



DIABETES PREDICTION ANALYSIS

TASK - 2 (USING MICROSOFT EXCEL)

BY VISHAL KUMAR

Project Description:

DIABETES IS A CHRONIC DISEASE AFFECTING MILLIONS WORLDWIDE, LEADING TO DEVASTATING COMPLICATIONS LIKE CARDIOVASCULAR DISEASE, NEPHROPATHY, AND NEUROPATHY. EARLY DIAGNOSIS AND INTERVENTION ARE CRUCIAL IN MANAGING THE DISEASE AND PREVENTING THESE COMPLICATIONS. HOWEVER, TRADITIONAL METHODS RELY ON REACTIVE DIAGNOSES AFTER SYMPTOMS EMERGE.

PROJECT AIM: THIS PROJECT AIMS TO DEVELOP A DATA-DRIVEN APPROACH FOR PREDICTING DIABETES AND ITS COMPLICATIONS AT AN EARLY STAGE USING SQL.

OBJECTIVES:

- DEVELOP A PREDICTIVE MODEL: UTILIZE PATIENT DATA INCLUDING DEMOGRAPHICS, MEDICAL HISTORY, LIFESTYLE FACTORS, AND CLINICAL METRICS TO BUILD A MACHINE LEARNING MODEL FOR ACCURATELY PREDICTING THE ONSET OF DIABETES AND ITS POTENTIAL COMPLICATIONS.
- IDENTIFY KEY RISK FACTORS: ANALYZE THE MODEL TO UNDERSTAND THE SIGNIFICANT CONTRIBUTING FACTORS FOR DIABETES AND ITS COMPLICATIONS, FACILITATING TARGETED PREVENTION STRATEGIES.
- PERSONALIZE INTERVENTIONS: DESIGN A SYSTEM THAT DELIVERS PERSONALIZED RECOMMENDATIONS FOR LIFESTYLE MODIFICATIONS, MEDICAL MONITORING, AND EARLY TREATMENT BASED ON INDIVIDUAL RISK PROFILES.
- IMPROVE EARLY DETECTION: DEVELOP AN INTERACTIVE PLATFORM FOR HEALTHCARE PROFESSIONALS TO ASSESS INDIVIDUAL RISK BASED ON READILY AVAILABLE DATA, ENABLING EARLIER INTERVENTION AND BETTER DISEASE MANAGEMENT.

1. Retrieve the Patient_id and ages of all patients.

• • •

Q.1

```
select Patient_id, age  
from diabetes;
```



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

• • •

Q.2

```
select * from diabetes  
where gender = 'female' AND age > 40;
```



3. Calculate the average BMI of patients.

Q.3

```
select avg(bmi) as average_bmi  
from diabetes;
```



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



Q.4

```
select EmployeeName, Patient_id, age, blood_glucose_level from diabetes  
order by blood_glucose_level desc;
```



5. Find patients who have hypertension and diabetes.

• • •

Q.5

```
select EmployeeName, gender, hypertension, diabetes from diabetes  
where hypertension = 1 AND diabetes = 1;
```

1

6. Determine the number of patients with heart disease.

• • •

Q.6

```
select count(Patient_id) as heart_disease_patient from diabetes  
where heart_disease=1;
```



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

...

Q.7

```
select Count(*) as total_count, smoking_history from diabetes  
group by smoking_history;
```



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

• • •

Q.8

```
select Patient_id, bmi from diabetes  
where bmi > (select AVG(bmi) from diabetes);
```

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

...

Q.9

```
select patient_id, MAX(HbA1c_level) as highest_level, min(HbA1c_level) as lowest_level  
from diabetes;
```

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

... Q.10

```
SELECT
    employeeename,
    age,
    CURDATE() - INTERVAL age YEAR AS approximate_date_of_birth
FROM diabetes;
```



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

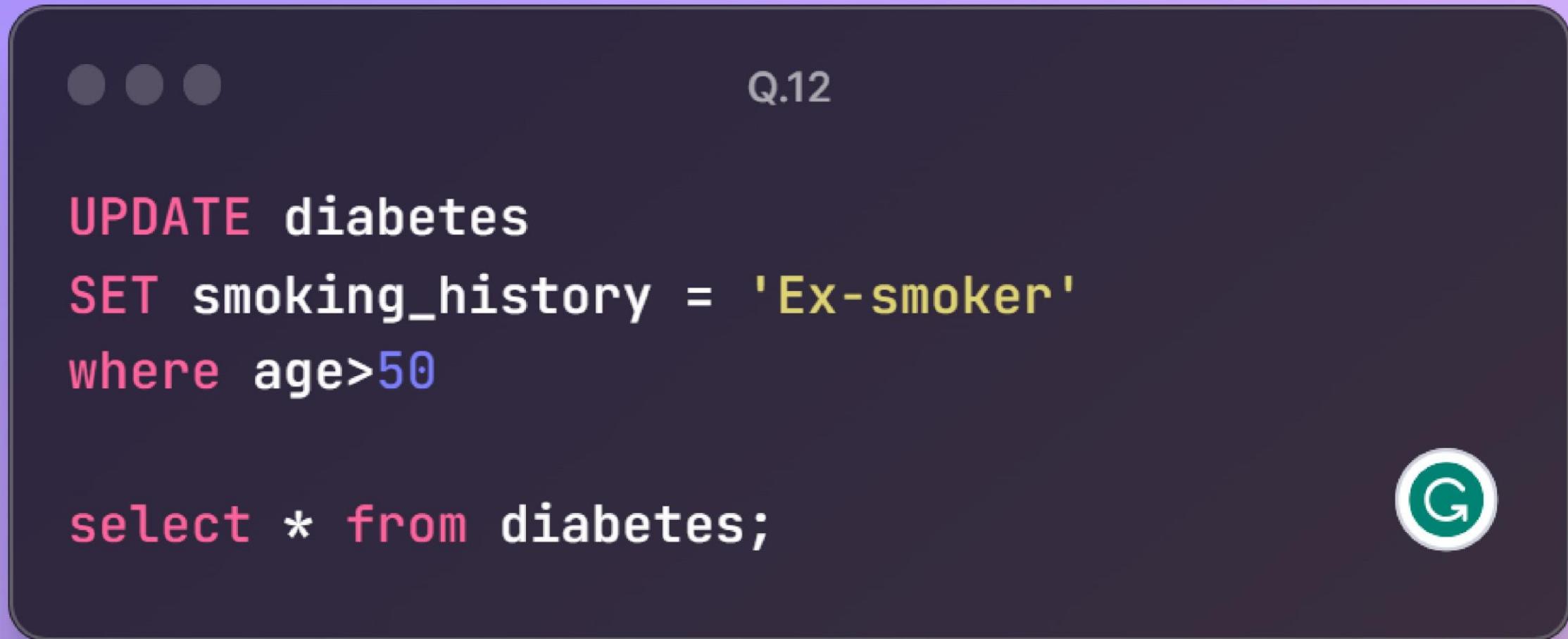


Q.11

```
select patient_id, blood_glucose_level, gender,  
rank() over(partition by gender order by blood_glucose_level) as glucose_level_rank  
from diabetes;
```



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



Q.12

```
UPDATE diabetes  
SET smoking_history = 'Ex-smoker'  
where age>50  
  
select * from diabetes;
```

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

• • •

Q.13

```
INSERT INTO diabetes  
VALUES('Neymar Junior', 'PT100', 'Male', 37, 0, 1, 'current', 23.46, 4.2, 152, 0);
```

○

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

• • •

Q.14

```
DELETE FROM diabetes  
where heart_disease=1;
```



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

• • •

Q.15

```
select * from diabetes  
where hypertension=1  
except  
select * from diabetes  
where diabetes=1;
```



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



Q.16

```
ALTER TABLE diabetes  
ADD UNIQUE(patient_id);
```



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

Q.17

```
CREATE VIEW Patient_info AS  
select patient_id, age, bmi from diabetes;  
  
select * from diabetes;
```

1

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

Some suggestions to improve database schema for reduced redundancy and better data integrity:

- 1. Normalization**
- 2. Use of primary and foreign keys:**
- 3. Avoiding null values**
- 4. Use of constraints**
- 5. Use of ER diagrams**
- 6. Regularly reviewing and updating the schema**
- 7. Use of version control**
- 8. Implementing referential integrity**
- 9. Ensuring security**

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

There are some ways to optimize the performance of SQL queries in the dataset:

- Minimize the use of wildcard characters**
- Use Index Correctly**
- Select only the columns you need:**
- Consider de-normalizing your database**
- Avoid loops and stick to set operations**
- Monitor query performance**
- Use a WITH statement to handle query filtering**
- Reduce field sizes to a minimum**
- Use uppercase for keywords**
- Run queries at off-peak hours**

THANK
YOU