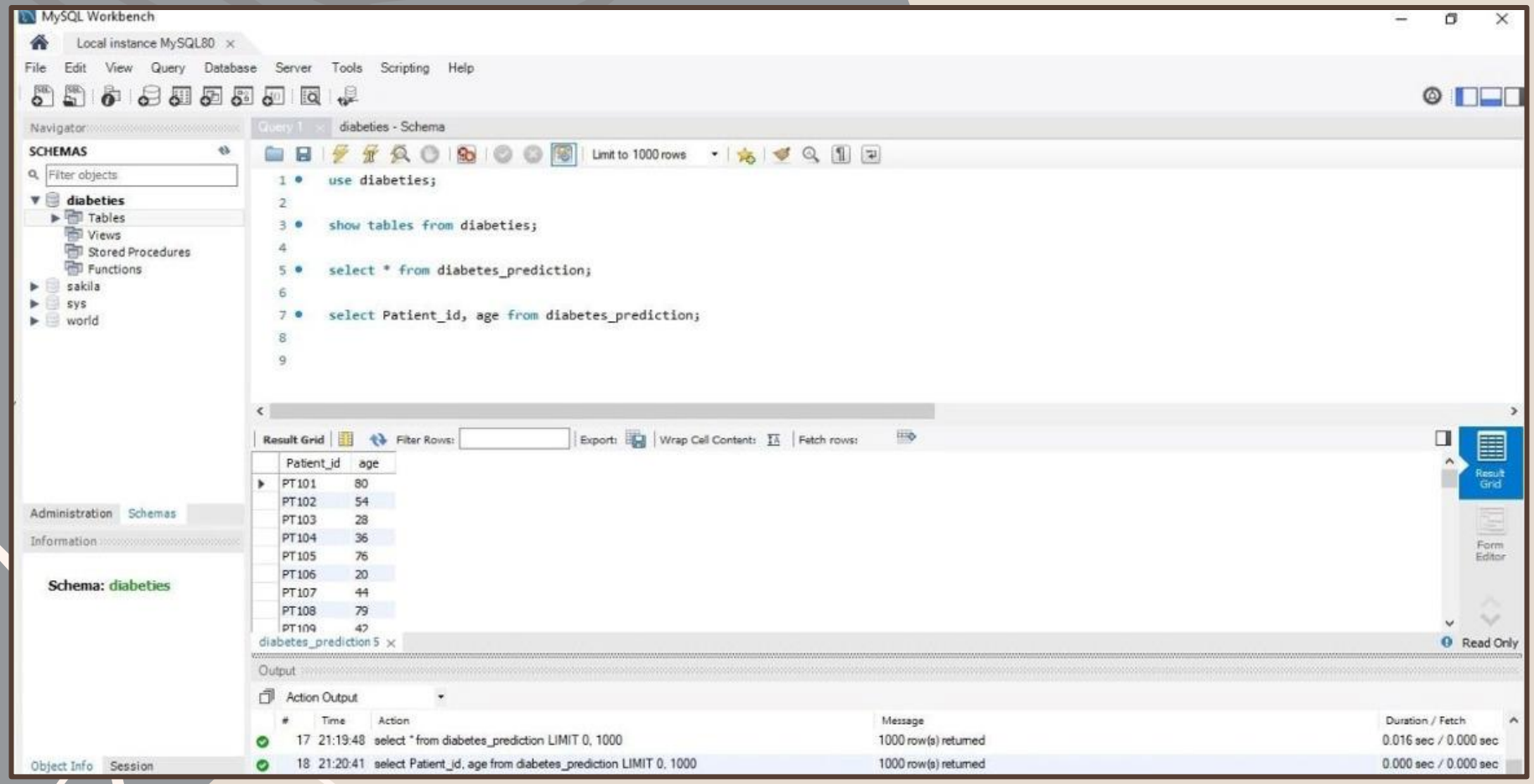
Vaishnavi C. Badgujar

DIABETES PREDICTION

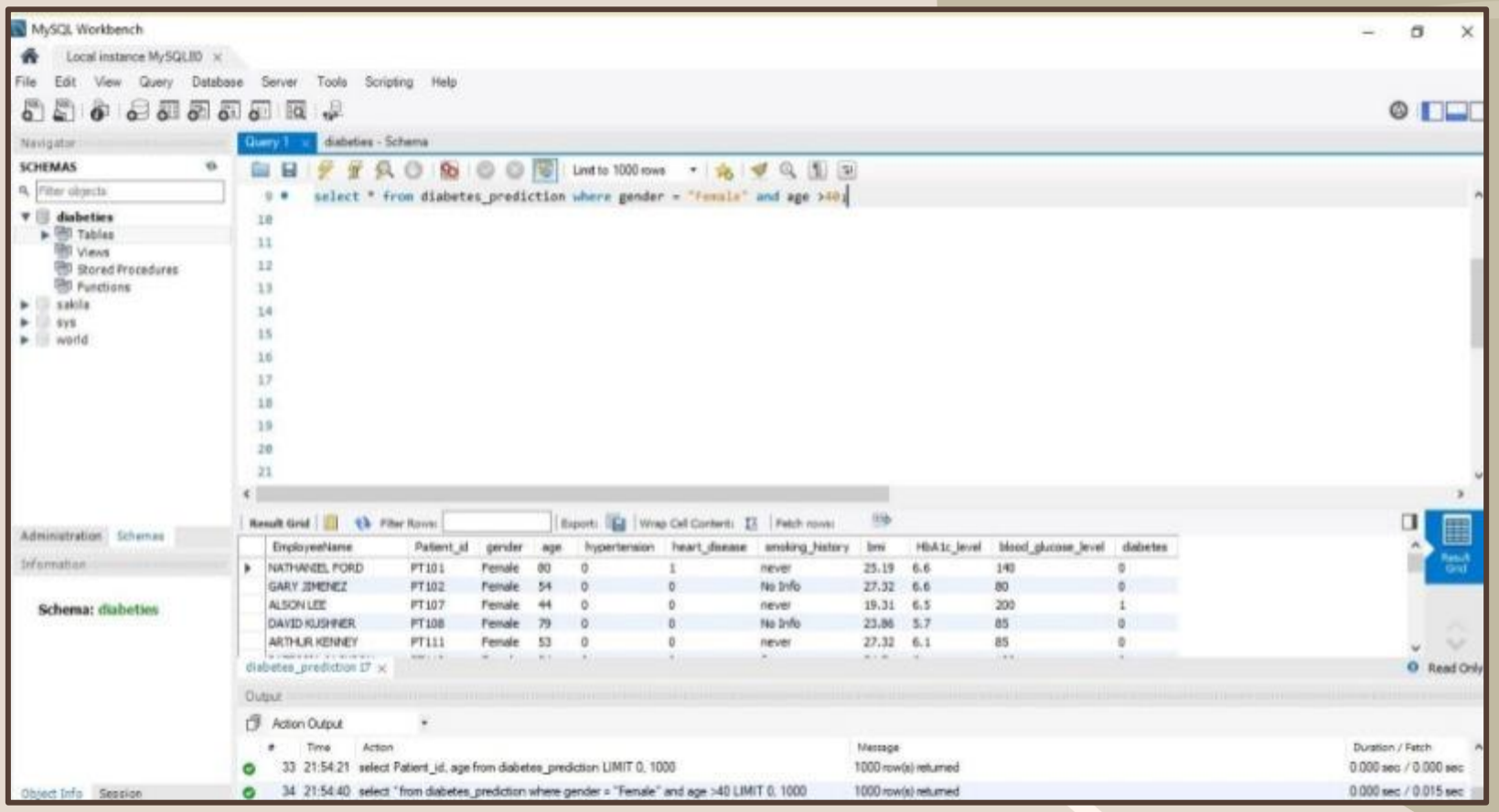
Using SQL

Psyliq Data Analyst Internship

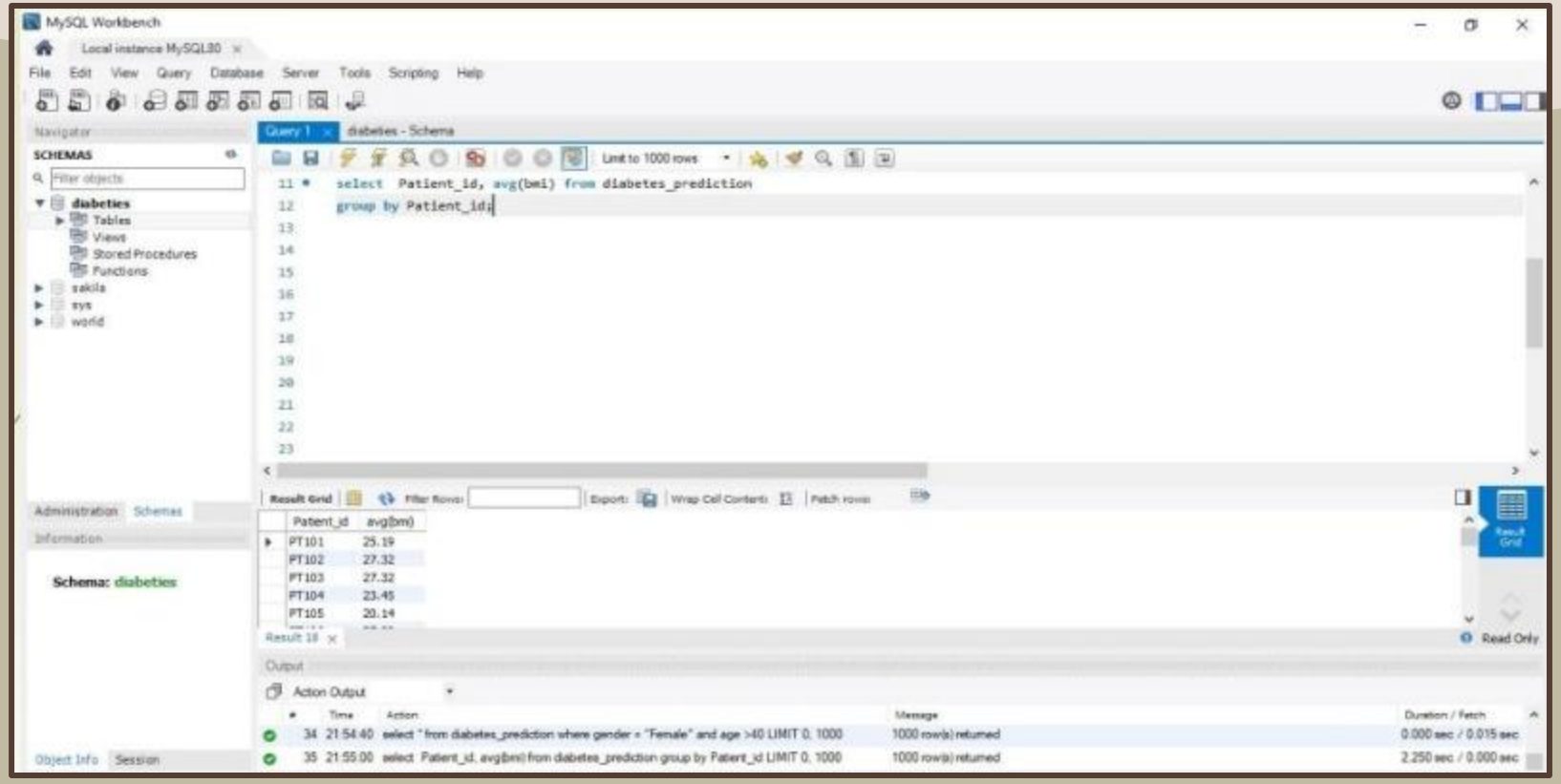
1. Retrieve the Patient_id and ages of all patients.



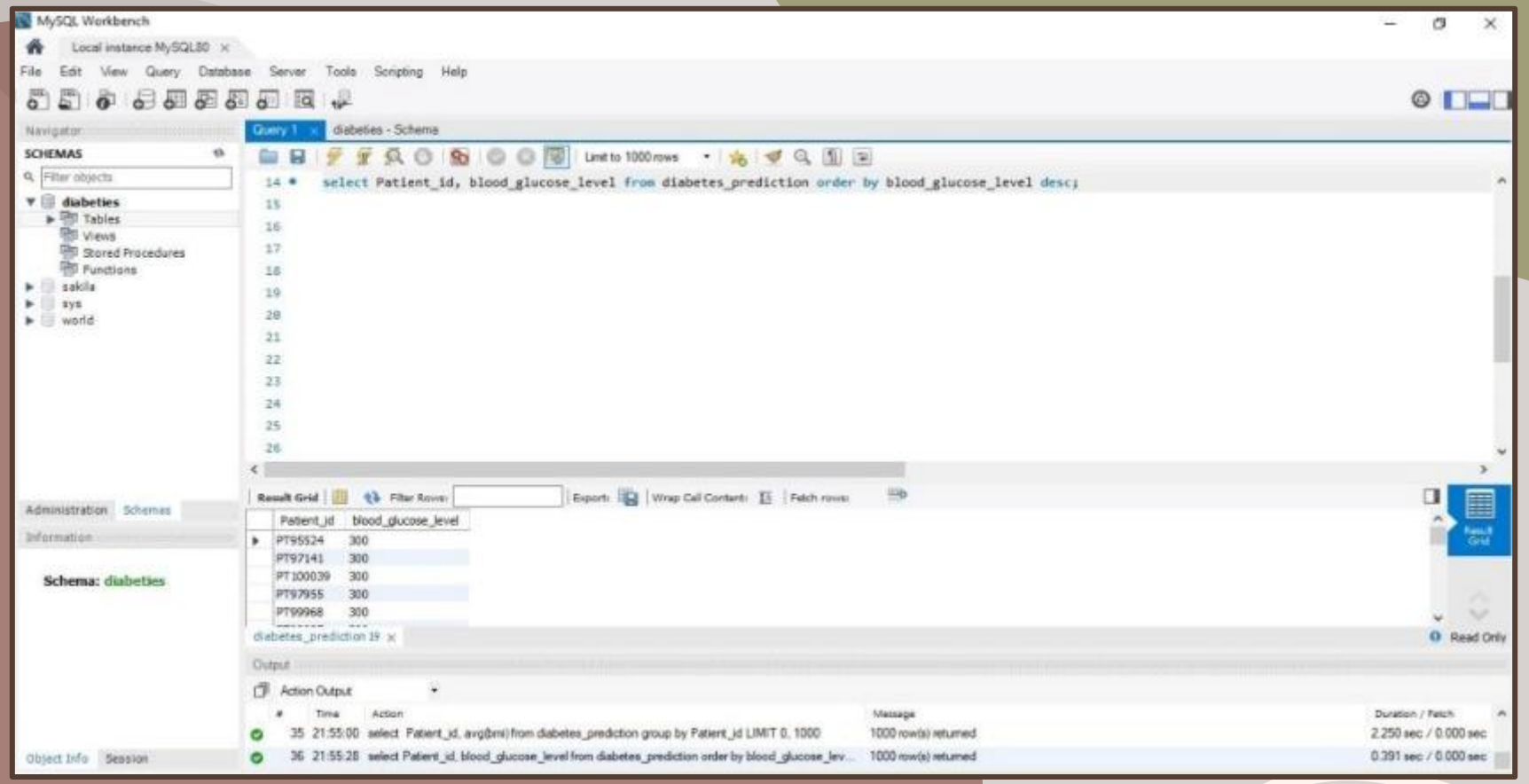
2. Select all female patients who are older than 40.



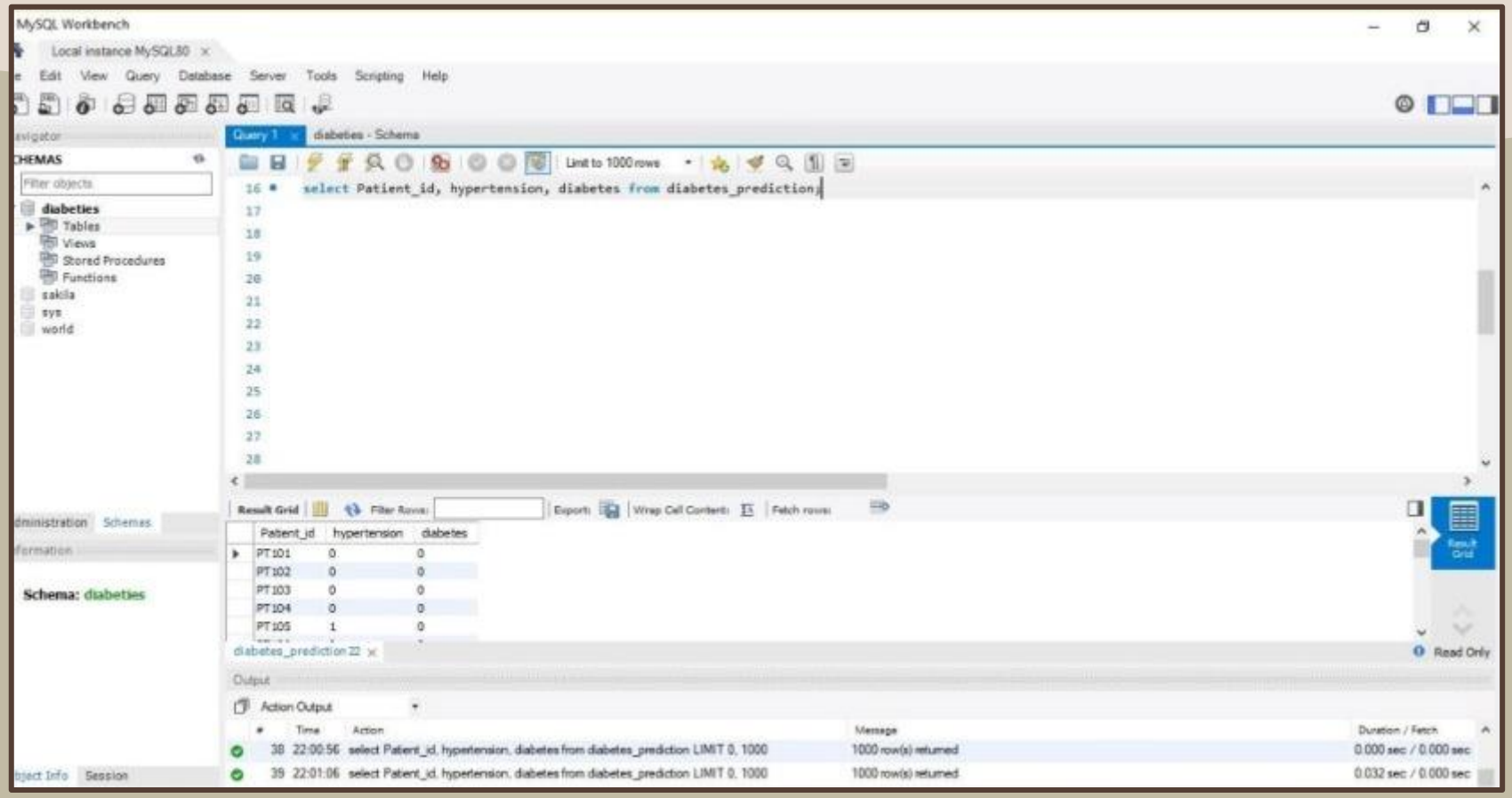
3. Calculate the average BMI of patients.



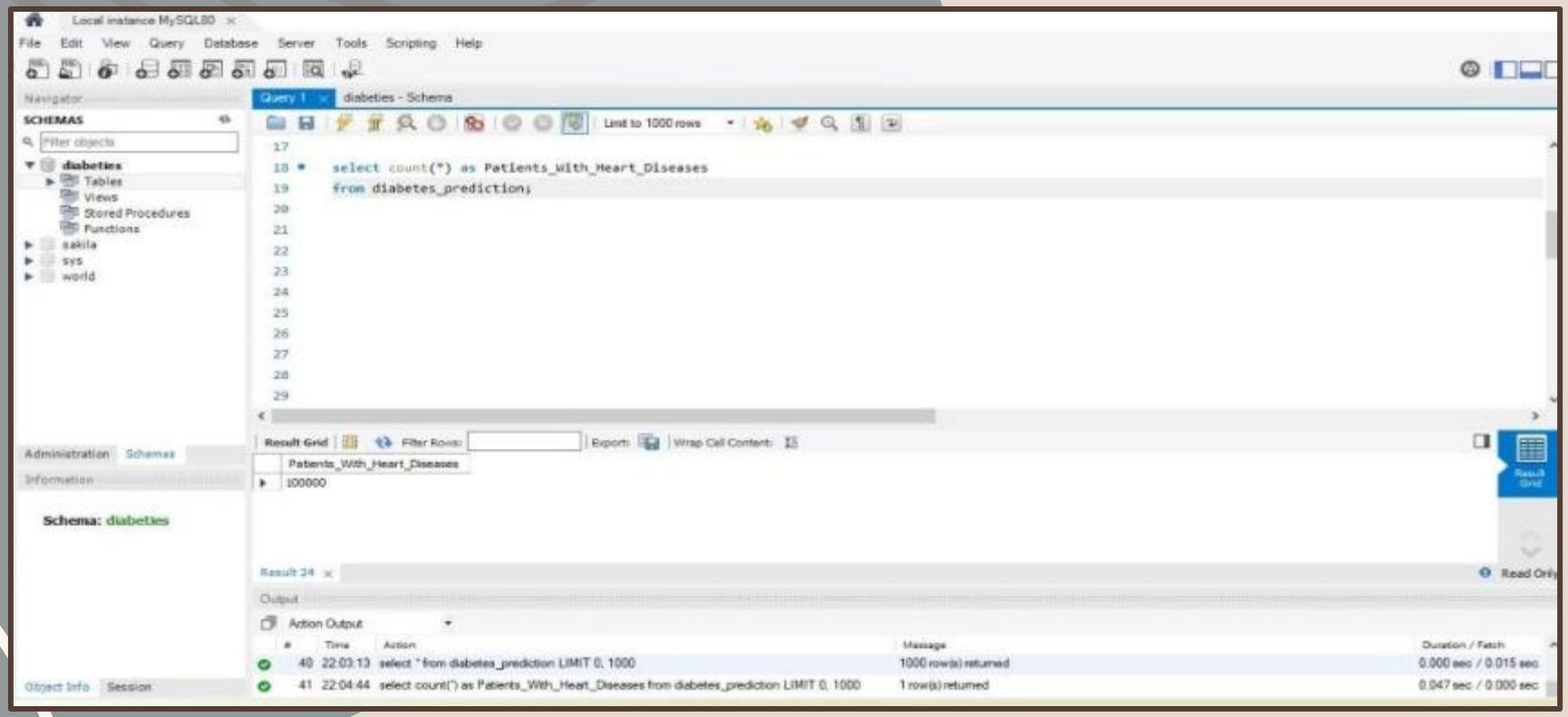
4. List patients in descending order of blood glucose levels.



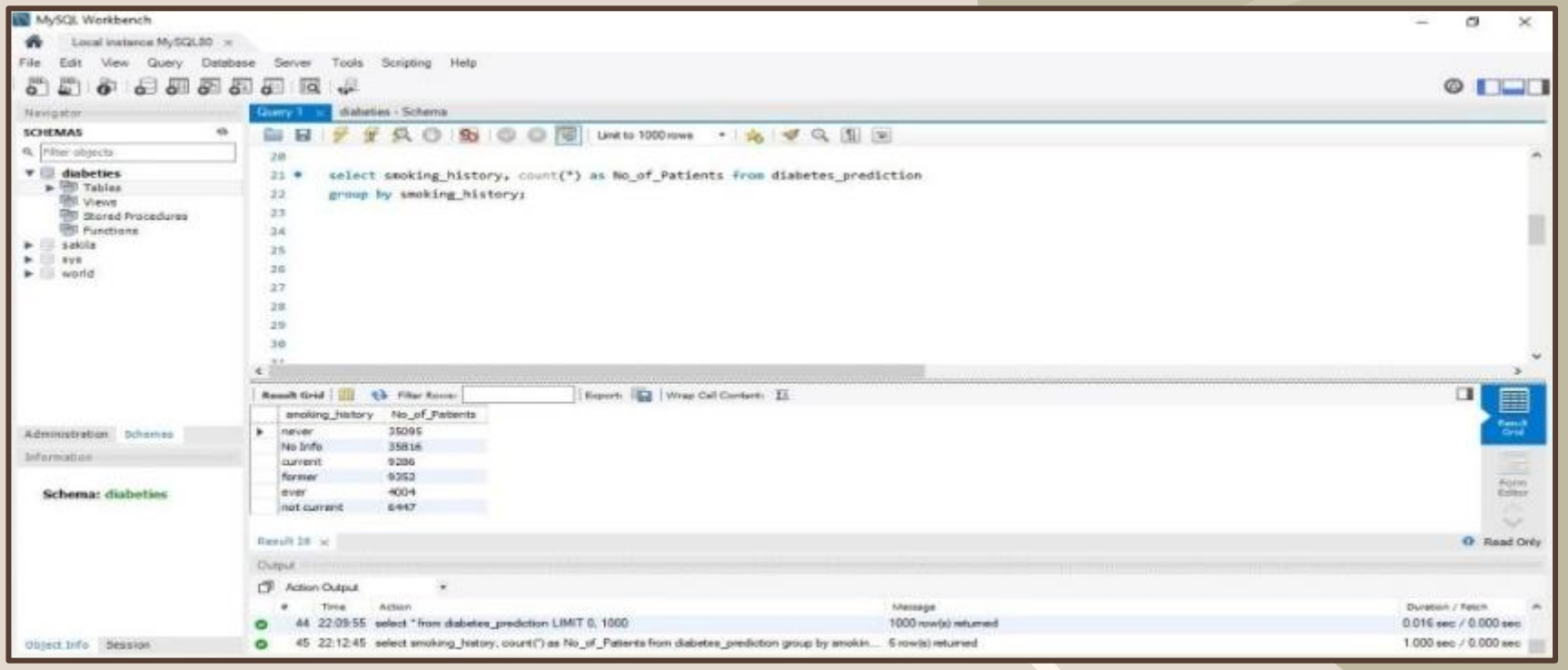
5. Find patients who have hypertension and diabetes.



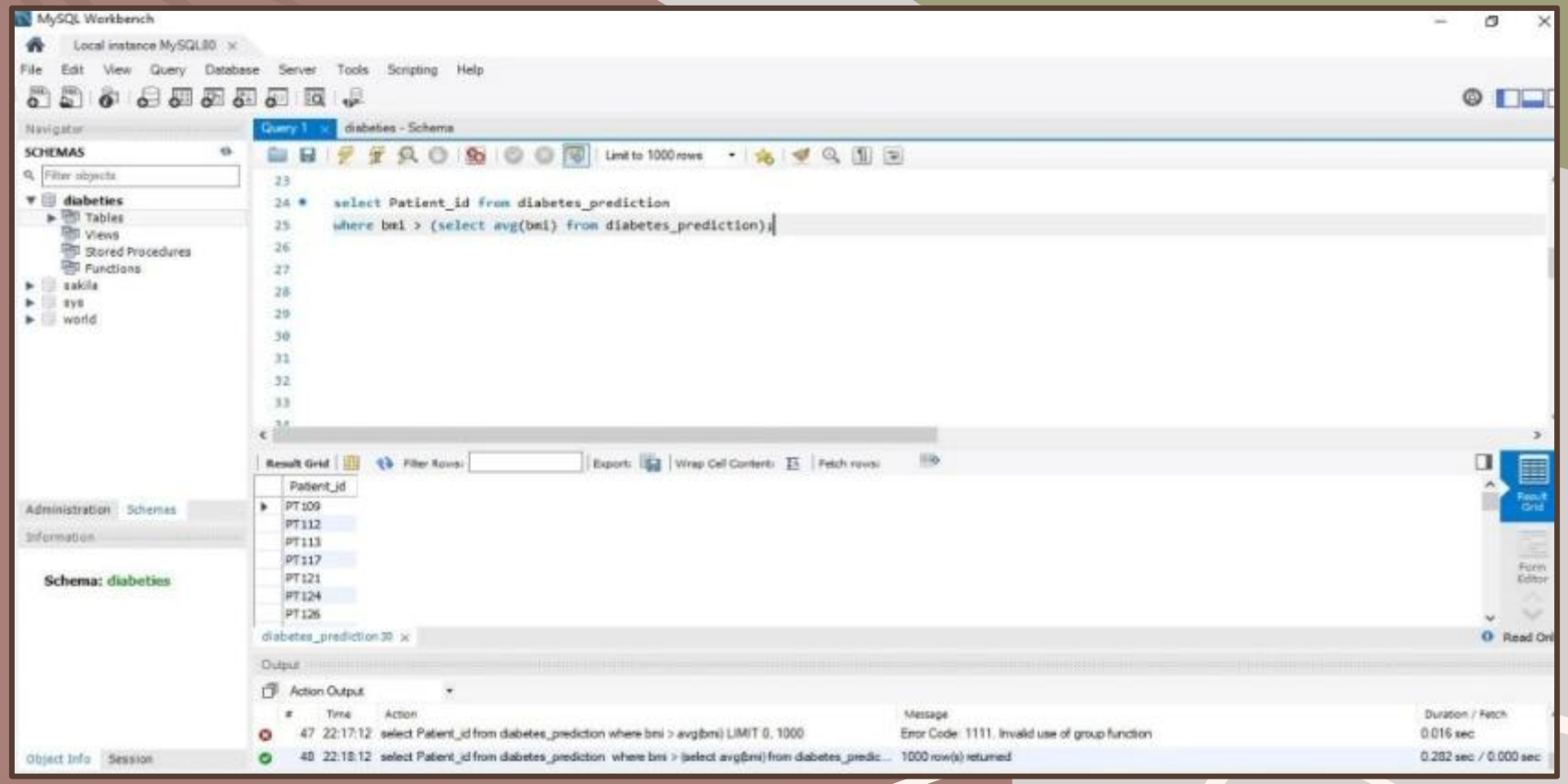
6. Determine the number of patients with heart disease.



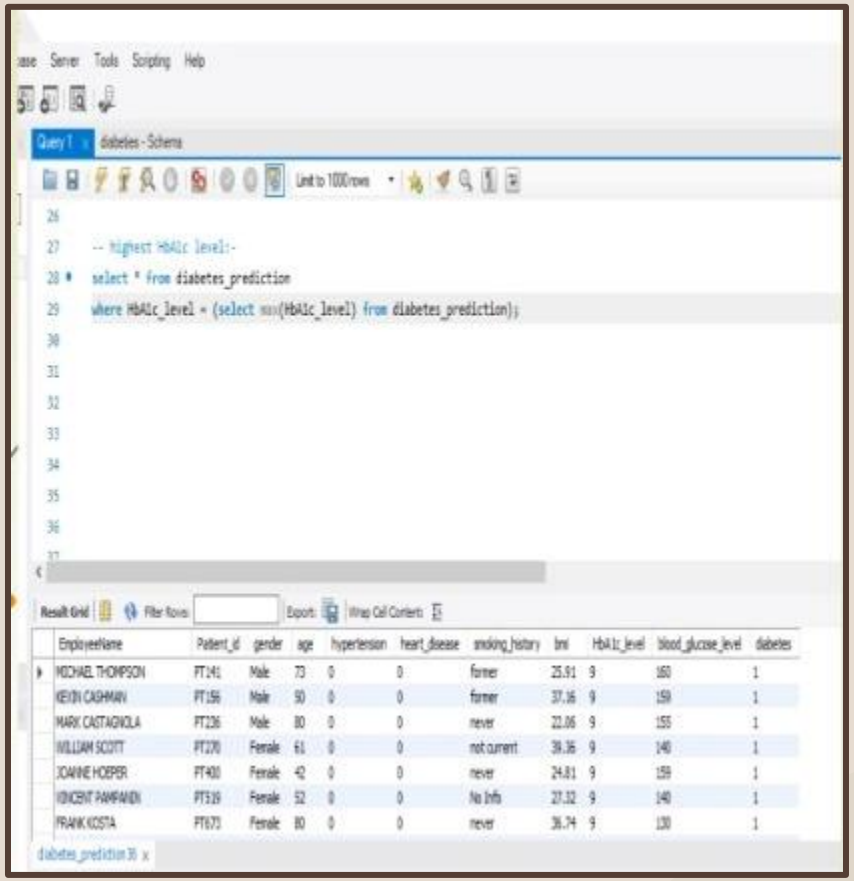
7. Group patients by smoking history and count how many smokers and nonsmokers there are.

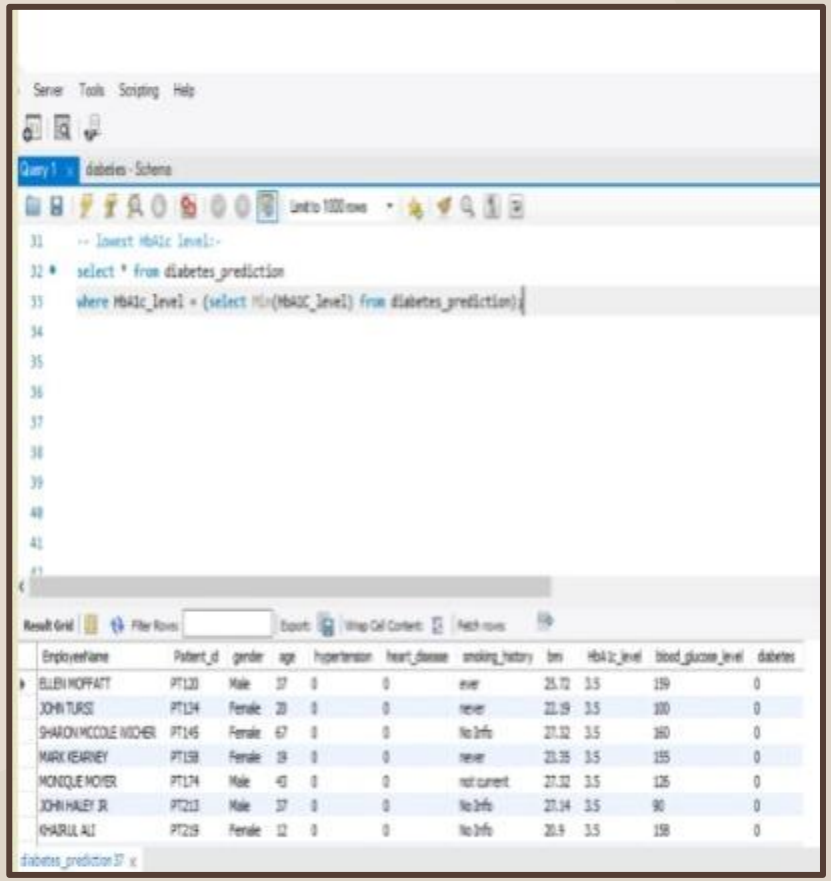


8. Retrieve the Patient_ids of patients who have a BMI greater than the average BMI.

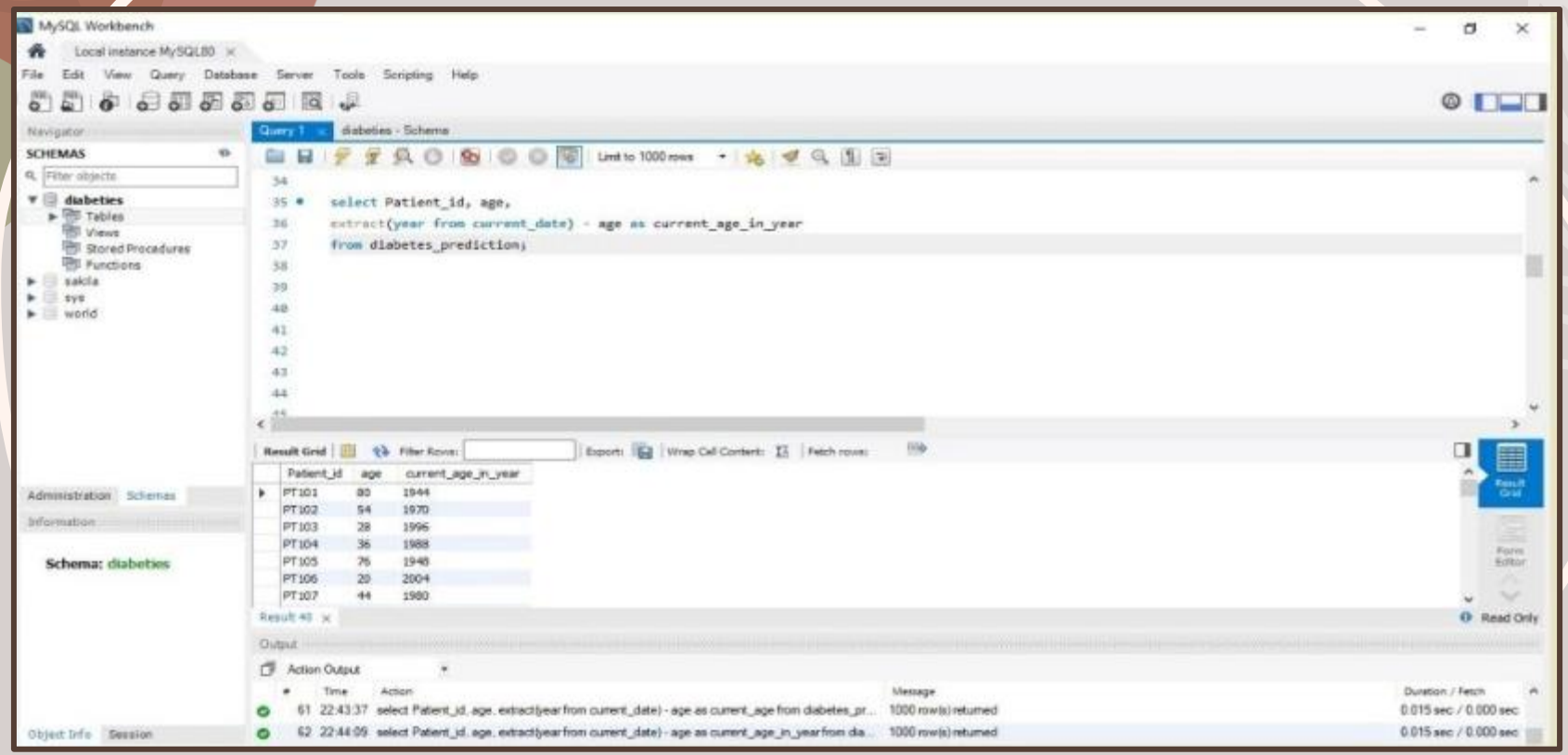


9. Find the patient with the highest HbA1c level and the patient with the lowest HbA1clevel.

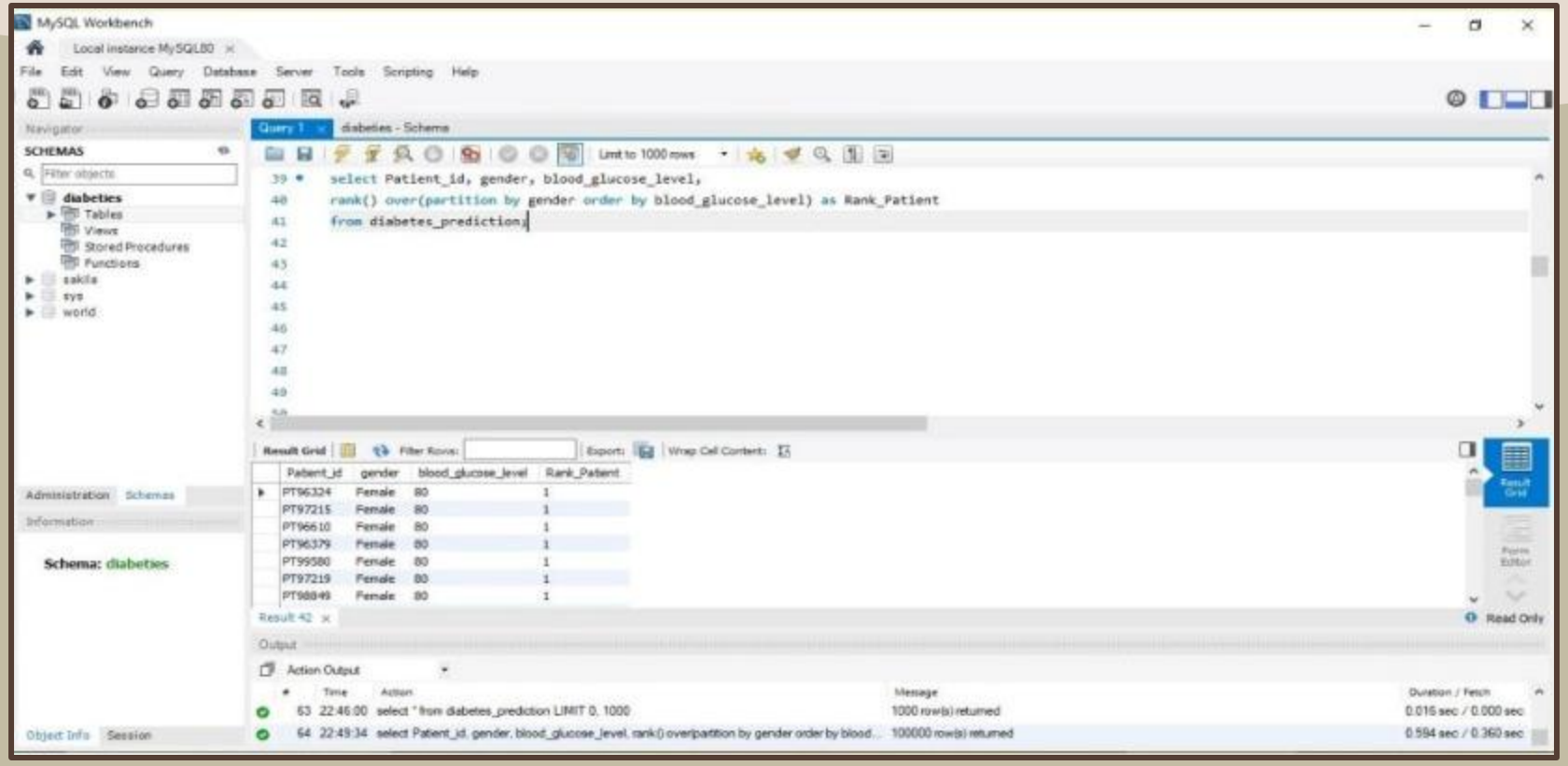




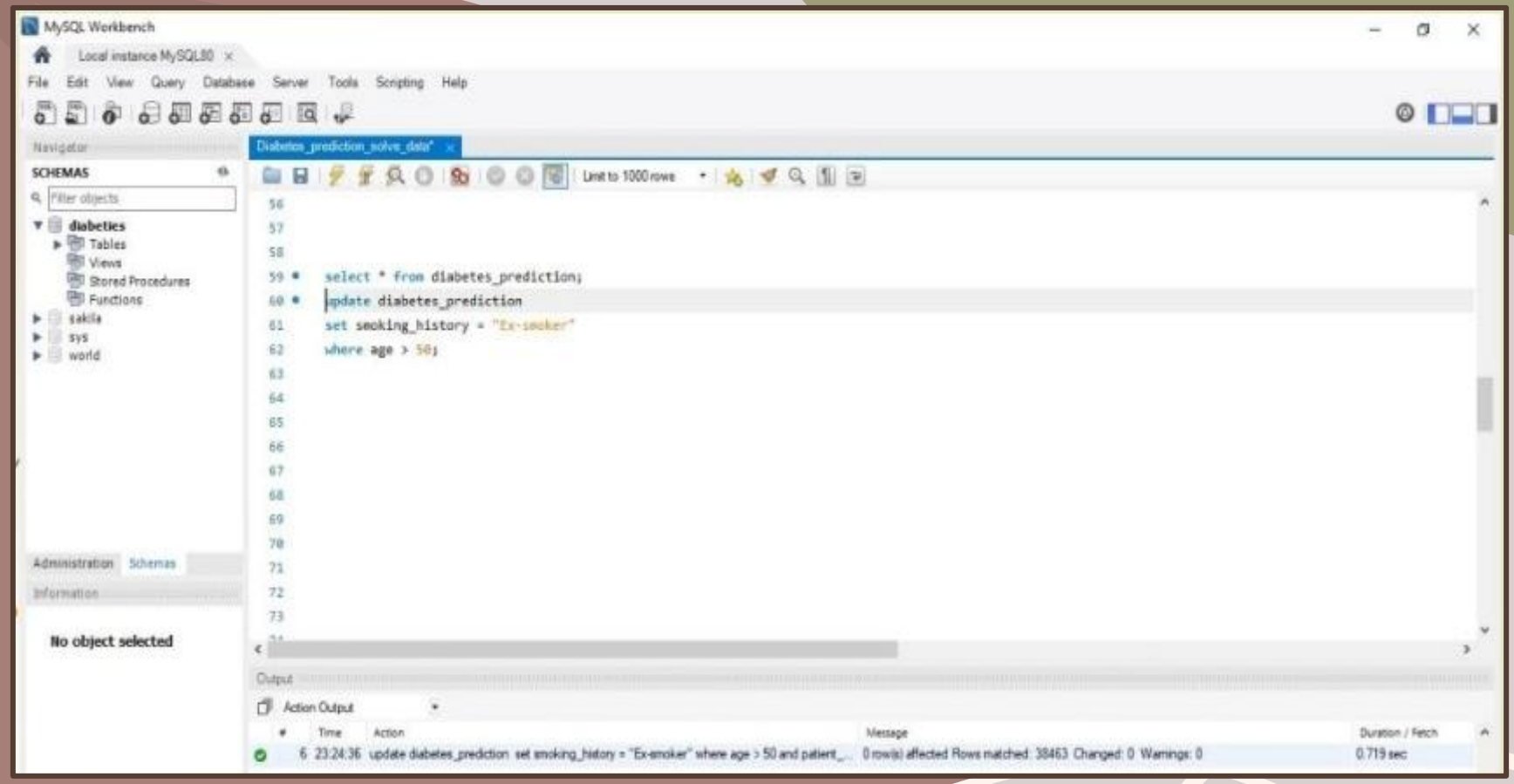
10. Calculate the age of patients in years (assuming the current date as of now).



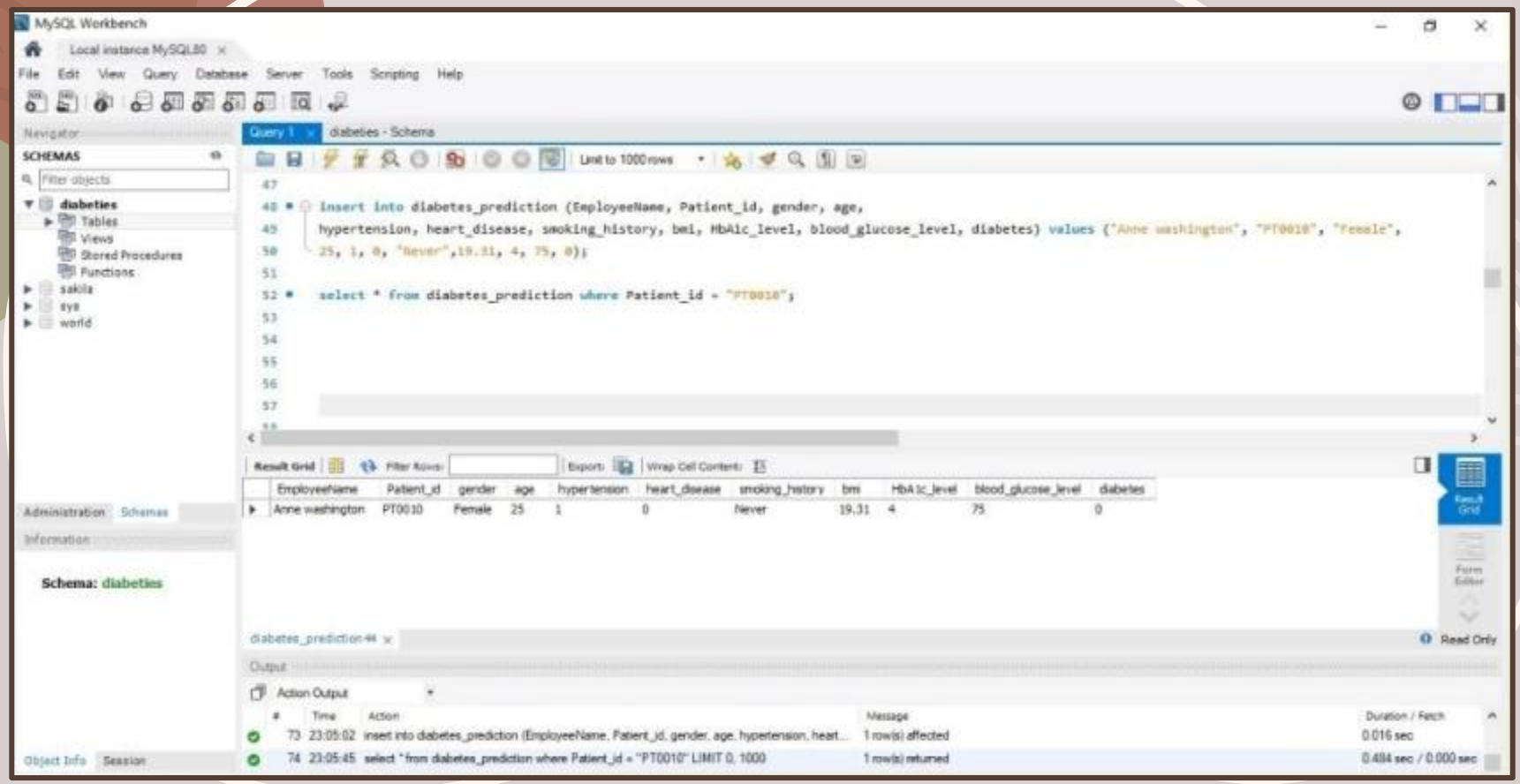
11. Rank patients by blood glucose level within each gender group.



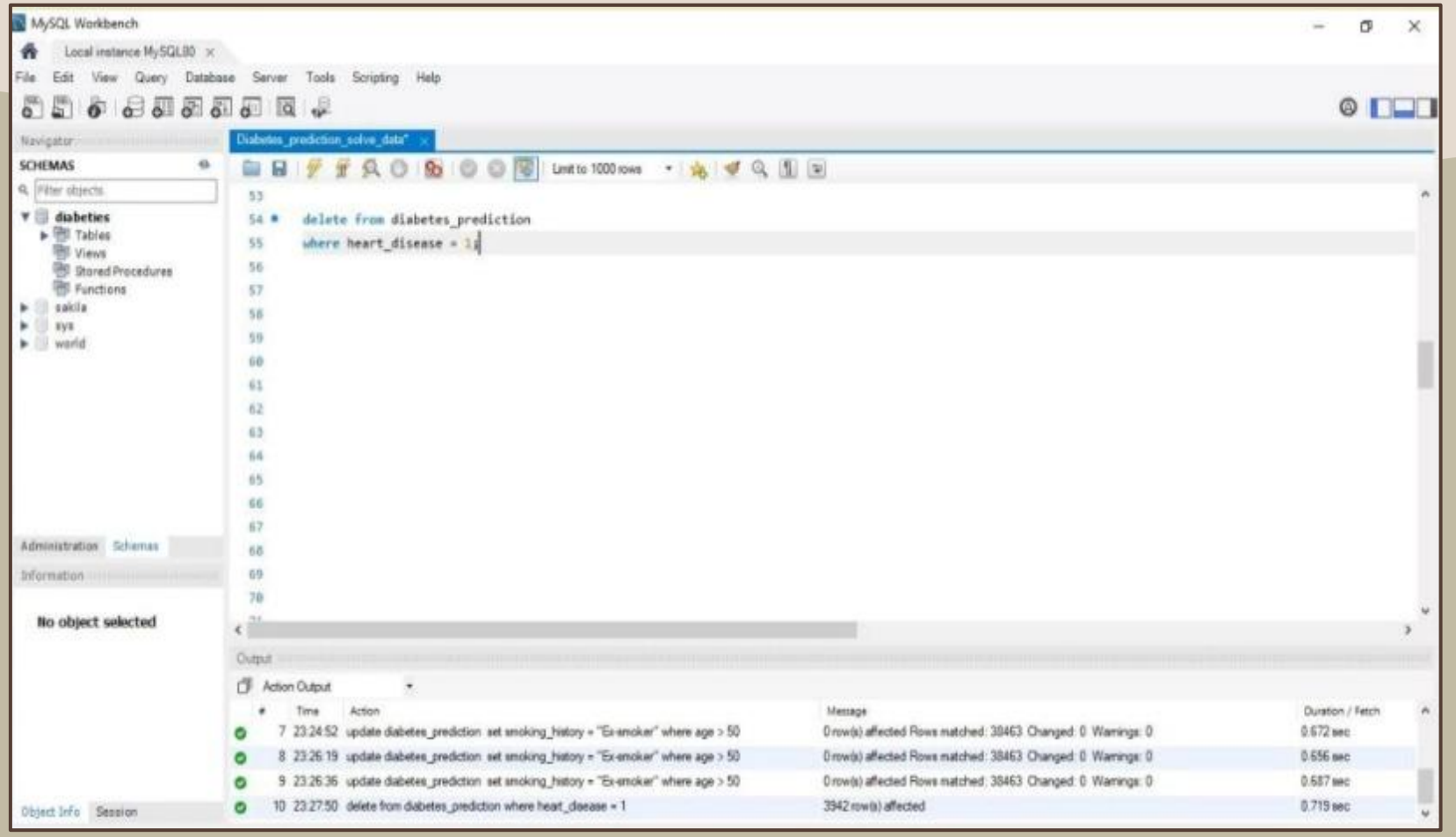
12. Update the smoking history of patients who are older than 50 to "Ex-smoker."



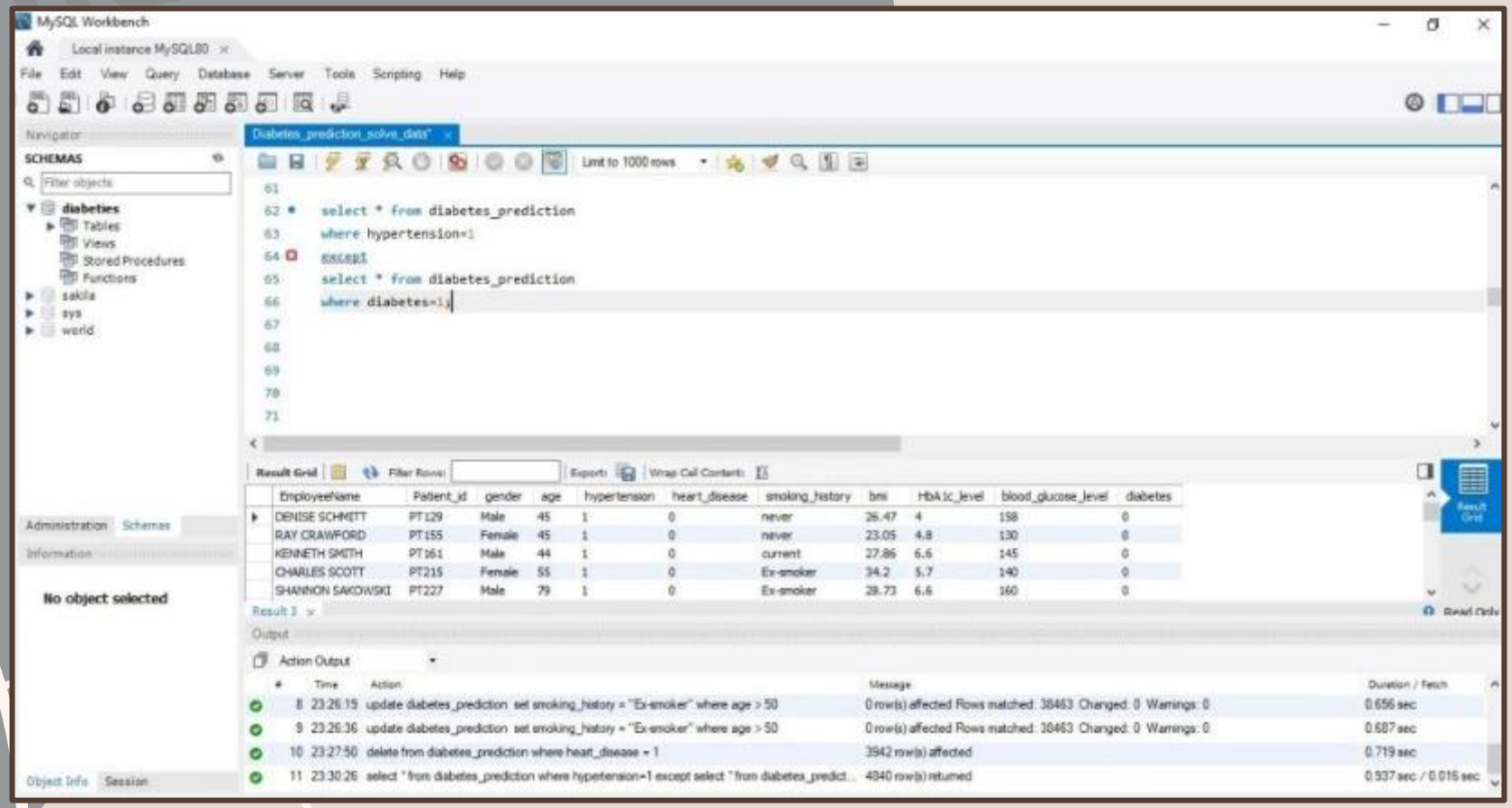
13. Insert a new patient into the database with sample data.



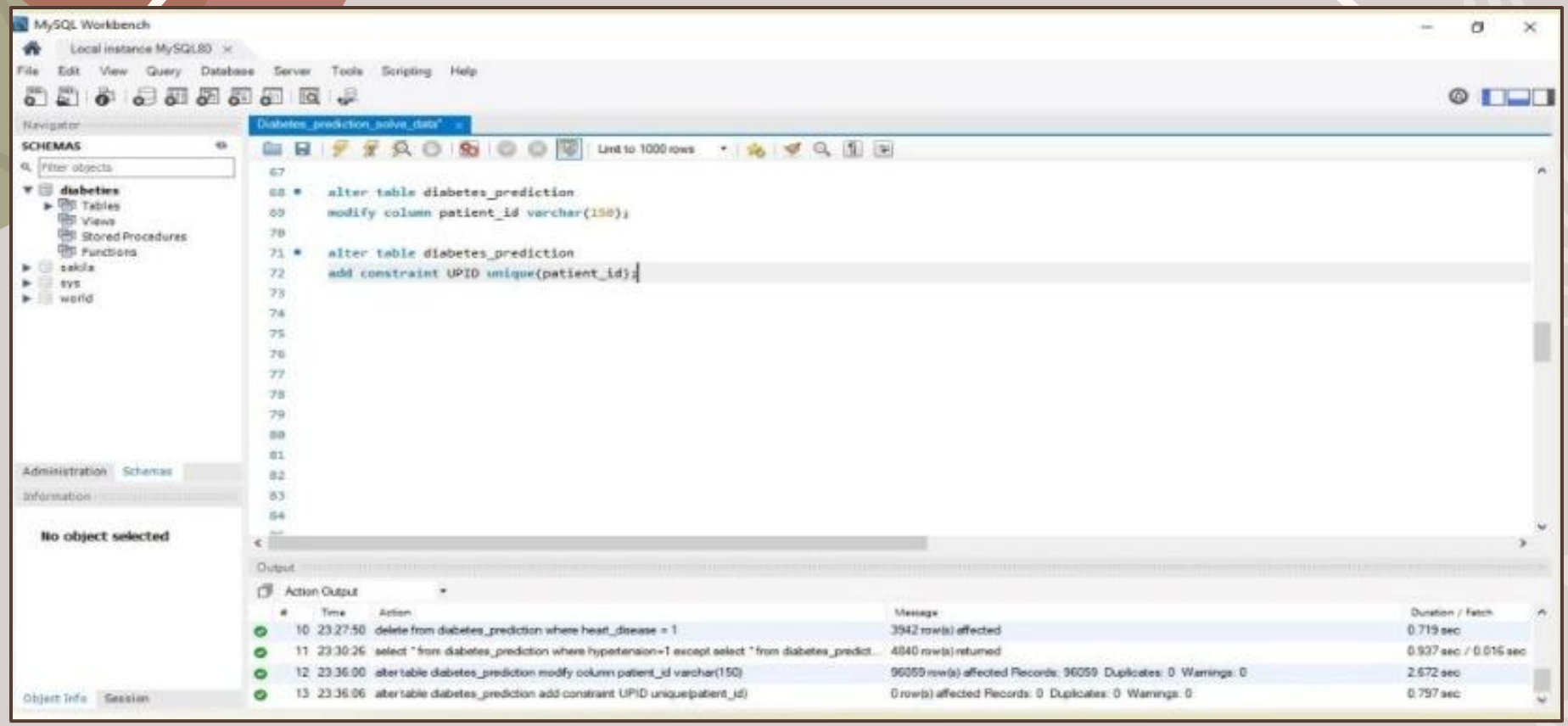
14. Delete all patients with heart disease from the database.



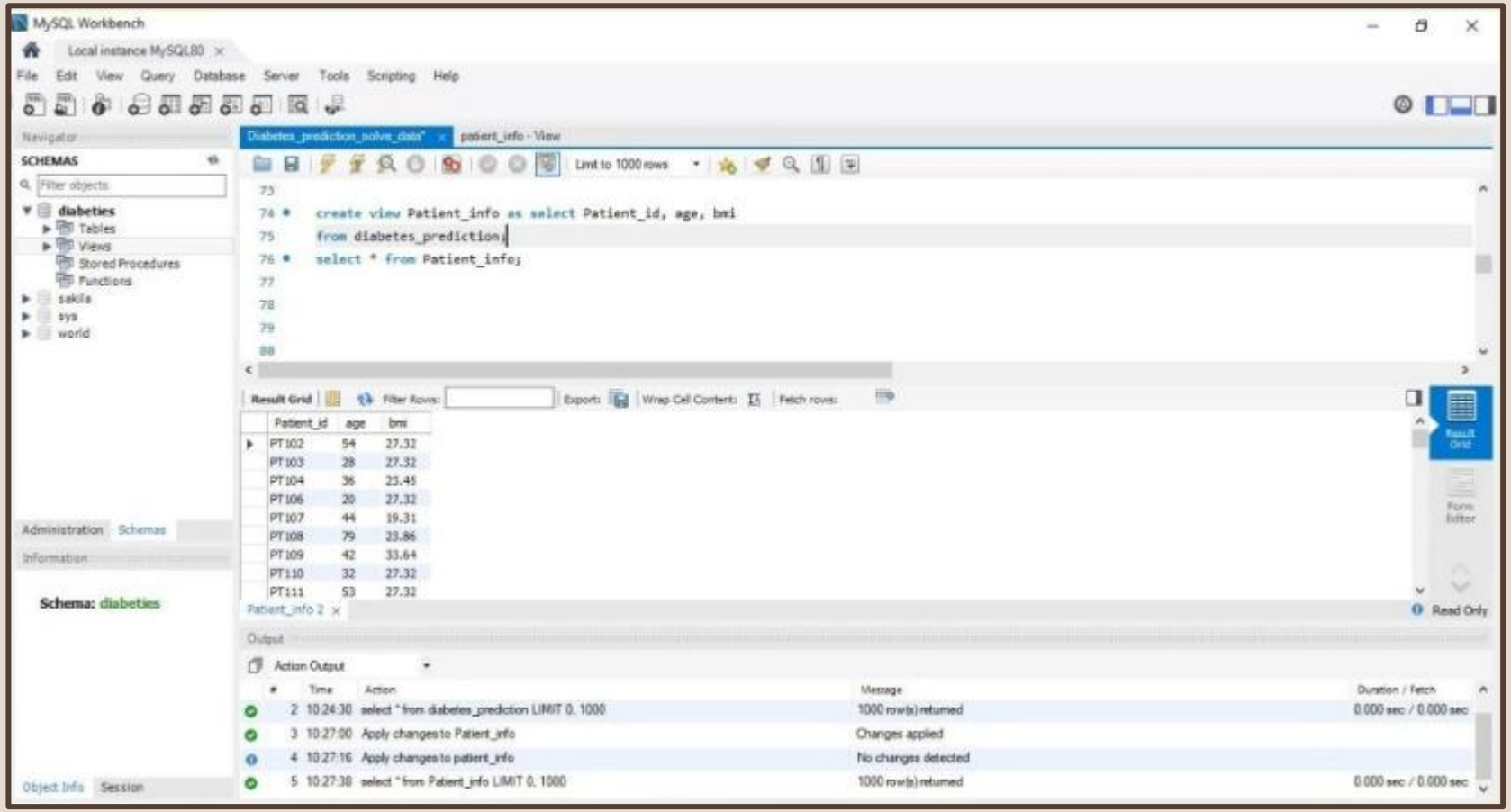
15. Find patients who have hypertension but not diabetes using the EXCEPT operator.



16. Define a unique constraint on the "patient_id" column to ensure its values are unique.



17. Create a view that displays the Patient_ids, ages, and BMI of patients.



18. Suggest improvements in the database schema to reduce data redundancy and improve data integrity.

Some of the improvements include,

- Moving employees name in another table & replacing them by Employees id.
- Adding a table for patients' name, age, gender and other information, hence removing them from data table.
- Using primary and foreign keys for patients, employees and general data.
- Using views for simplifying queries
- Setting up constraints such as unique, not null and check constraints.

19. Explain how you can optimize the performance of SQL queries on this dataset.

- Using Correct data types.
- We can use indexing, if there is frequently used columns.
- Writing simple queries & avoiding sub-queries.
- Avoiding [select*] query instead retrieving only necessary columns.

thank you

Vaishnavi C. Badgujar

badgujarvaishnavi7520@gmail.com