DIABETES PREDICATION ASSESSMENT QUESTIONS

Q Retrieve the Patient_id and ages of all patients.

SELECT Patient_id,age from turing-clock-400911.project.Diabetes_prediction;

Q Select all female patients who are older than 40.

SELECT * from turing-clock-400911.project.Diabetes_prediction where gender = 'Female' AND age > 40;

Q Calculate the average BMI of patients.

SELECT avg(bmi) from turing-clock-400911.project.Diabetes_prediction;

QList patients in descending order of blood glucose levels.

SELECT EmployeeName from turing-clock-400911.project.Diabetes_prediction order by blood_glucose_level desc;

QFind patients who have hypertension and diabetes.

SELECT EmployeeName,Patient_id from turing-clock-400911.project.Diabetes_prediction where hypertension=1 and diabetes = 1;

QDetermine the number of patients with heart disease.

SELECT count(*) from turing-clock-400911.project.Diabetes_prediction where heart_disease = 1;

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

SELECT count(*),smoking_history from `turing-clock-400911.project.Diabetes_prediction` group by smoking_history;

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

SELECT patient_id from turing-clock-400911.project.Diabetes_prediction where bmi > (SELECT avg(bmi) from turing-clock-400911.project.Diabetes_prediction) ;

QFind the patient with the highest HbA1c level and the patient with the lowest HbA1clevel.

select patient_id,EmployeeName from `turing-clock-400911.project.Diabetes_prediction` where HbA1c_level =

(select max(HbA1c_level) from turing-clock-400911.project.Diabetes_prediction) OR HBA1c_level = (select min(HbA1c_level) from turing-clock-400911.project.Diabetes_prediction);

QCalculate the age of patients in years (assuming the current date as of now)

SELECT age from `turing-clock-400911.project.Diabetes_prediction`;

QRank patients by blood glucose level within each gender group.

SELECT *,rank() over(Partition by gender ORDER BY blood_glucose_level) glucose_level from `turing-clock-400911.project.Diabetes_prediction`;

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

update `turing-clock-400911.project.Diabetes_prediction` set smoking_history = 'Ex-smoker' where age > 50;

QInsert a new patient into the database with sample data.

INSERT INTO `turing-clock-400911.project.Diabetes_prediction`values('HARRY','PT100101','Male',66,0,1,'never',27.30,6.1,130,1);

QDelete all patients with heart disease from the database.

Delete from `turing-clock-400911.project.Diabetes_prediction` where heart_disease =1;

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

SELECT * from `turing-clock-400911.project.Diabetes_prediction`where hypertension = 1

except Distinct

SELECT * from `turing-clock-400911.project.Diabetes_prediction`where diabetes = 1;

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

Alter table turing-clock-400911.project.Diabetes_prediction ADD CONSTRAINT UNIQUE(Patient_id);

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

Create view turing-clock-400911.project.summary as select Patient_id,age,bmi from `turing-clock-400911.project.Diabetes_prediction`;

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

- Use constraint like primary key and unique key.
- Apply Conditional Formatting to highlight the wrong values.
- Use Data Validation steps.

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

We can use the result of above sql queries in analysing the health of the given patients and providing the correct medication also we can predict or monitor the health parameters which can lead to diabetes in the patients.

URL
https://console.cloud.google.com/bigquery?sq=720355025104:d84d5792eb034a1998ff5bf645ed28b9