

**Healthcare Stroke Prediction**

**SQL Project**

**December 7, 2022**

**Data Associate-03**

**PRN: 2201207101**

**Guide by: Miss. Vaishnavi Satav Ma’am**

**Name: Neha Parameshwar Sonavane**

**Course Name: Data Associate-03**

**Project Title: Healthcare stroke prediction**

**Stroke prediction dataset**

**Problem statement: According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.**

**Aim: This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status.**

**(Each row in the data provides relevant information about the patient.)**

Dataset: **https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset**

**Attribute Information**

1) id: unique identifier

2) gender: "Male", "Female" or "Other"

3) age: age of the patient

4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension

5) heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease

6) ever\_married: "No" or "Yes"

7) work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"

8) Residence\_type: "Rural" or "Urban"

9) avg\_glucose\_level: average glucose level in blood

10) bmi: body mass index

11) smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*

12) stroke: 1 if the patient had a stroke or 0 if not

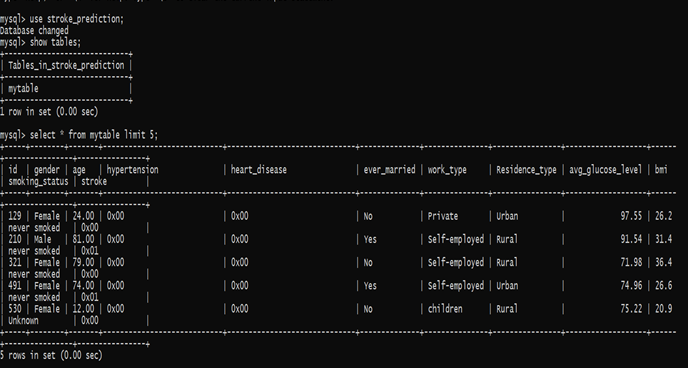
\*Note: "Unknown" in smoking\_status means that the information is unavailable for this patient

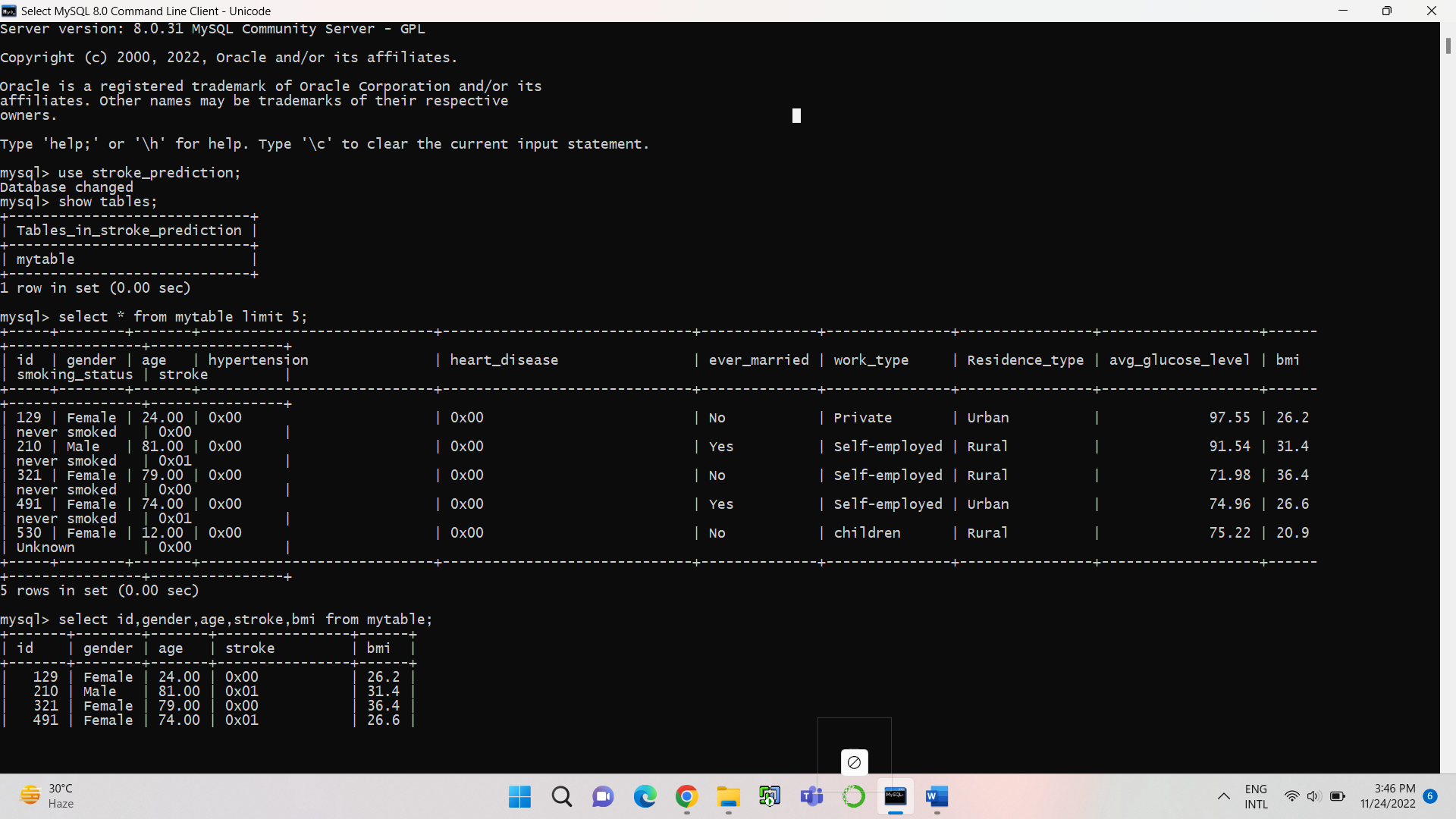
Here I use MYSQL and Excel for the analysis of the data

First, I load the data in MYSQL and then start doing an analysis the graphs (pie chart and bar charts are drawn by using Excel)

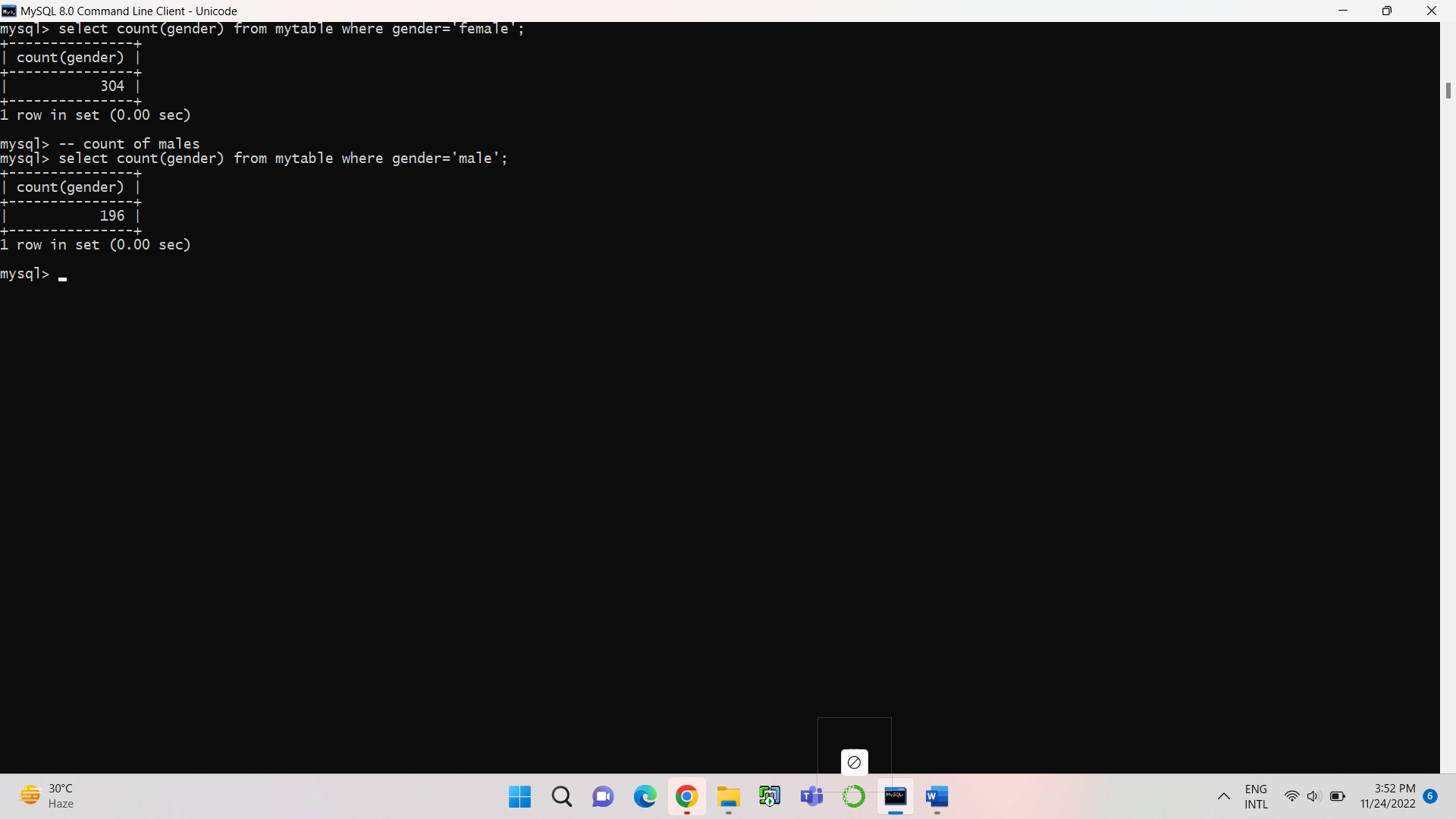
Here the data contains 500 rows and 12 columns.

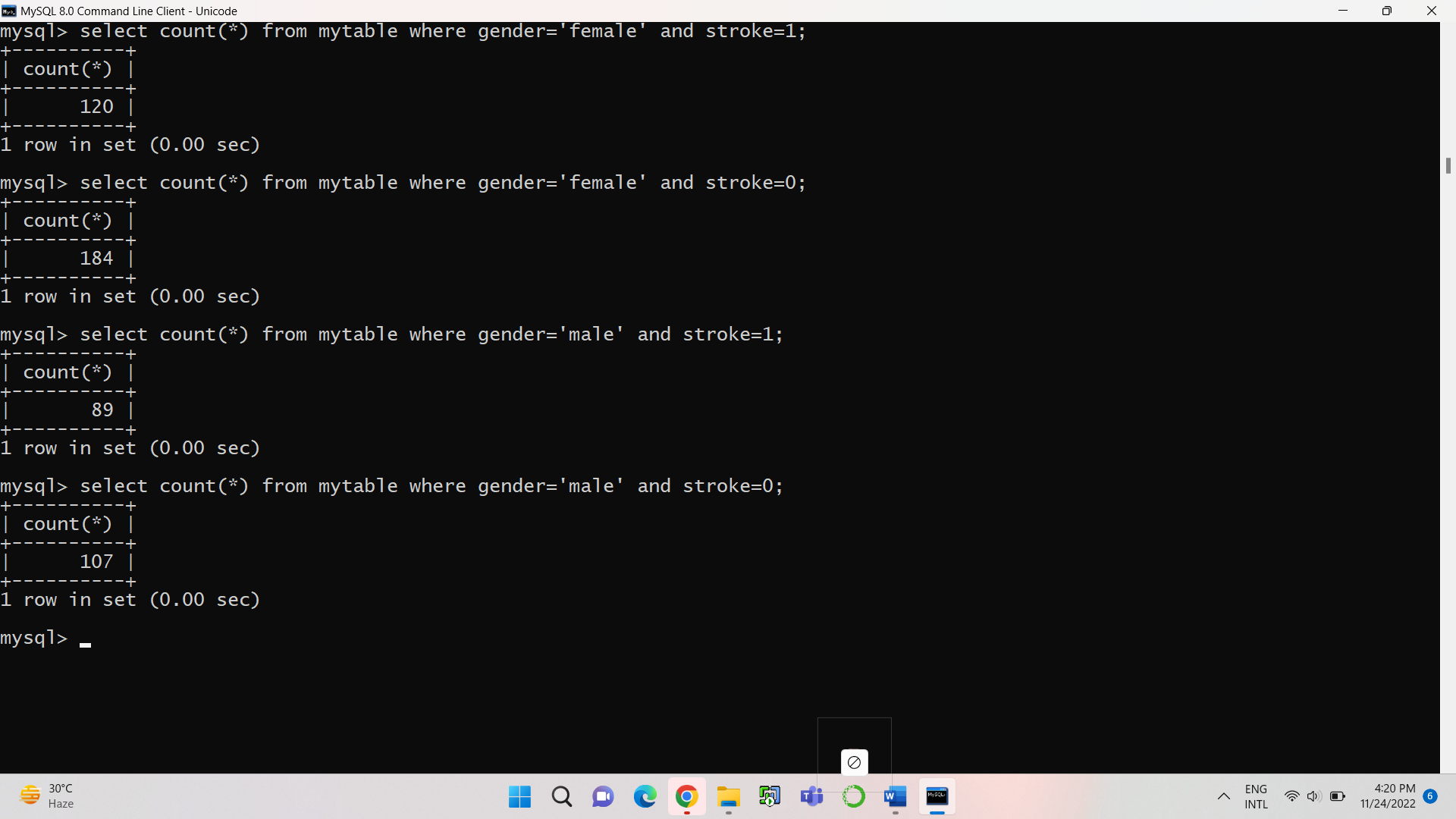
Command: load data outfile into path fields terminated by ‘,’





**#Here we can see that in our data there are 304 female patients and 196 male patients**

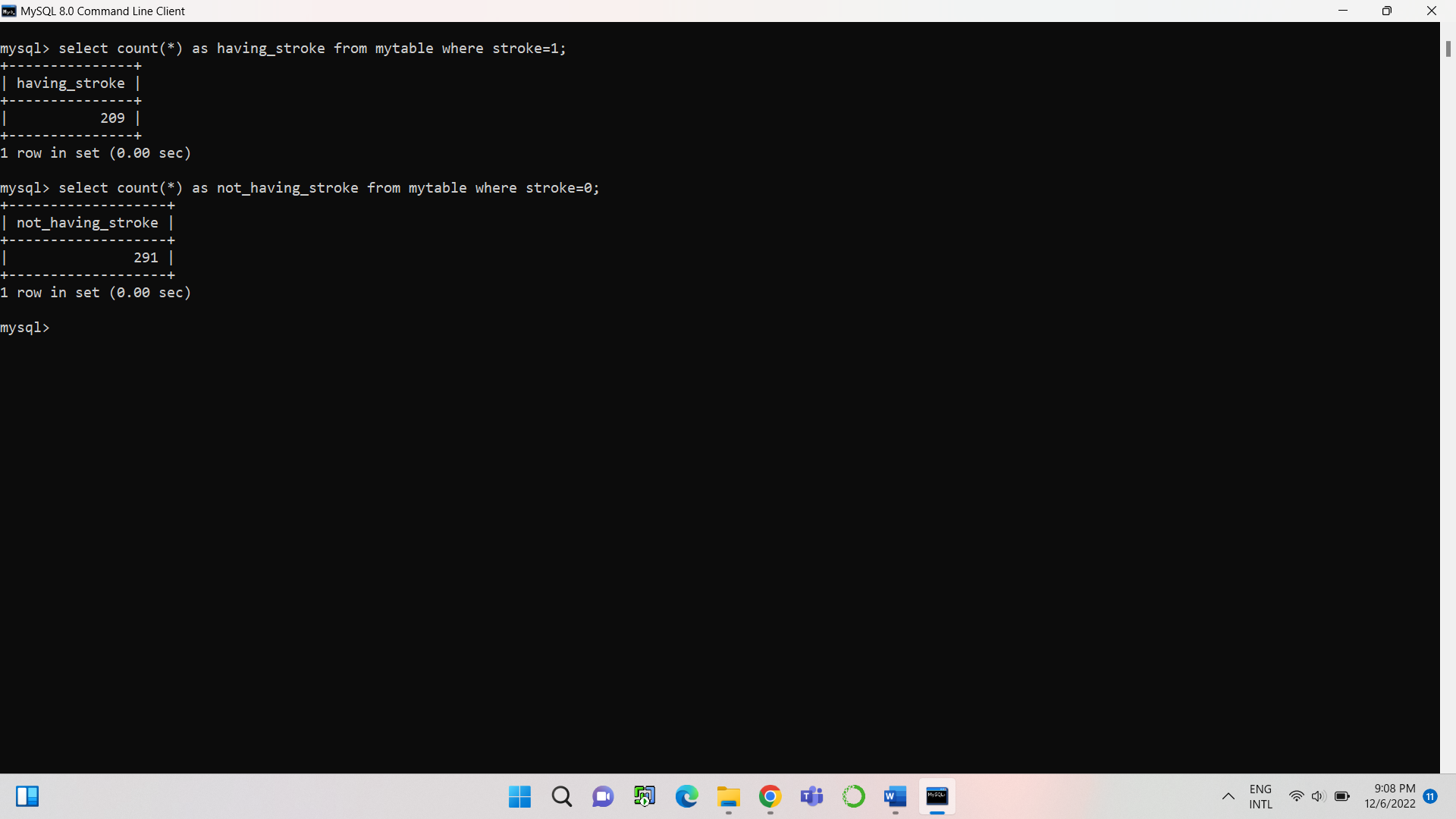


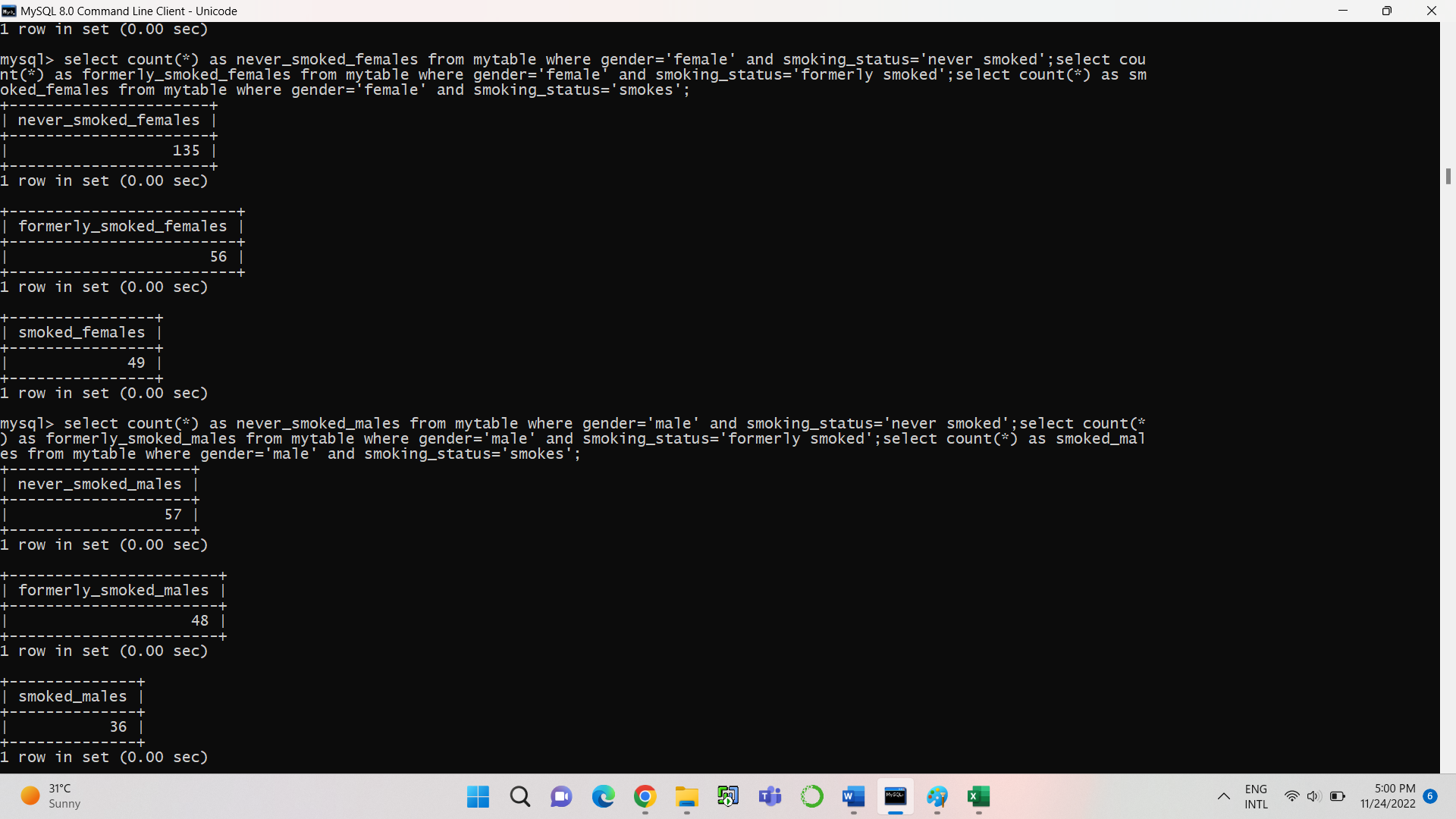


**#Here we can see that there are 120 females who were having stroke and 184 not having stroke also we can see that out of 196 males 107 not having stroke and 89 having stroke.**

|  |  |
| --- | --- |
| **females** | **stroke count for Females** |
| **yes** | **120** |
| **no** | **184** |

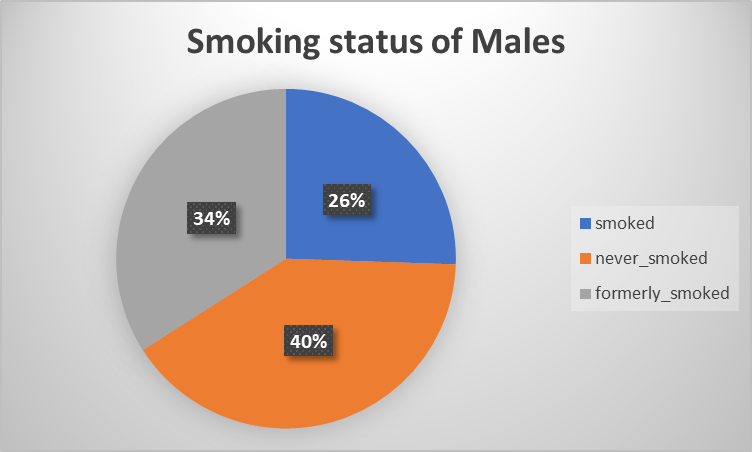
|  |  |
| --- | --- |
| **males** | **stroke count for Males** |
| **yes** | **89** |
| **no** | **107** |



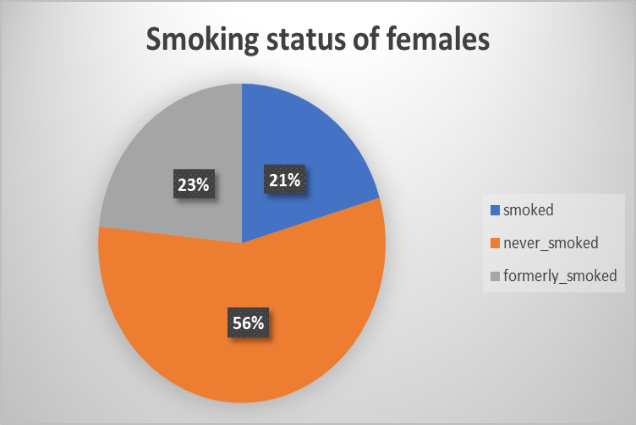


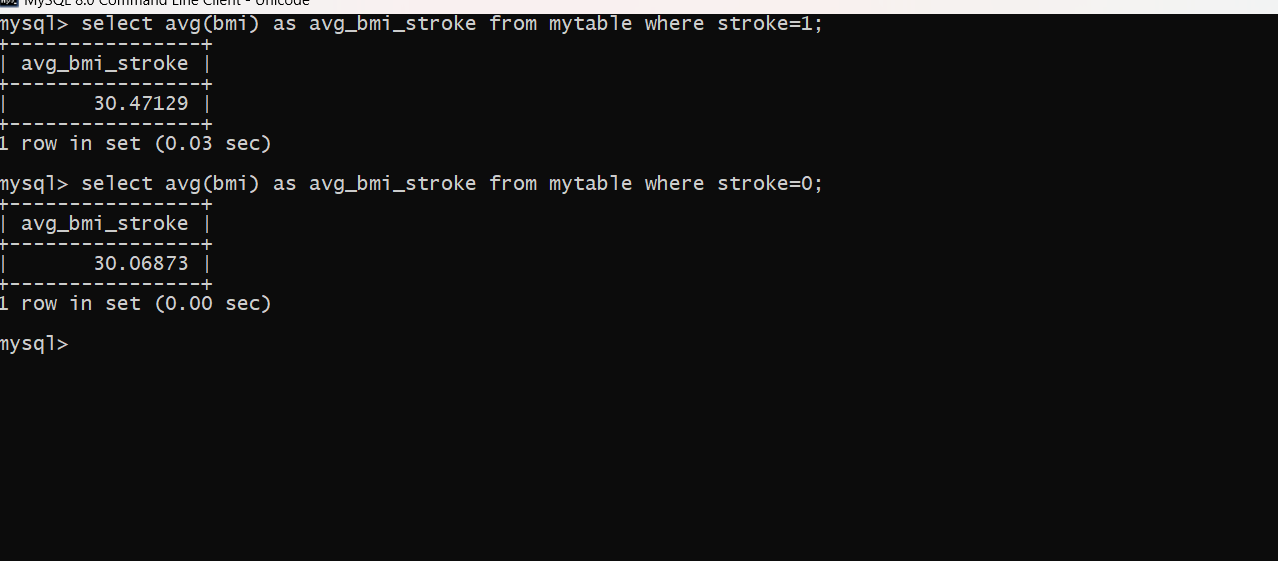
**#Total no. of peoples not having stroke out of 500 are 209 and having stroke are 291. So, we can say that the people who are having stroke are more in number.**

|  |  |
| --- | --- |
| **smoking status** | **Number of males** |
| **smoked** | **36** |
| **never\_smoked** | **57** |
| **formerly\_smoked** | **48** |

****

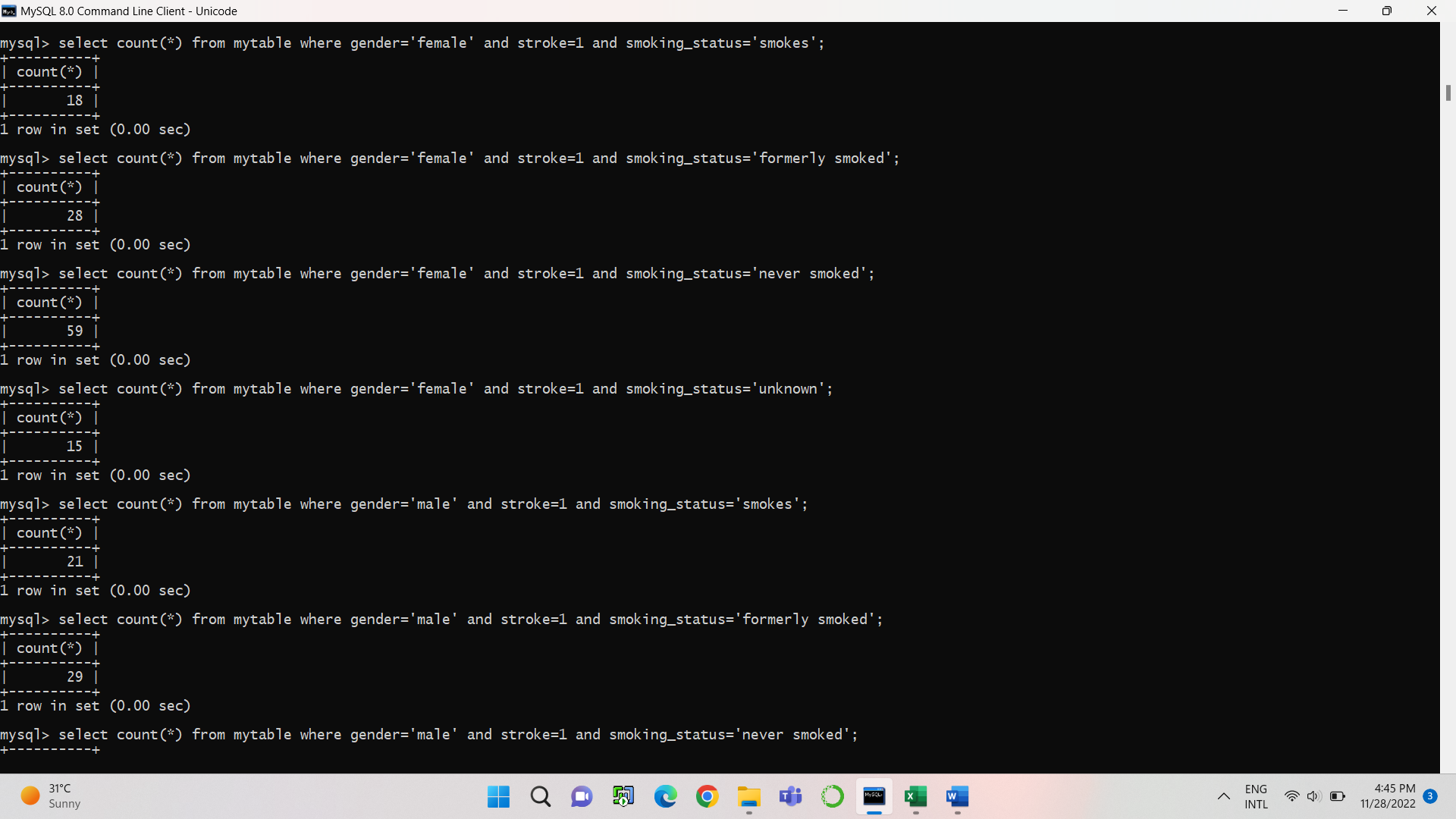
|  |  |
| --- | --- |
| **smoking status** | **Number of females** |
| **smoked** | **49** |
| **never\_smoked** | **135** |
| **formerly\_smoked** | **56** |

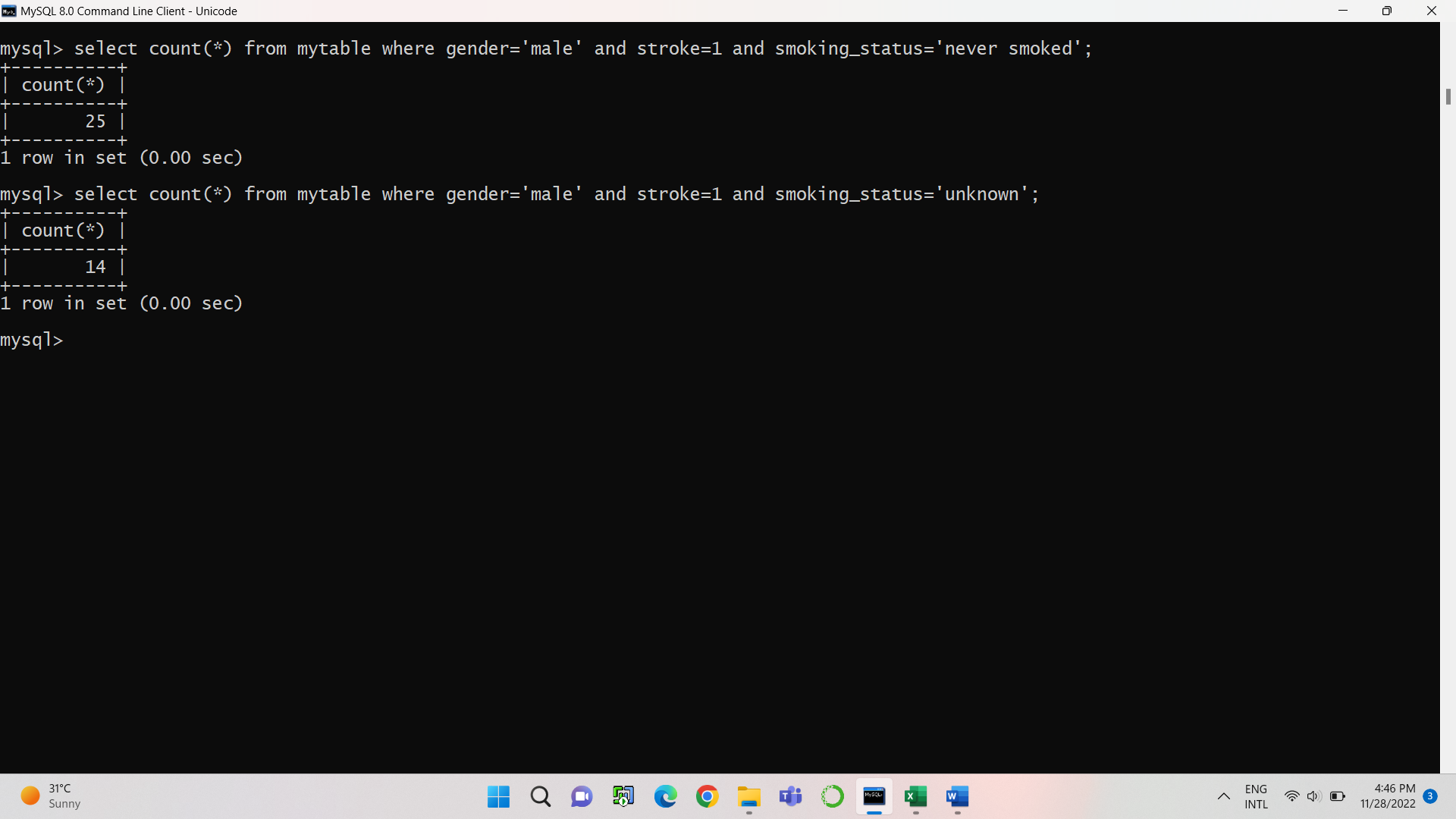
****

****

**#We can say that the patients who are having more bmi are more chances of having stroke**

|  |  |
| --- | --- |
| **stroke** | **avg\_bmi** |
| **yes** | **30.47129** |
| **no** | **30.06873** |





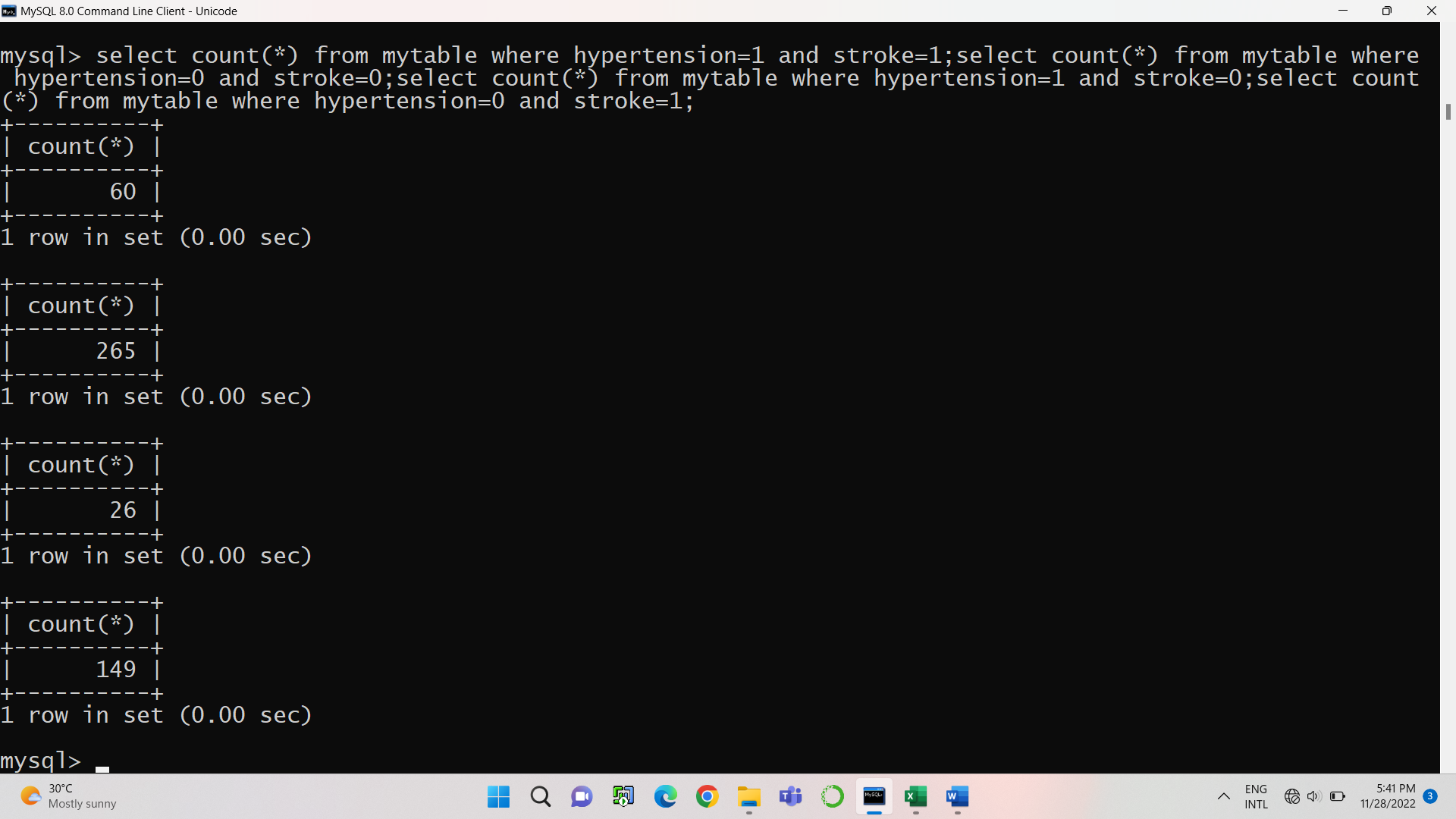
|  |  |
| --- | --- |
| **Smoking Status of Female patients** | **Number of Female patients having stroke** |
| **Smokes** | **18** |
| **Never Smoked** | **59** |
| **Formerly Smoked** | **28** |
| **Unknown** | **15** |

|  |  |
| --- | --- |
| **Smoking Status of Male patients** | **Number of Male patients having stroke** |
| **Smokes** | **21** |
| **Never Smoked** | **25** |
| **Formerly Smoked** | **29** |
| **Unknown** | **14** |

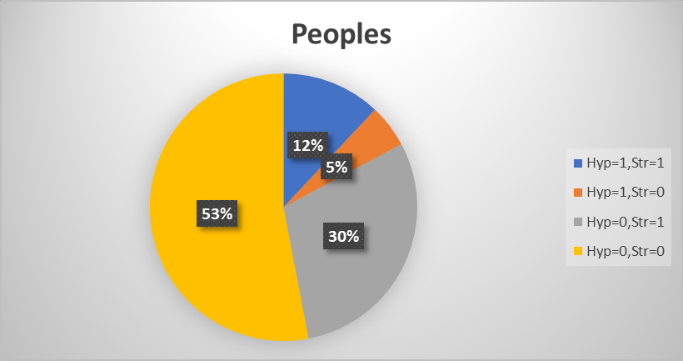
**#Here we can say that the female patients who are having stroke they don’t belong to any category they are having Unknown as smoking status and also the females who formerly smoked is more in percent that is 23% having stroke.**

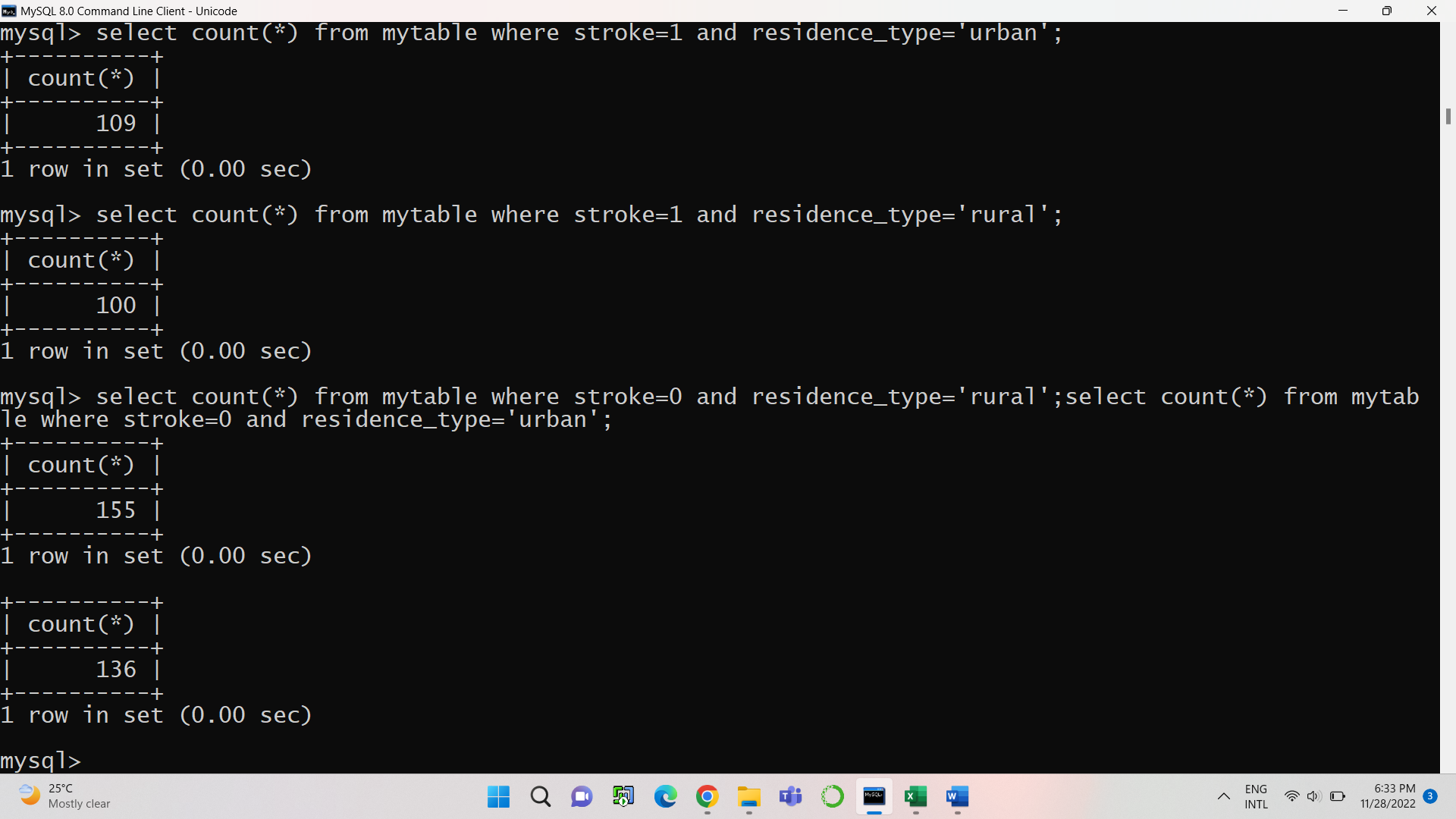
**#Here in Males also we see that the males who are having stroke their smoking status is formerly smoked that means the males who smoke formerly are more chances of having stroke.**

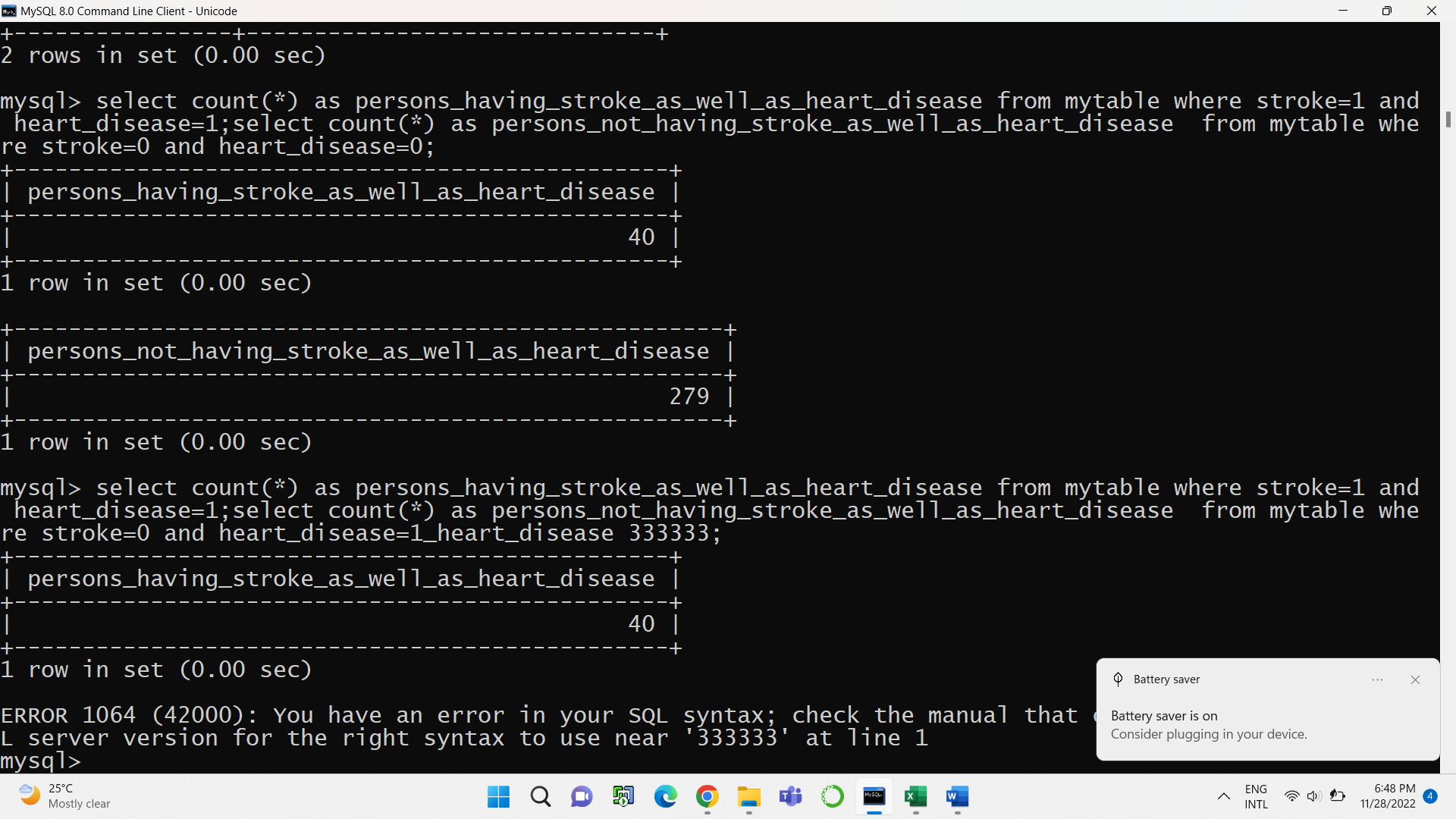
**#From Bar graph also, we can say that the males who smoke formerly are more chances of having stroke.**

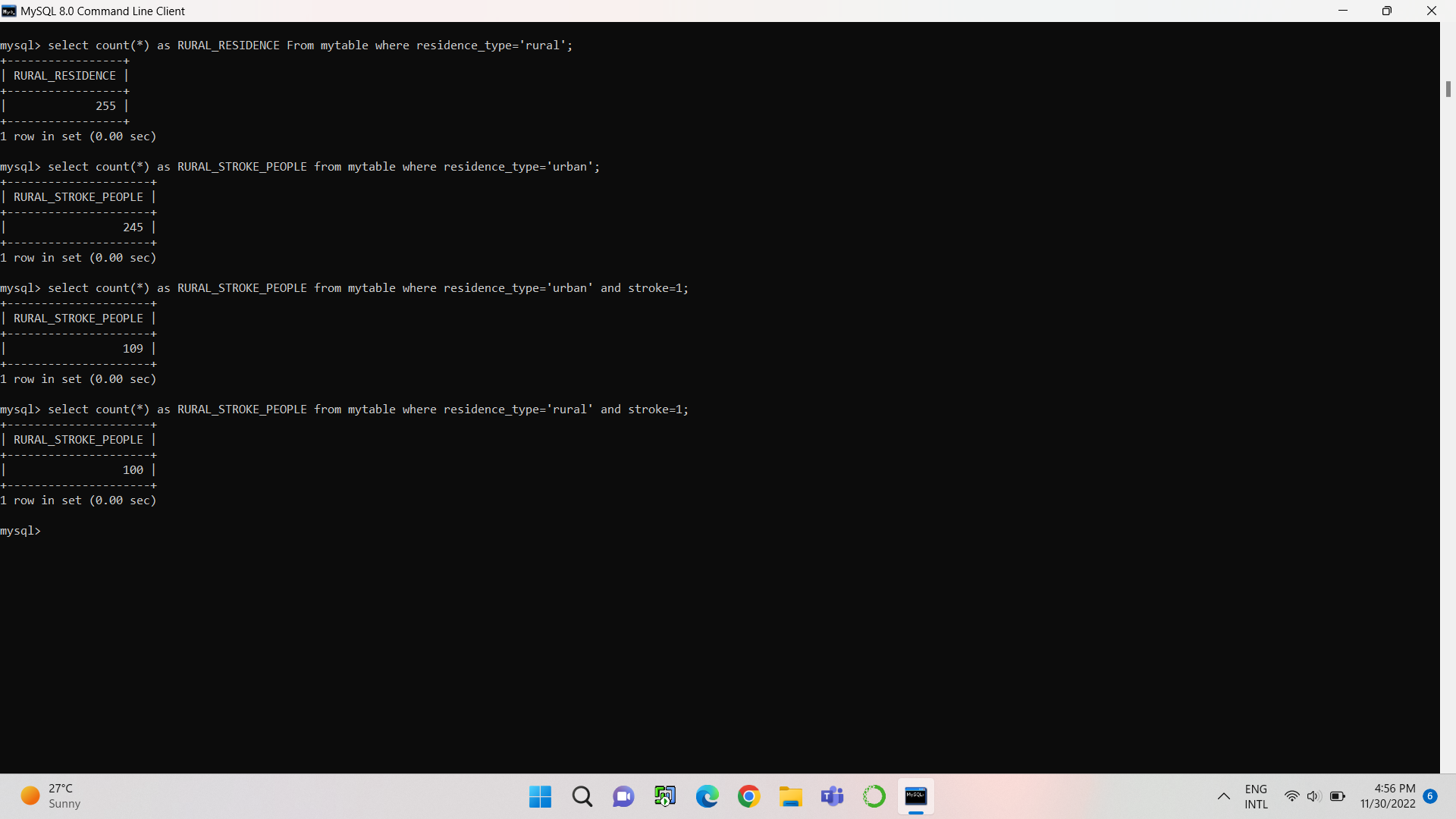


|  |  |
| --- | --- |
| **Classes** | **Peoples** |
| **Hyp=1, Str=1** | **60** |
| **Hyp=1, Str=0** | **26** |
| **Hyp=0, Str=1** | **149** |
| **Hyp=0, Str=0** | **265** |

****

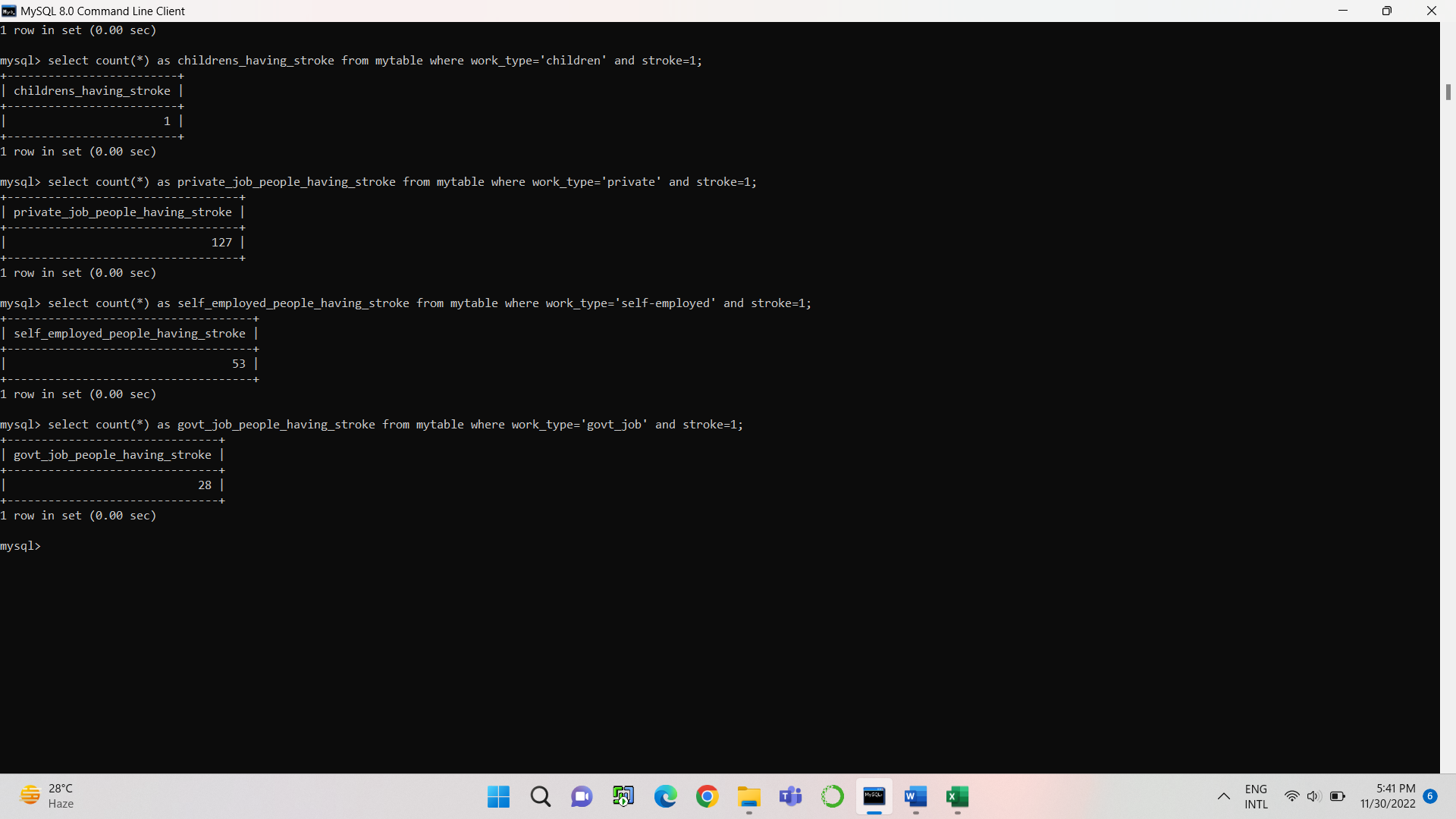




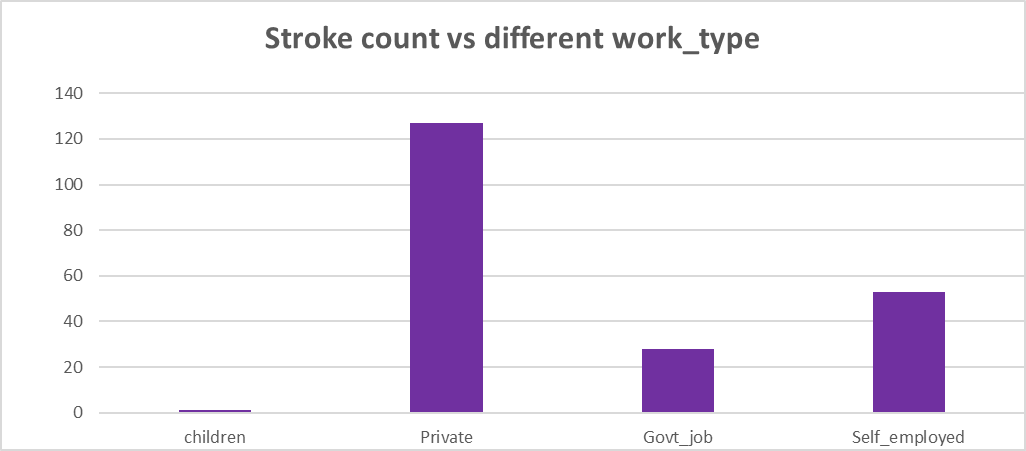


|  |  |
| --- | --- |
| **Residence type** | **stroke count residence wise** |
| **Rural** | **100** |
| **Urban** | **109** |

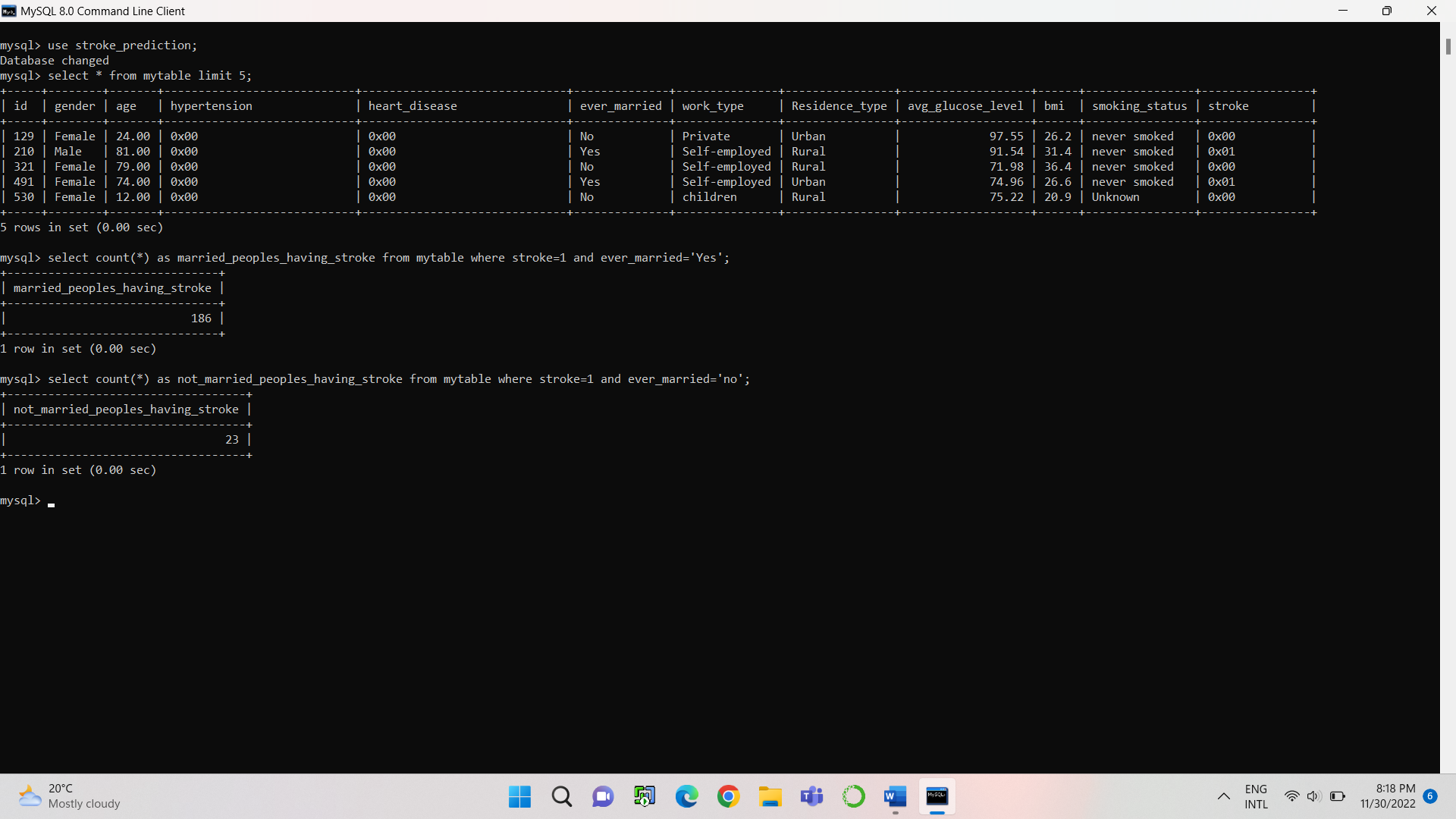
**# Here we clearly see that stroke percentage is more in Urban area rather than Urban area which is 52%.**



|  |  |
| --- | --- |
| **Work\_6type** | **Stroke count for different work\_type** |
| **children** | **1** |
| **Private** | **127** |
| **Govt\_job** | **28** |
| **Self\_employed** | **53** |

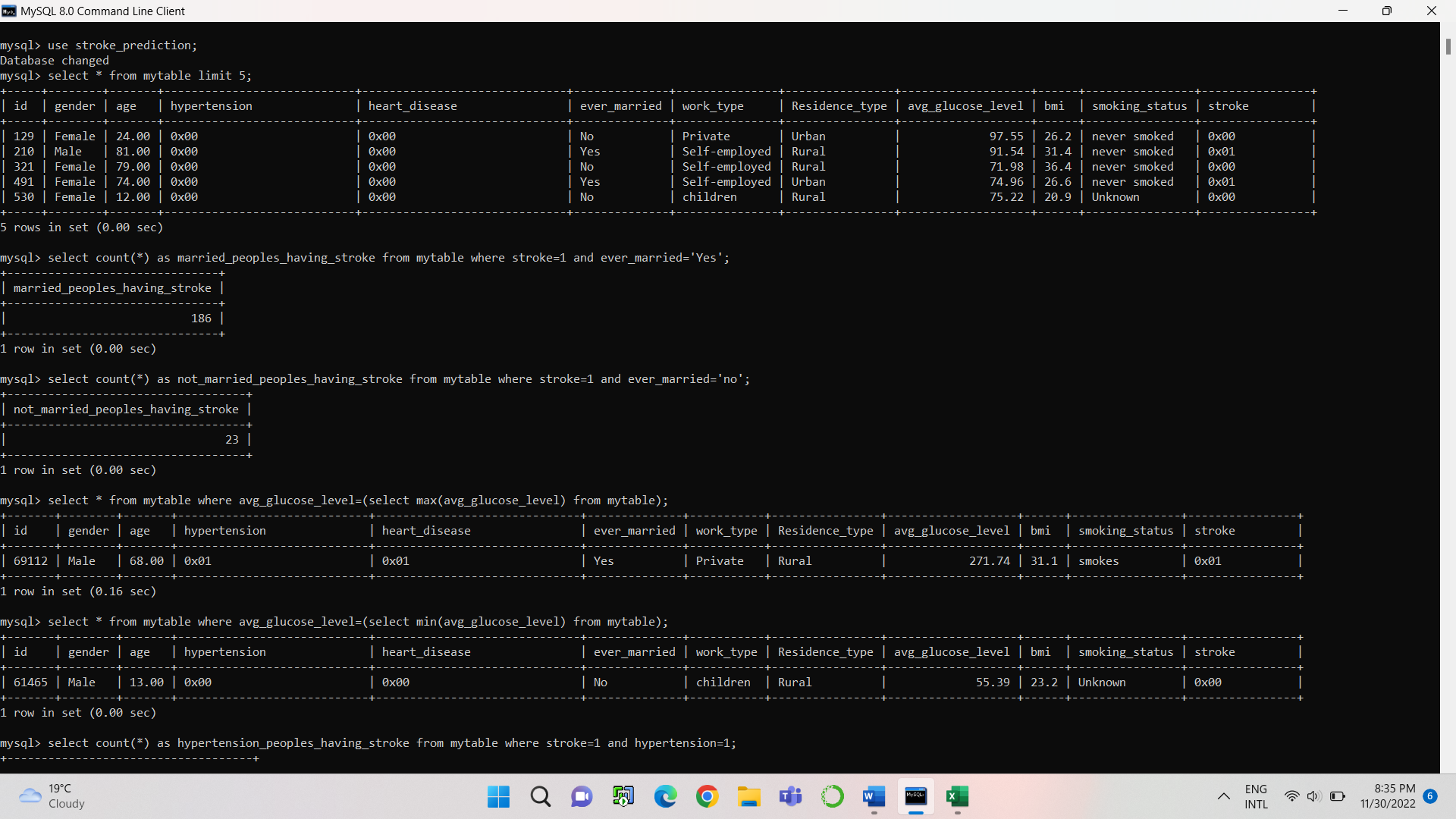
****

**#Here we clearly say that the stroke % is more in people who are self-employed which is 61%.**

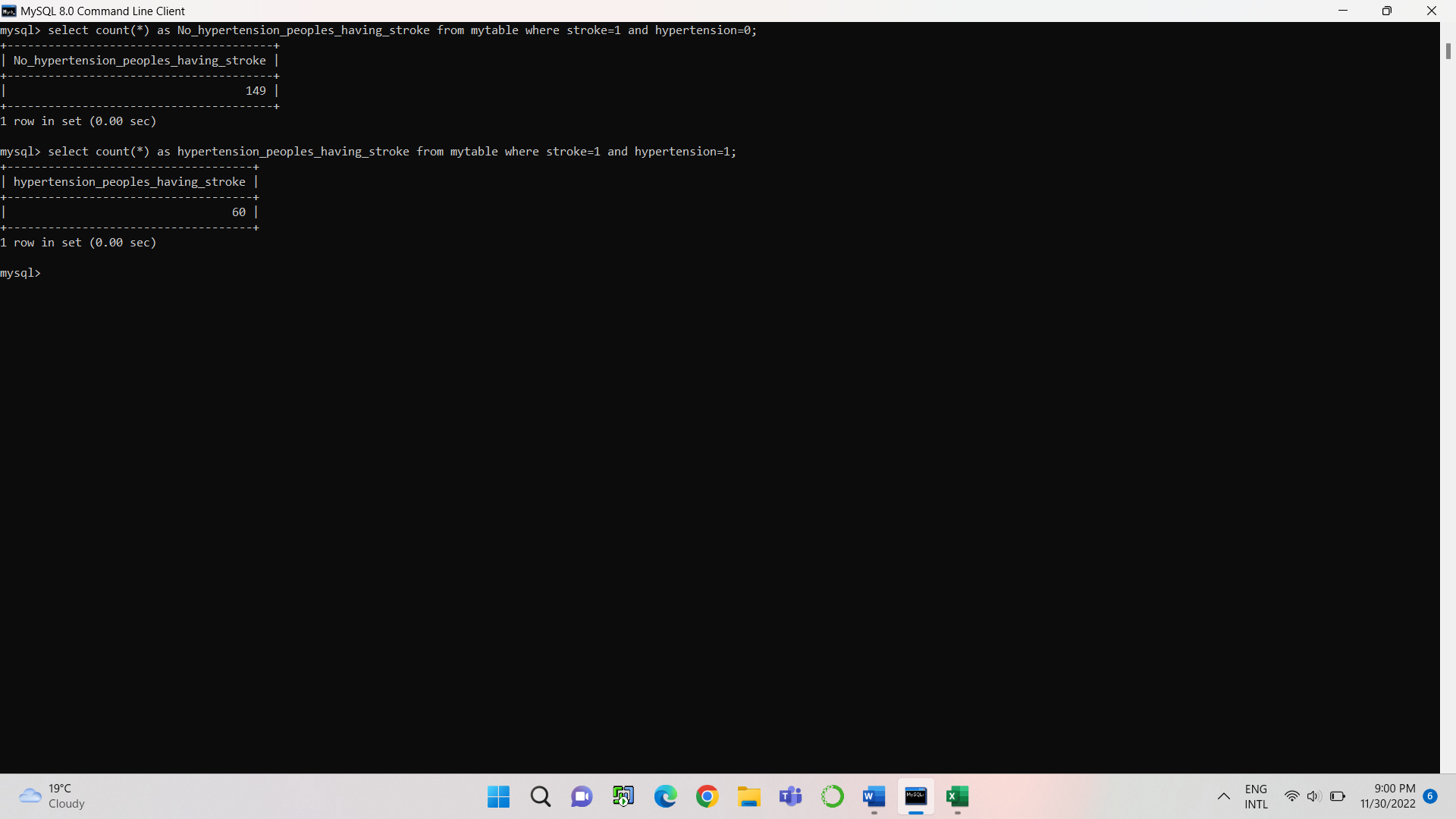


|  |  |
| --- | --- |
| **married\_status** | **stroke count for married status** |
| **Yes** | **186** |
| **No** | **23** |

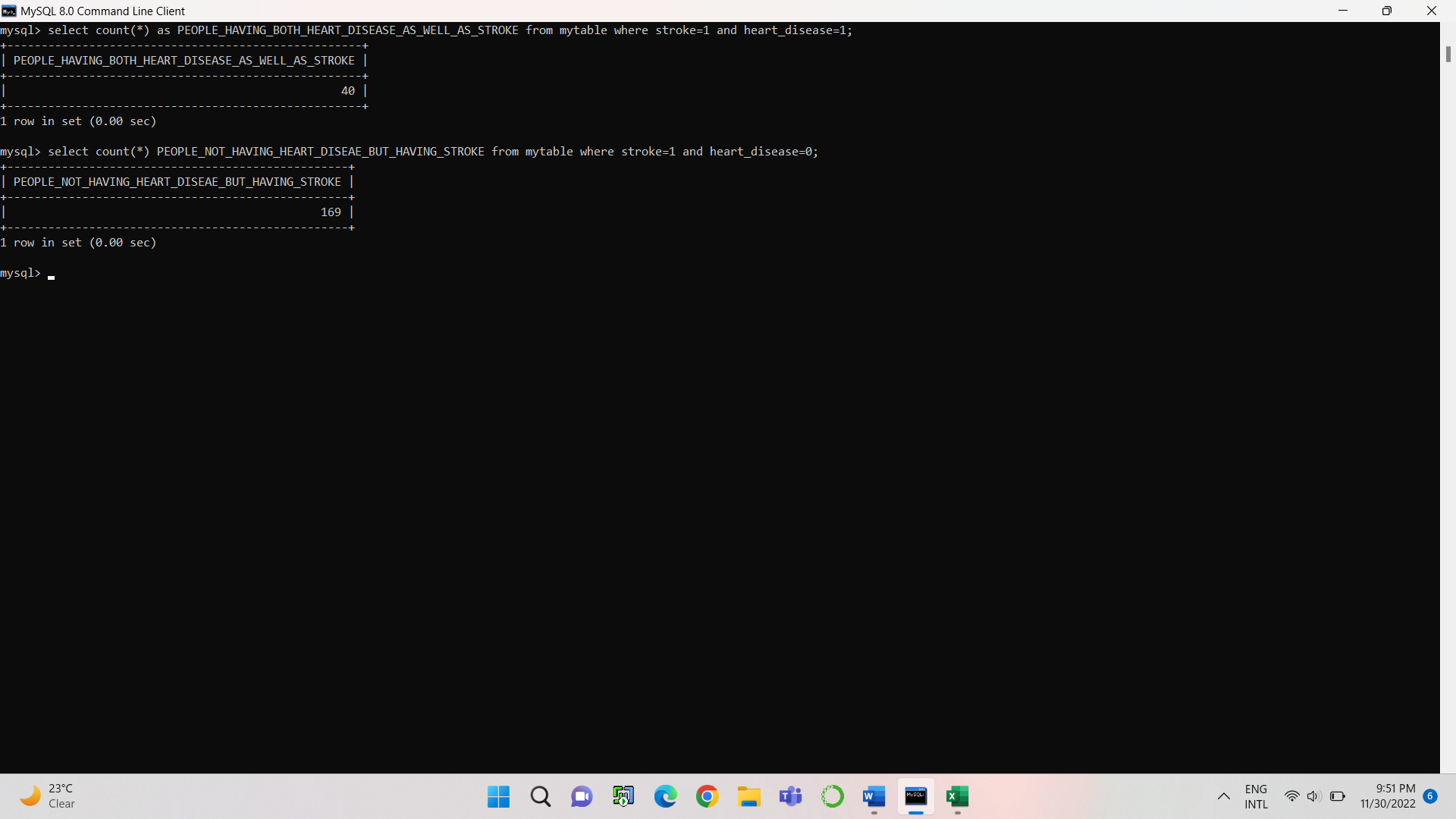
**#So, our data says that the peoples who are married are more chances of having stroke because their percentage is 89%.**



**#Here we see that the person who is having maximum average glucose level having stroke his age is more than the children who is having minimum average glucose level but that children doesn’t having stroke. Also, the person is doing private job and having more BMI than the children. They both belong to Rural area.**

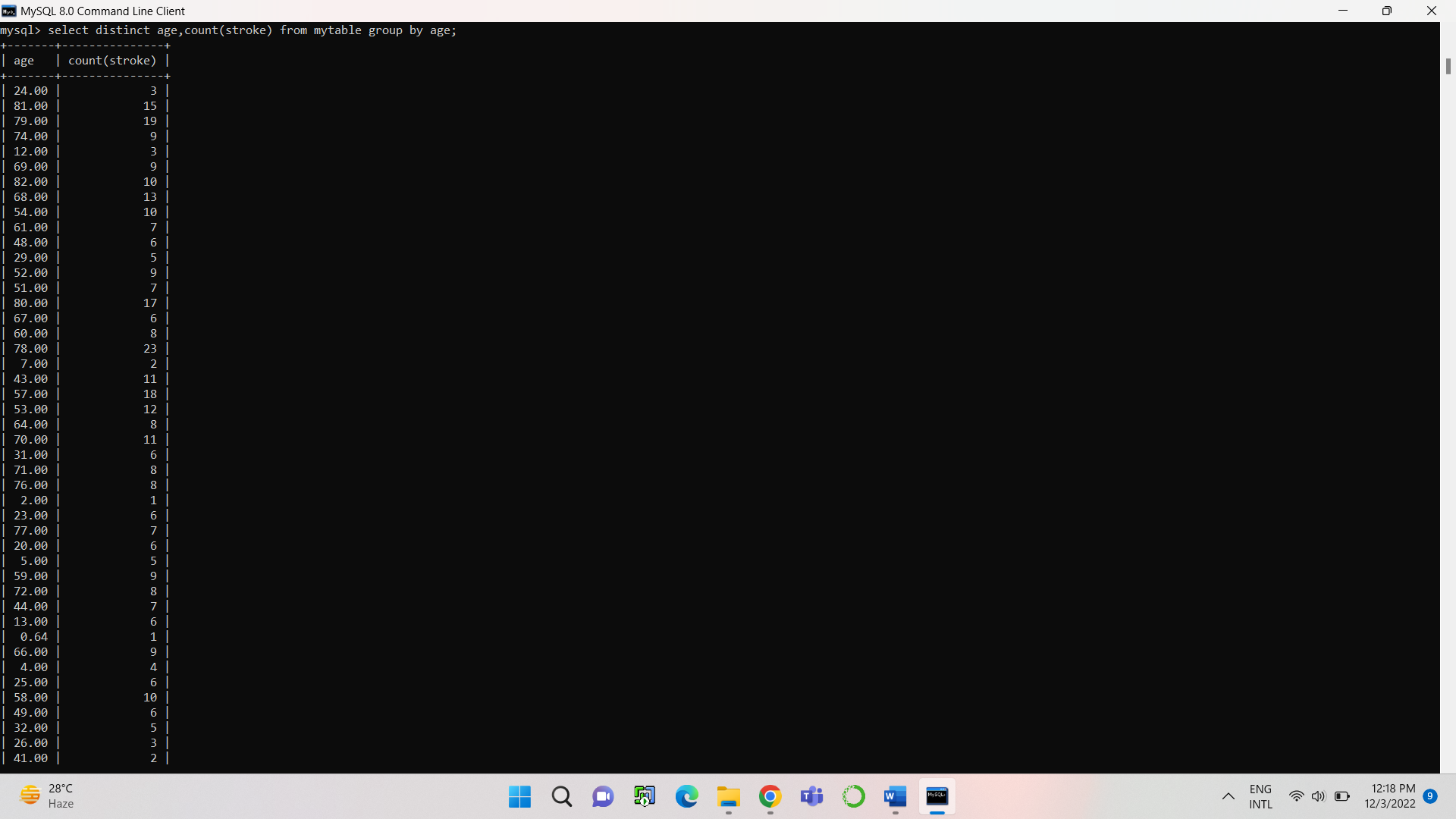


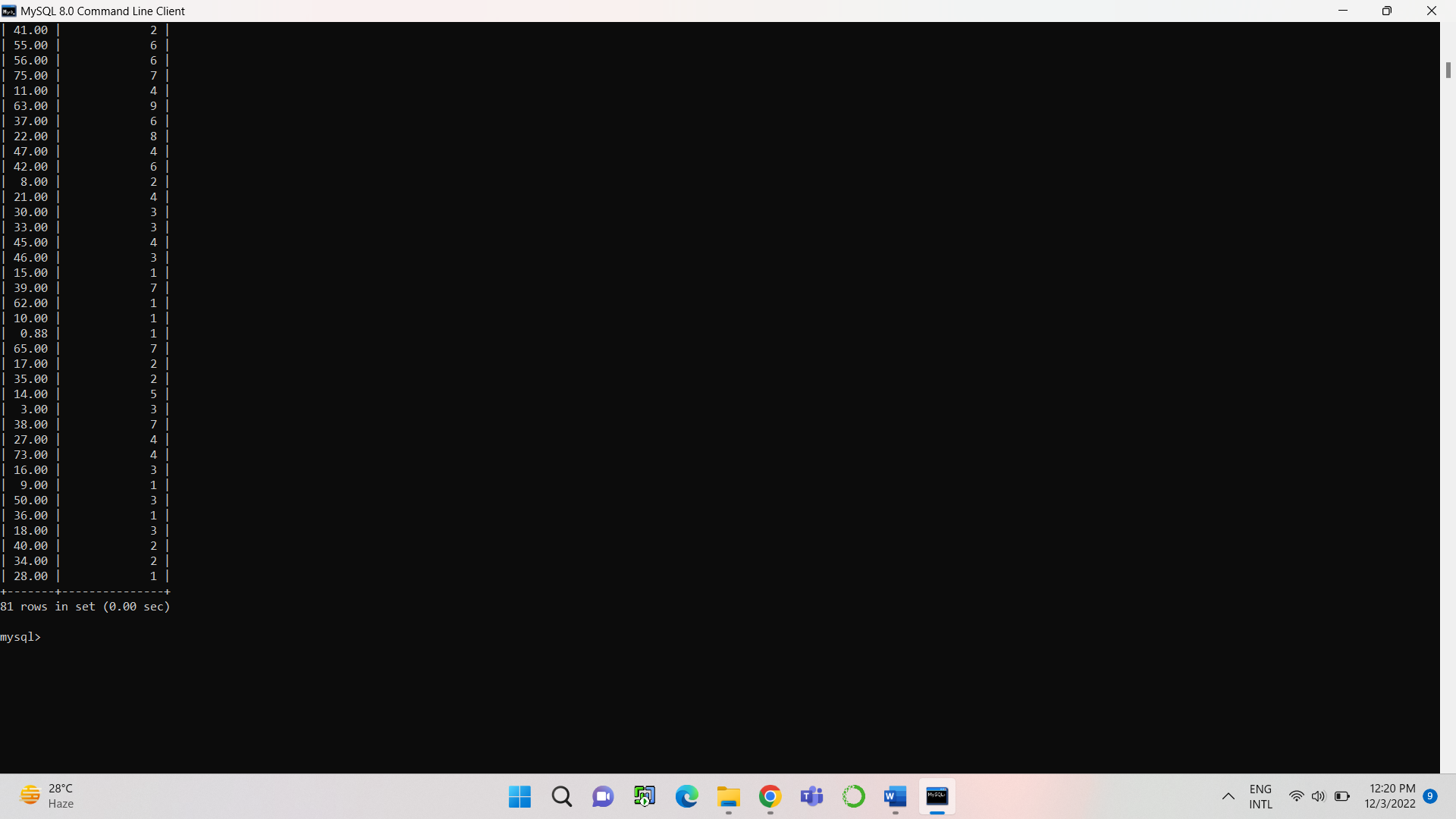
**# Our data says that those who are not having hypertension having more chances of having stroke rather than having hypertension.**



|  |  |
| --- | --- |
| **Disease** | **Number of Patients** |
| **Heart disease as well as Stroke** | **40** |
| **Only Stroke not heart disease** | **169** |

**s#Here we see that peoples who are not having heart disease are more chances of having stroke as the bar for the same is more.**



**#Here we clearly see that the age group where maximum stroke occurs is nothing but 78.**

**Conclusion**: The data contains 500 patients from analysis we saw that 209 patients are having stroke and the BMI for these patients is more from which the smoking status of females is unknown (that is not known) and the smoking status of males is formerly smoked. These patients are belonging to Urban area as the analysis says that the patients who are having stroke are more likely belongs to the Urban area.

These peoples are self-employed and they are married also they do not have hypertension and heart disease problem also their glucose level is high and having more age.

So clearly, we say that the factors which are responsible for the stroke are (As our data says)

1) Residence-type is Urban area

2) Work-type is Self-employed

3) High glucose level

4) High BMI

5) More Age

6) Marital Status: Married