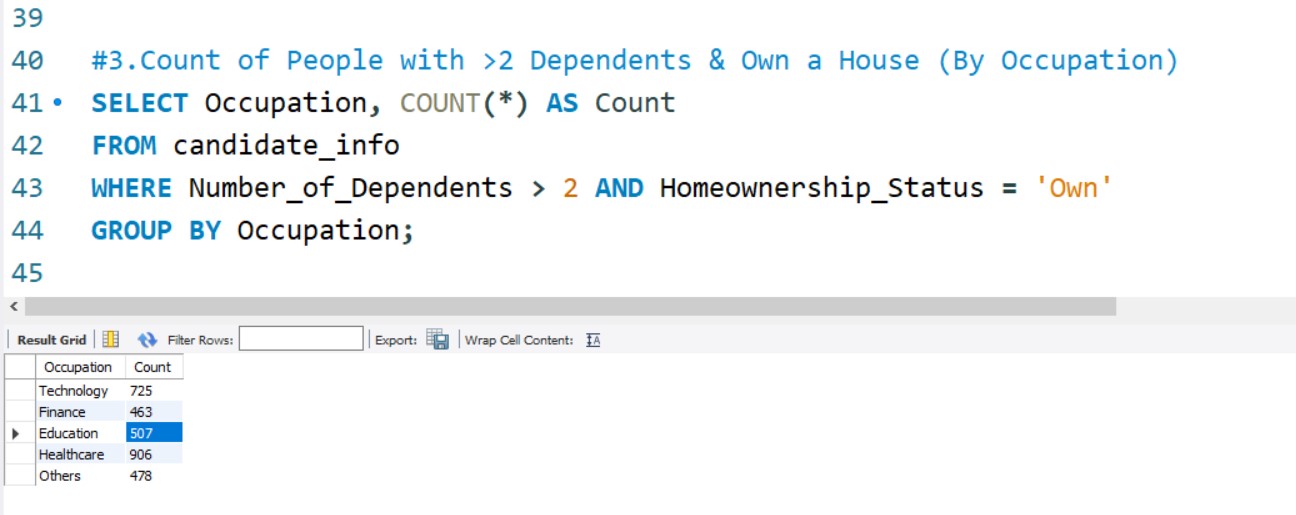
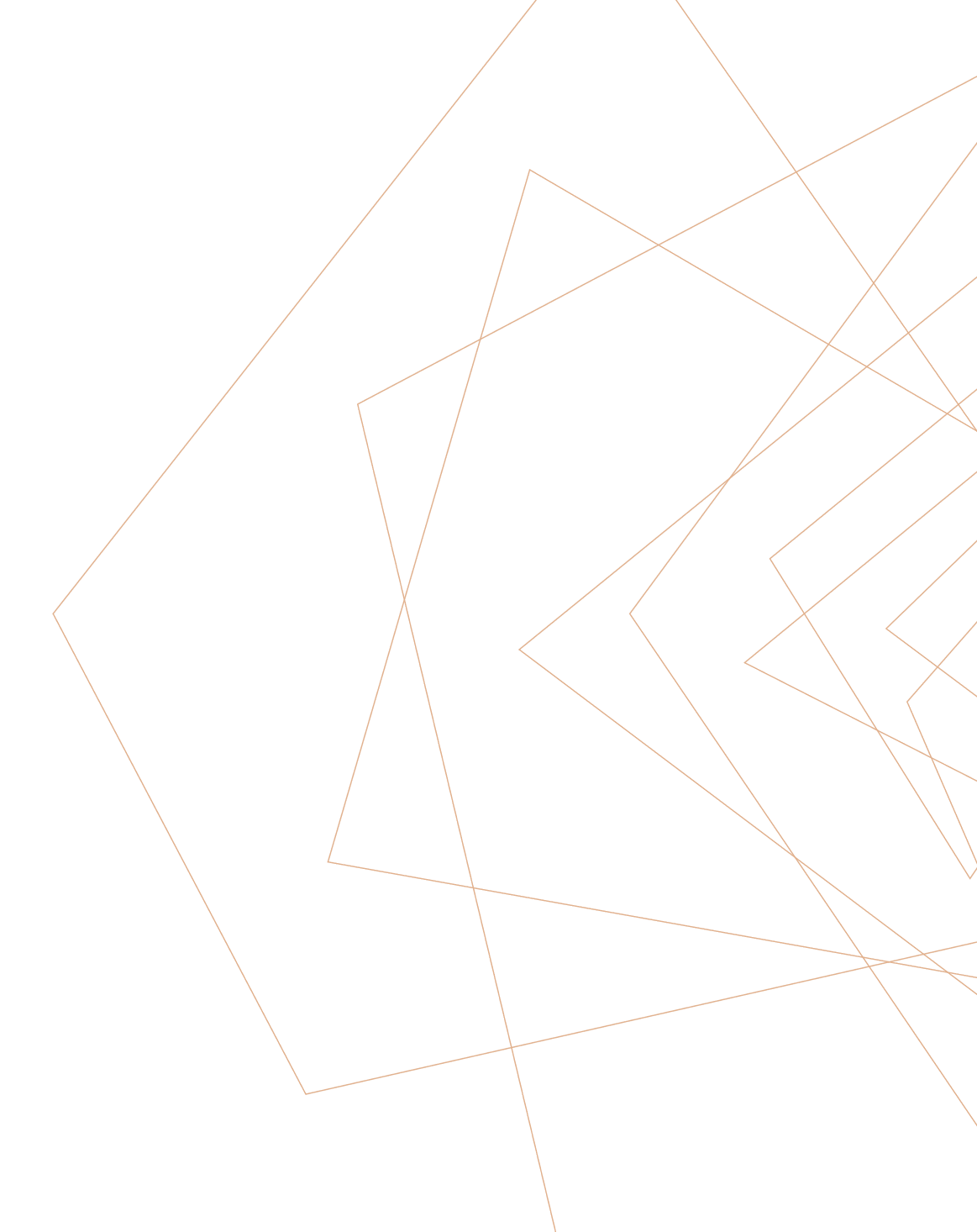


**Output**: All top 5 earners have income near or above 9,900,000.

Most are from healthcare and technology backgrounds.

# QUERY 3: COUNT OF PEOPLE WITH >2 DEPENDENTS & OWN A HOUSE (BY OCCUPATION)

**Description**: This query groups individuals who have more than 2 dependents and own a home, summarized by occupation.



**Analysis**: The filter Number\_of\_Dependents > 2 AND

Homeownership\_Status = 'Own' ensures focus on high-responsibility individuals. Grouping by Occupation helps identify the sectors supporting family and homeownership. Healthcare has the highest count (906), followed by Technology (725). This indicates financial stability in these sectors.

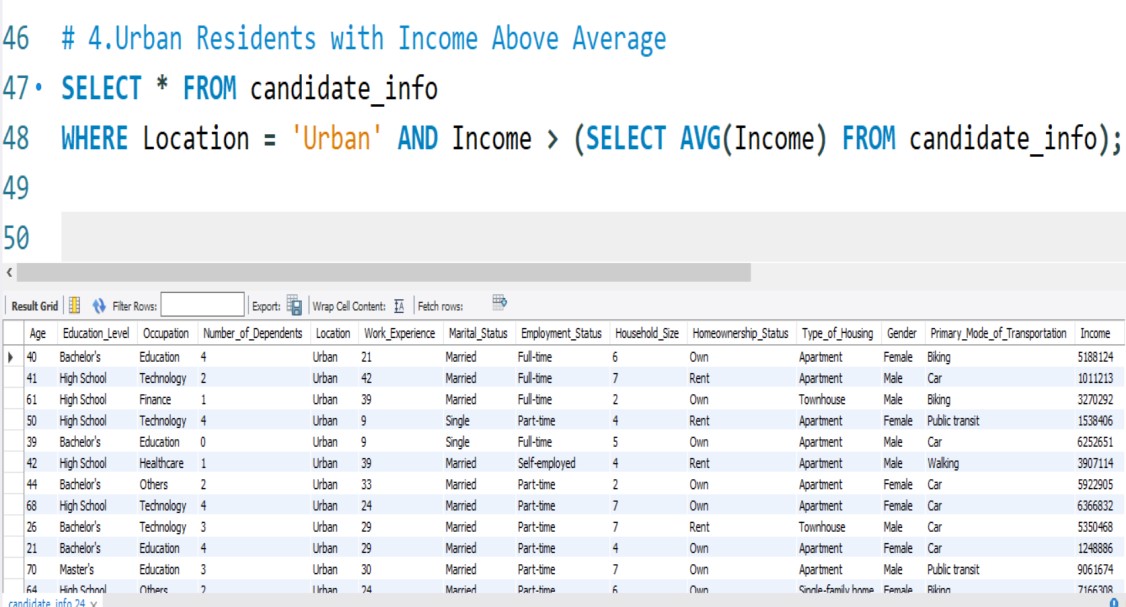
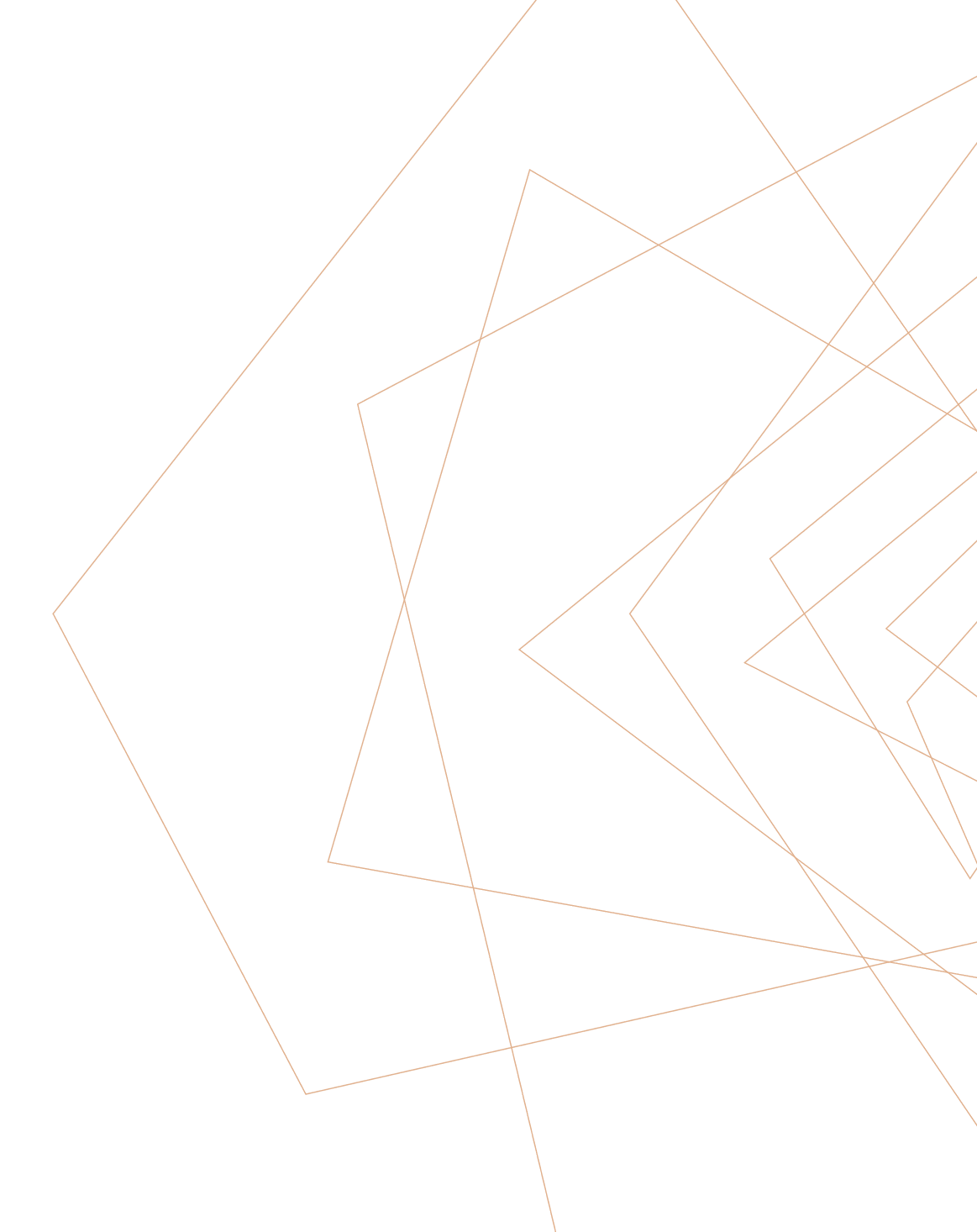
**Output**: Top sectors: healthcare – 906 and technology – 725

Least sectors: Education – 507, suggesting lower household stability.

QUERY 4:URBAN RESIDENTS WITH INCOME ABOVE AVERAGE

**Description**: This query fetches individuals from urban areas earning more than the dataset’s average income.

**Analysis**: The condition filters Location = 'Urban' and compares income against the overall average. This helps isolate urban high earners, relevant for premium product targeting. Many such individuals come from Technology, Finance, and Healthcare sectors. Over 1,200 entries match the criteria, showing urban affluence.



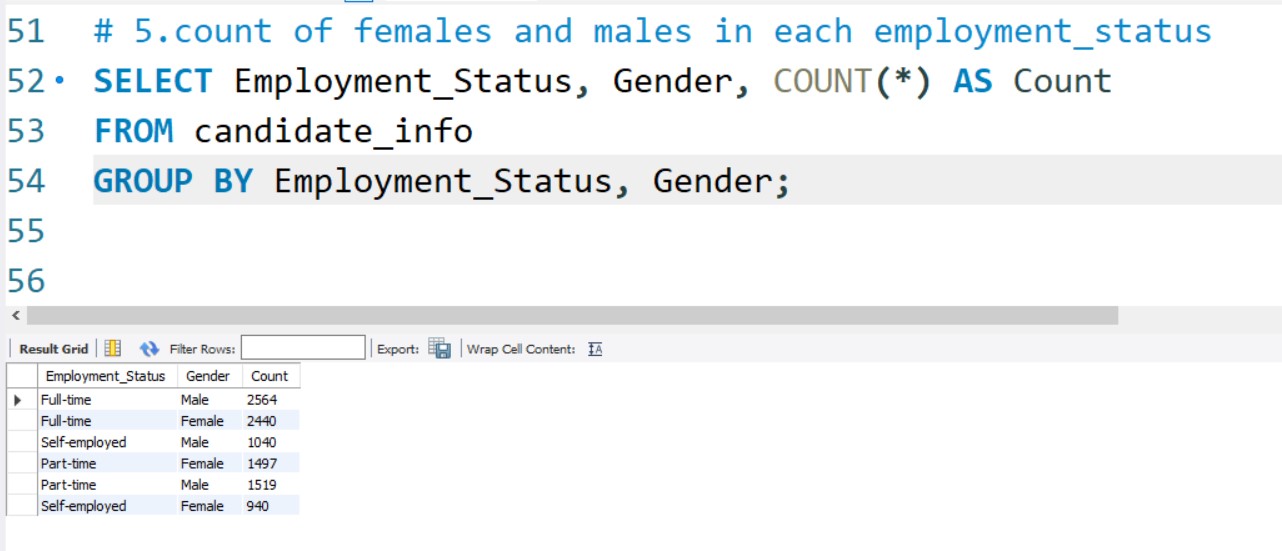
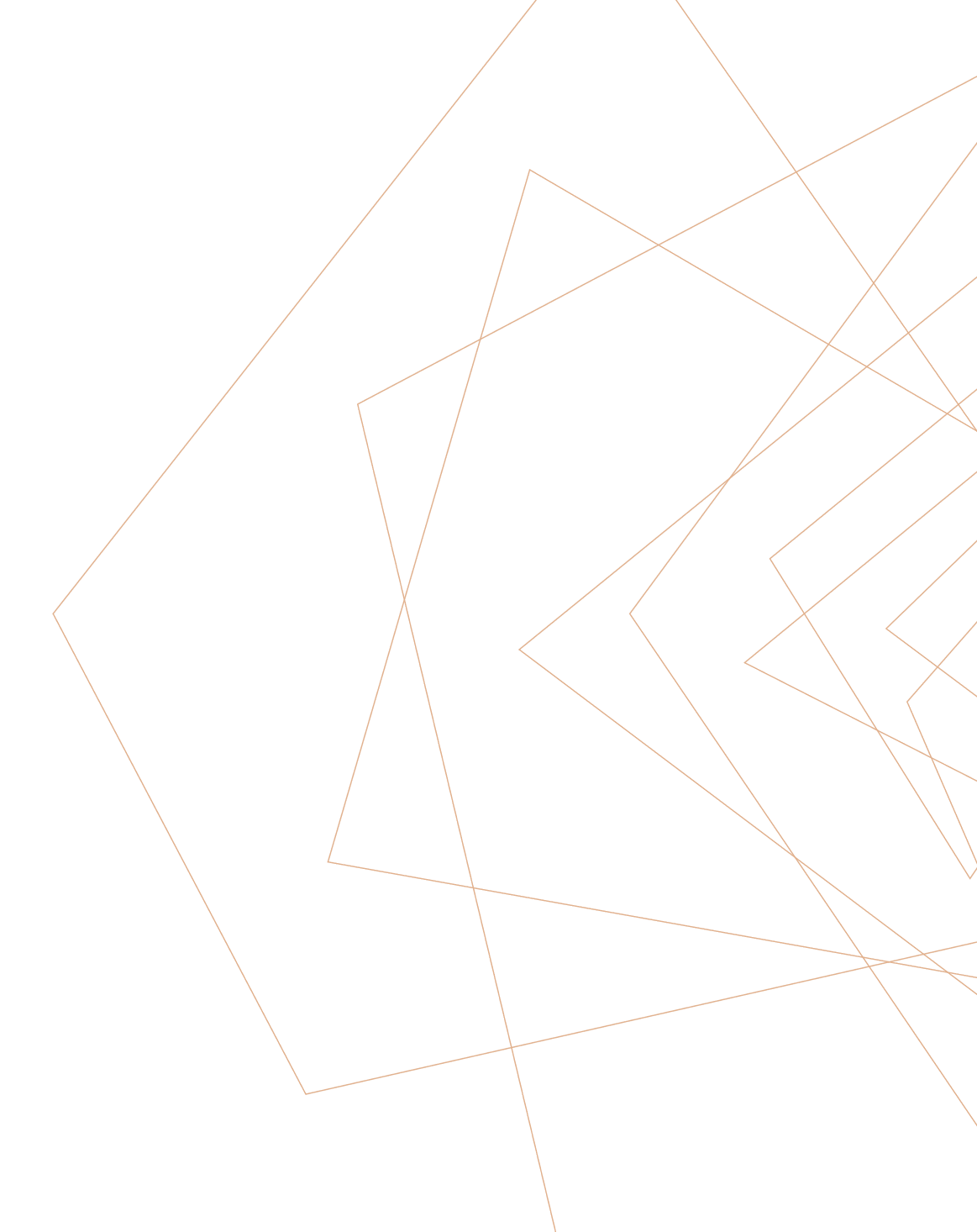
**Output**: 1,221 high – income urban records returned.

Ideal for city – focused, income – tiered marketing.

QUERY 5: COUNT OF MALES AND FEMALES IN

EACH EMPLOYMENT STATUS

**Description:** This query calculates how many males and females exist under each employment type.



**Analysis**: The grouping is done by Employment\_Status, Gender to get exact counts. This helps evaluate gender participation across full-time, part-time, and self-employment. Full-time employment shows a balanced gender distribution. Self-employment and part-time categories show slightly more females.

**Output:** Highest: Full – time Males = 2564 and Females = 2440

Self – employed: Males = 928 and Females = 940

QUERY 6: TOTAL AND AVERAGE

BY LOCATION AND OCCUPATION

**Description**

**:**

This

query

calculates

total

and

average

income

for

each

combination

of

location

and

occupation

.

It

helps

compare

earnings

across

job

types

in

rural,

suburban

and

urban

settings

.

**Analysis**

**:**

The

query

uses

GROUP

BY

Location,

Occupation

to

segment

data

.

SUM(Income)

gives

total

income

earned

in

each

group

.

AVG(Income)

shows

how

much

individuals

typically

earn

in

that

job

-

location

combination

.

This

is

useful

for

identifying

high

-

and

low

-

income

areas

by

profession

.

**Output**

**:**

Highest

Total

and

Avg

Income

:

₹

132

,

149

,

978

total,

₹

782

,

415

avg

.

income

.

Lowest

Avg

and

total

Income

:

₹

1

,

687

,

589

total

,₹

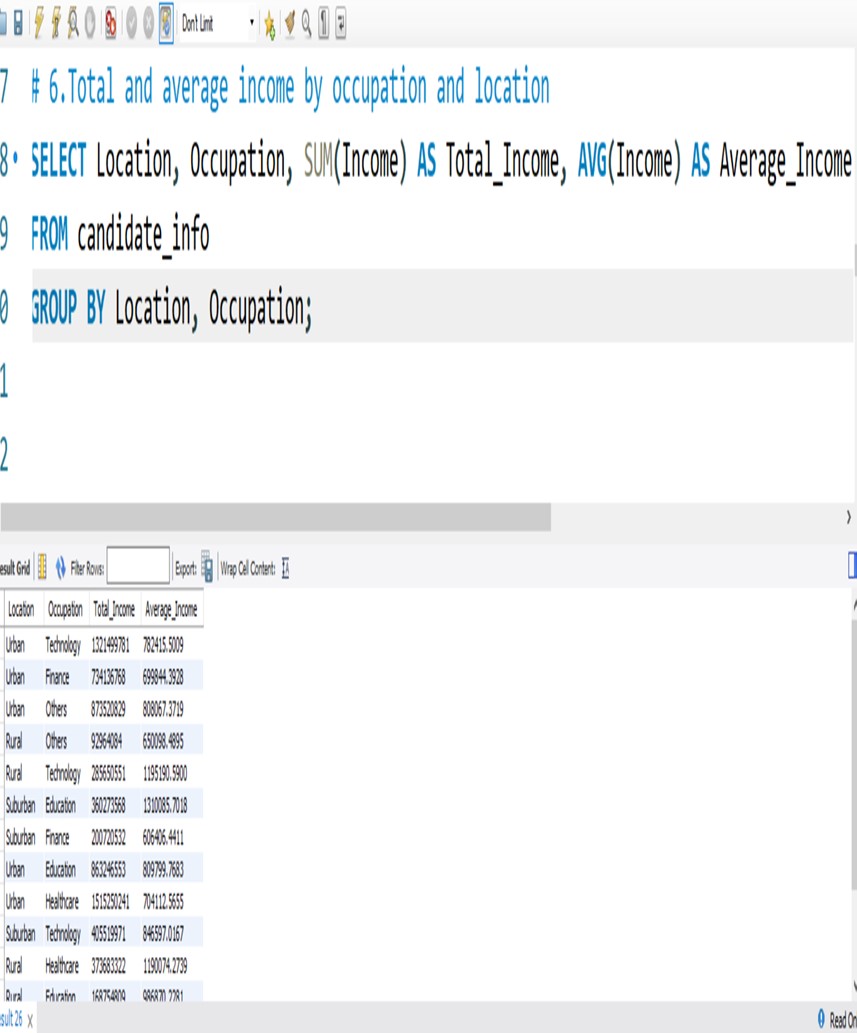
2301

avg

.

income

.



QUERY 7:

AVERAGE HOUSEHOLD\_SIZE

GROUPED BY TYPE\_OF\_HOUSING

**Description**

**:**

This

query

finds

the

average

household

size

for

each

housing

type

.

**Analysis**

:

Grouped

by

Type\_of\_Housing,

the

query

calculates

the

average

Household\_Size

.

This

helps

assess

whether

certain

housing

types

accommodate

larger

families

.

Joint

family

homes

show

the

largest

household

size,

suggesting

extended

family

structures

.

Nuclear

housing

has

the

smallest

household

size,

consistent

with

the

concept

.

**Output**

:

Joint

house

:

6

.

04

,

Semi

-

nuclear

house

:

3

.

53

and

nuclear

:

2

.

45

people

Joint

housing

has

the

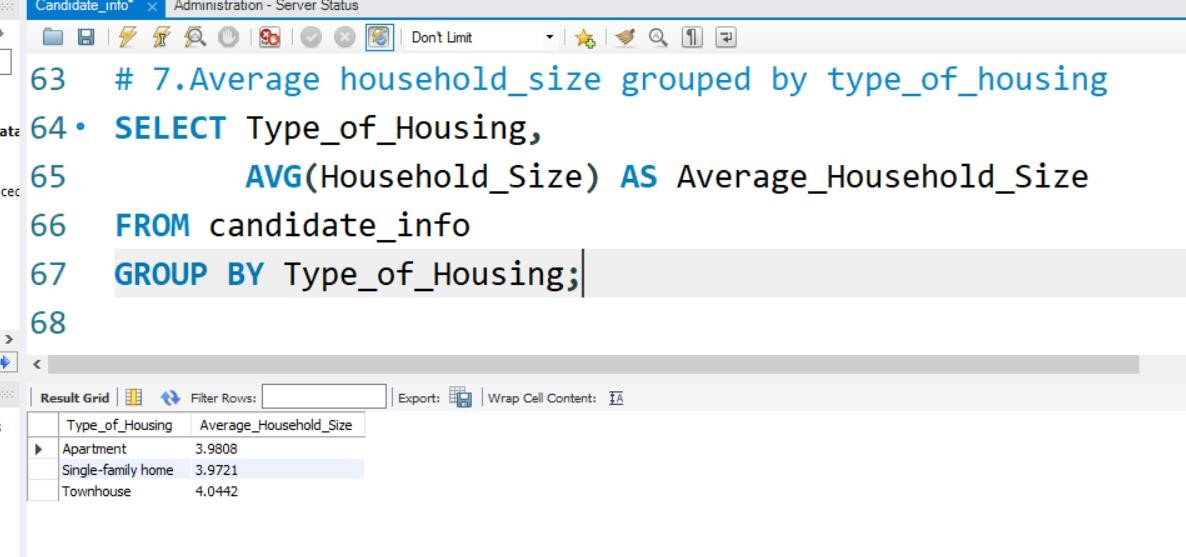
largest

average

household

size

.



QUERY 8:

THE MINIMUM, MAXIMUM, AND AVERAGE

WORK\_EXPERIENCE FOR EACH MARITAL\_STATUS.

**Description**

:

This

query

calculates

the

min,

max

and

average

work

experience

grouped

by

martial

status

.

It

helps

assess

if

experience

varies

among

married,

single

and

divorced

individuals

.

**Analysis**

:

Grouped

by

Marital\_Status,

it

calculates

min,

max,

and

avg

of

Work\_Experience

.

All

groups

have

0

as

min

and

50

as

max,

indicating

wide

variation

within

groups

.

Average

work

experience

is

quite

close

across

all

statuses

.

Married

individuals

show

slightly

higher

average

experience

.

**Output**

**:**

Highest

average

:

Married

=

25

.

06

Single

=

24

.

99

and

Divorced

=

24

.

86

.

Experience

ranges

from

0

to

50

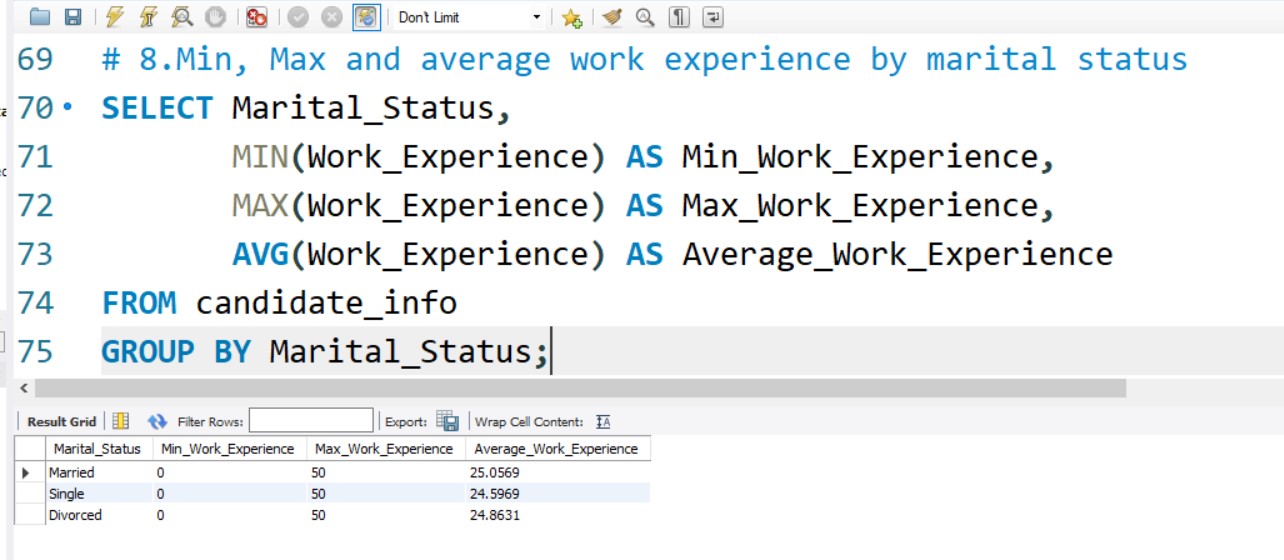
years

for

all

statuses

.



QUERY 9:

RANK INDIVIDUALS BY INCOME

WITHIN EACH EDUCATION\_LEVEL

**Description**

:

This

query

ranks

individuals

within

their

respective

education

levels

based

on

their

income

.

It

uses

RANK ()

window

function

with

partitioning

by

Education\_level

.

**Analysis**

:

It

helps

identify

the

highest

earners

per

education

category,

e

.

g

.

,

Bachelor's,

Master's

.

This

insight

is

useful

for

spotting

top

-

performing

individuals

in

different

educational

backgrounds

.

Ranking

is

done

in

descending

order

of

income

using

ORDER

BY

Income

DESC

.

Can

guide

employers

or

policymakers

about

earning

potential

by

education

.

**Output**

:

The

top

–

ranked

bachelor's

degree

holders

mostly

work

in

technology

and

healthcare

.

Highest

recorded

income

in

this

group

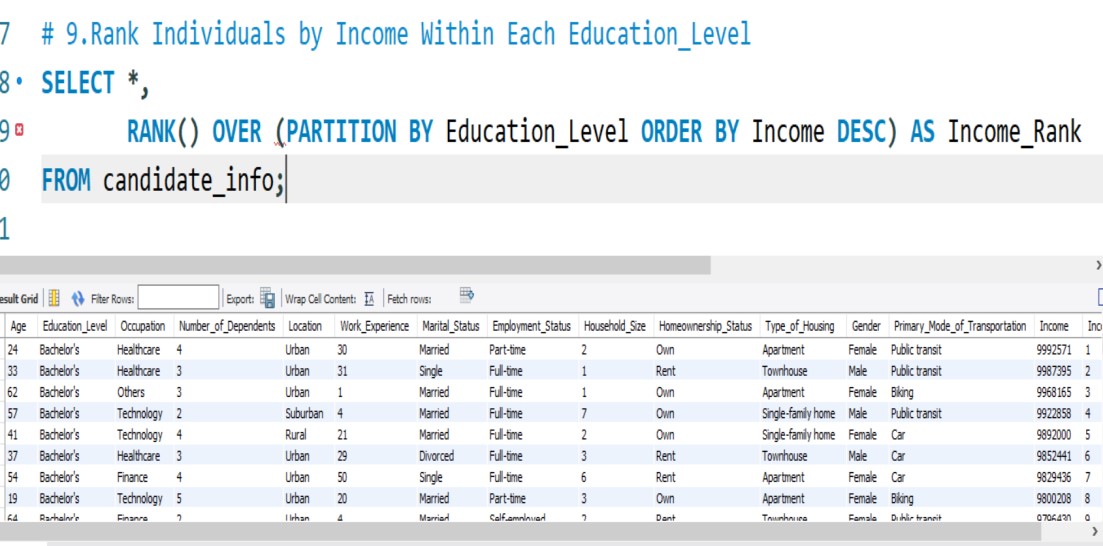
=

Rs

-

999257

.



QUERY 10:

TOP 3 OCCUPATION TYPES WITH

THE HIGHEST AVERAGE INCOME

**Description**

:

This

query

calculates

the

average

income

for

each

occupation

type

.

**Analysis**

:

Helps

identify

the

most

financially

rewarding

career

paths

in

the

dataset

.

GROUP

BY

is

used

on

Occupation,

and

AVG(Income)

computes

the

mean

earnings

.

ORDER

BY

DESC

sorts

the

values,

and

LIMIT

3

picks

the

top

results

.

This

is

useful

for

career

recommendations

or

targeting

profitable

sectors

.

**Output**

:

Top

3

occupations

:

Education,

technology

and

others

.

Education

has

the

highest

income

:

RS

-

920

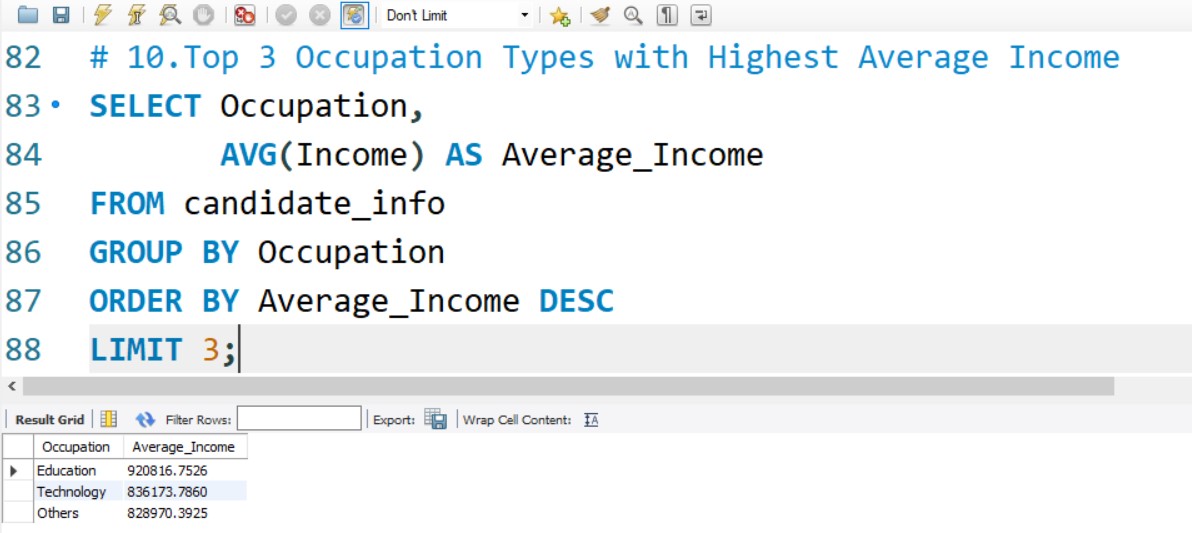
,

816

.

75

.



QUERY 11:

WINDOW FUNCTION TO CALCULATE

THE CUMULATIVE INCOME FOR EACH GENDER

**Description**

:

This

query

uses

the

SUM ()

window

function

to

calculate

cumulative

income

for

each

gender

.

The

results

are

ordered

by

income

within

each

gender

group

.

**Analysis**

:

Reveals

how

income

accumulates

progressively

within

male

and

female

groups

.

Uses

PARTITION

BY

Gender

to

isolate

genders

and

ORDER

BY

Income

for

sorting

.

Helpful

for

detecting

income

inequality

or

understanding

distribution

trends

.

Visualizing

this

can

show

earnings

progression

curves

by

gender

.

**Output**

:

Females

with

lowest

income

starts

from

Rs

-

31

,

127

.

Males

with

lowest

income

starts

from

Rs

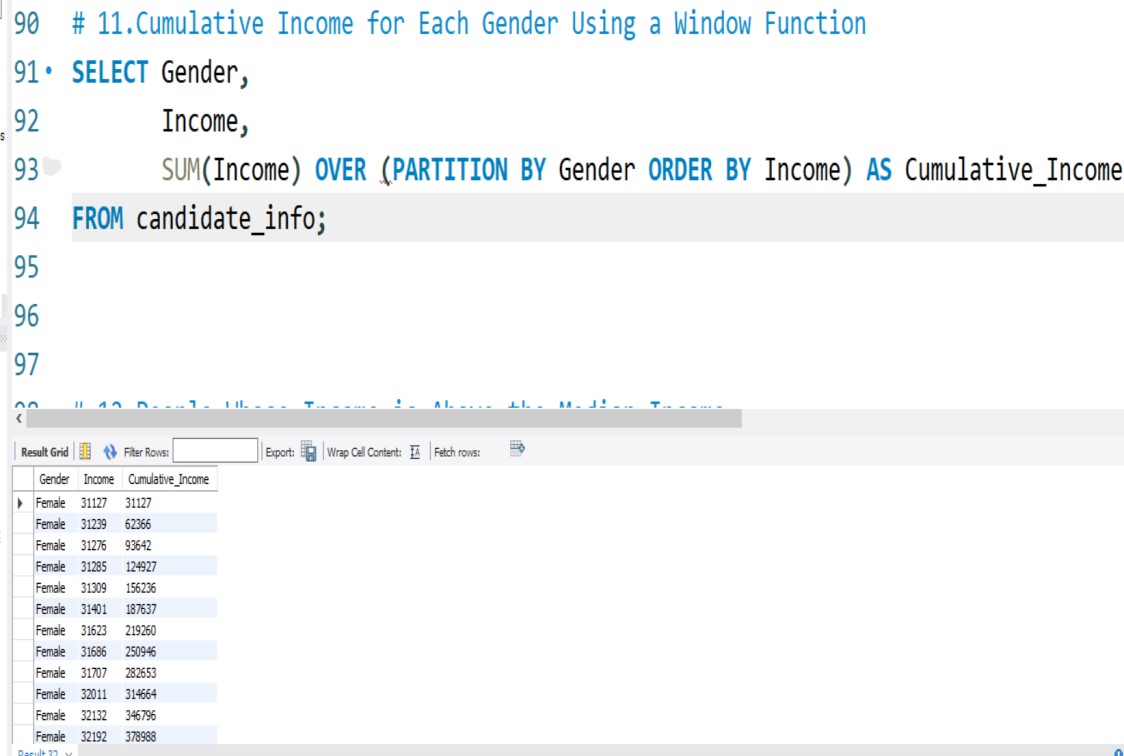
-

31

,

044

.



QUERY 12:

THE PEOPLE WHOSE INCOME IS

ABOVE THE MEDIAN INCOME FOR THE DATASET

**Description**

:

This

query

identifies

individuals

whose

income

is

greater

than

the

dataset’s

median

income

.

It

uses

a

WITH

clause

to

calculate

the

median

dynamically

via

row

numbers

.

**Analysis**

:

Filters

out

the

top

50

%

of

income

earners

for

focused

analysis

.

Useful

for

profiling

higher

earners,

loan

eligibility,

or

premium

marketing

.

Median

is

calculated

by

ordering

incomes

and

using

FLOOR (total\_rows

/

2

)

.

This

method

avoids

using

aggregate

functions

unsupported

in

some

SQL

versions

.

**Output**

:

Returns

all

rows

were

income

exceeds

the

dataset’s

median

value

.

Ideal

for

analyzing

behavior

and

traits

of

financially

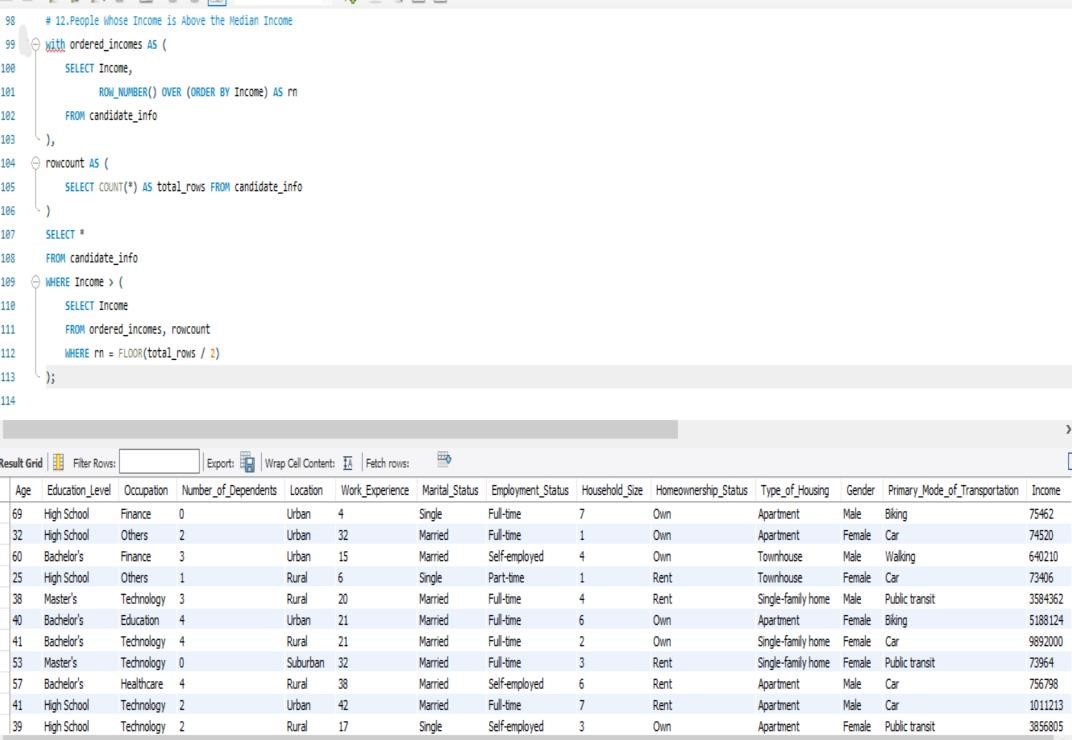
well

-

off

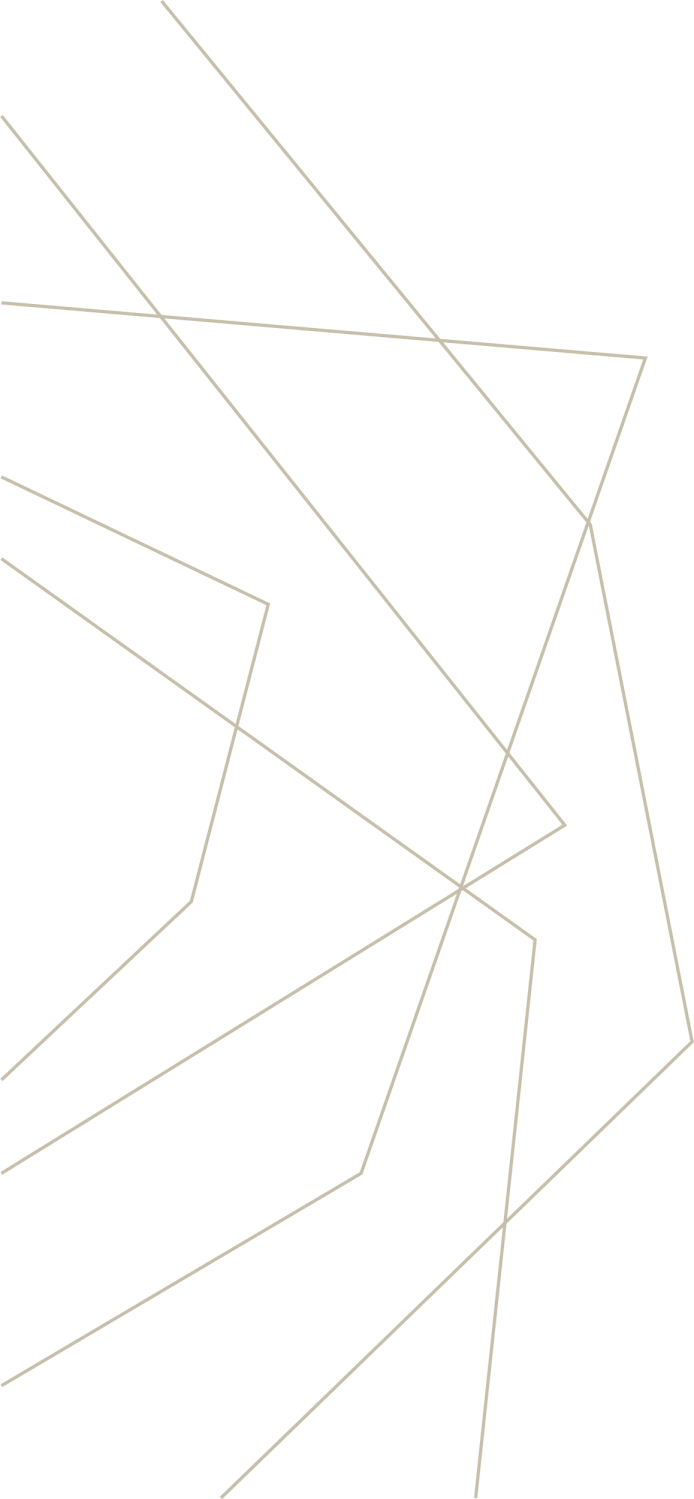
candidates

.



# CONCLUSION

This project enabled a thorough exploration of customer demographics, income levels, and lifestyle patterns using SQL. By querying the data, I uncovered insights into how factors like education, employment, and housing relate to financial outcomes. The results reveal key trends in gender distribution, urban wealth concentration, and occupational income. These insights can be applied in real-world scenarios such as strategic marketing, product targeting, and policy formulation.



THANK YOU

20

XX

Pitch Deck