ANALYSIS OF THE DIABETES DATASET USING POSTGRESQL

Count the number of diabetic and non-diabetic individuals in the dataset.

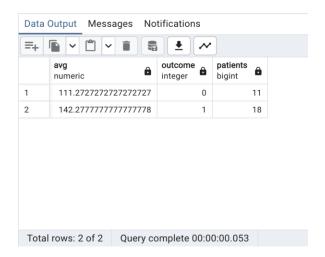
SELECT Outcome, COUNT(*) as PATIENTS FROM diabetes

GROUP BY Outcome;



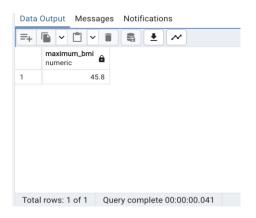
Calculate the average glucose level for diabetic and non-diabetic individuals.

SELECT avg(glucose), Outcome, COUNT(*) AS patients FROM diabetes GROUP BY Outcome;



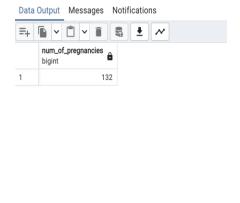
Find the maximum BMI (Body Mass Index) in the dataset.

SELECT max(bmi) AS Maximum_Bmi from diabetes;



Determine the number of pregnancies for individuals aged 30 or above.

SELECT SUM(pregnancies) AS Num_of_Pregnancies FROM diabetes WHERE age >=30;



Total rows: 1 of 1 Query complete 00:00:00.032

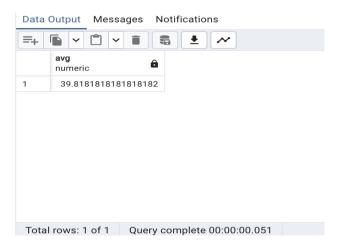
Identify individuals with a BMI greater than 30 and glucose level above 150.

SELECT * FROM diabetes WHERE bmi>=30 AND glucose >150;



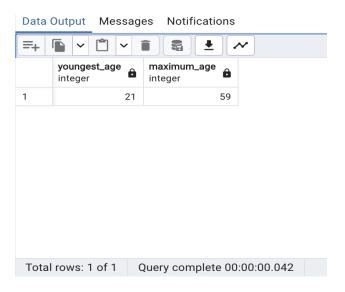
Calculate the average age of diabetic individuals with more than 5 pregnancies.

SELECT AVG(age) FROM diabetes WHERE Outcome= 1 AND pregnancies > 5;



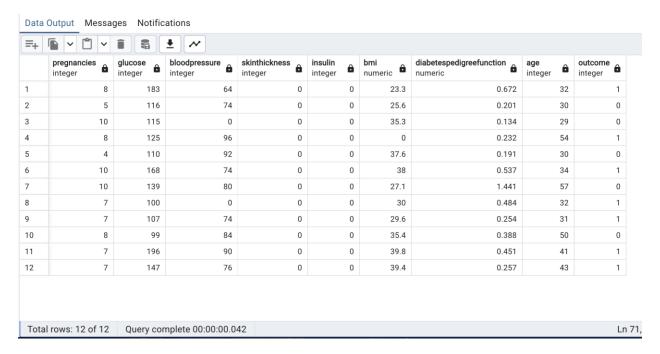
Find the youngest and oldest individual in the dataset.

SELECT MIN(age) as Youngest_Age ,MAX(age) as Maximum_Age FROM diabetes;

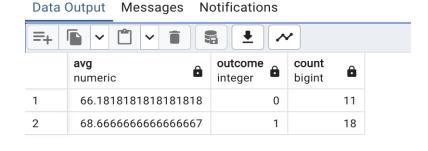


Identify individuals with missing values for SkinThickness or Insulin.

SELECT * FROM diabetes WHERE skinthickness = 0 AND insulin=0;



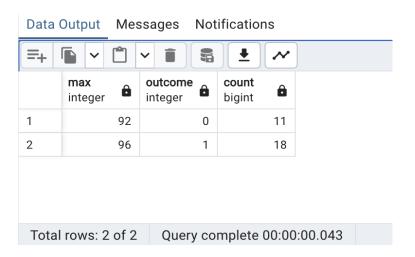
What is the distribution of blood pressure among diabetic and non-diabetic individuals? SELECT avg(bloodpressure), outcome, COUNT(*) FROM diabetes GROUP BY outcome;



Total rows: 2 of 2 Query complete 00:00:00.039

From the above results it is evident that the average blood pressure for non-diabetic individuals is 66.18 and the average blood pressure for diabetic individuals is 68.66.

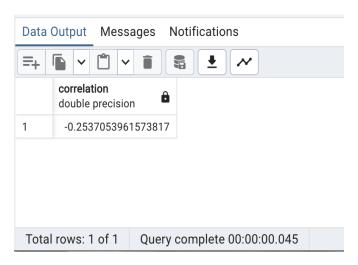
SELECT max(bloodpressure), outcome, COUNT(*) FROM diabetes GROUP BY outcome;



The maximum value of blood pressure for non-diabetic and diabetic individuals are 92 and 96 respectively.

Is there a correlation between the number of pregnancies and BMI?

SELECT CORR(pregnancies, BMI) AS correlation from diabetes;



From the above, it is evident that there is a negative correlation between the number of pregnancies and BMI.

How does the distribution of age differ between diabetic and non-diabetic individuals?

SELECT outcome as Patients, AVG(age) as average_age, MAX(age) as maximum_age, MIN(age) as minimum_age, COUNT(outcome) AS total_count FROM diabetes GROUP BY outcome;

Data	Output Mes	ssages Notifications			
	patients integer	average_age numeric	maximum_age integer	minimum_age integer	total_count bigint
1	0	35.1818181818181818	57	21	11
2	1	40.166666666666667	59	26	18

Total rows: 2 of 2 Query complete 00:00:00.050

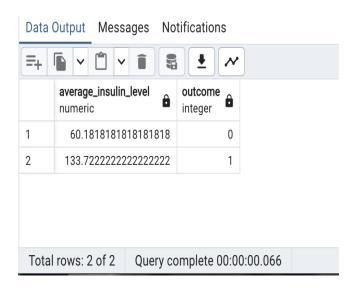
What is the average Diabetes Pedigree Function value for individuals with and without diabetes?

SELECT AVG(diabetespedigreefunction), Outcome as patients, Count(Outcome) AS number_of_patients FROM diabetes GROUP BY Outcome;



Total rows: 2 of 2 Query complete 00:00:00.055

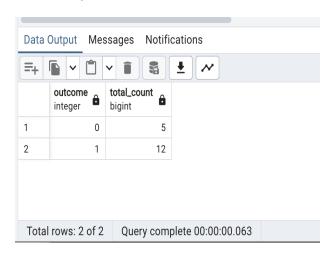
Are there any significant differences in insulin levels between individuals with and without diabetes?



The average insulin level is 60.18 for individuals who are non-diabetic and the average insulin level is 133.72 for diabetic individuals.

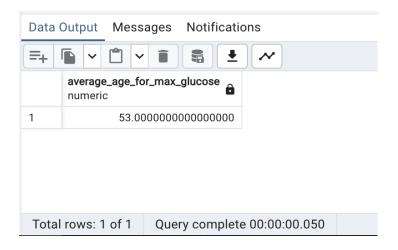
How many individuals have a BMI classified as "obese" (BMI > 30), and what is the distribution of diabetic and non-diabetic individuals within this group?

SELECT Outcome, Count(*) AS total_count from diabetes WHERE BMI > 30 GROUP BY Outcome;



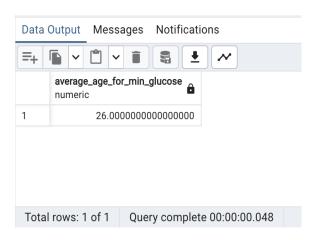
What is the average age of individuals with the highest and lowest glucose levels?

SELECT AVG(Age) FROM diabetes WHERE Glucose = (SELECT MAX(Glucose) FROM



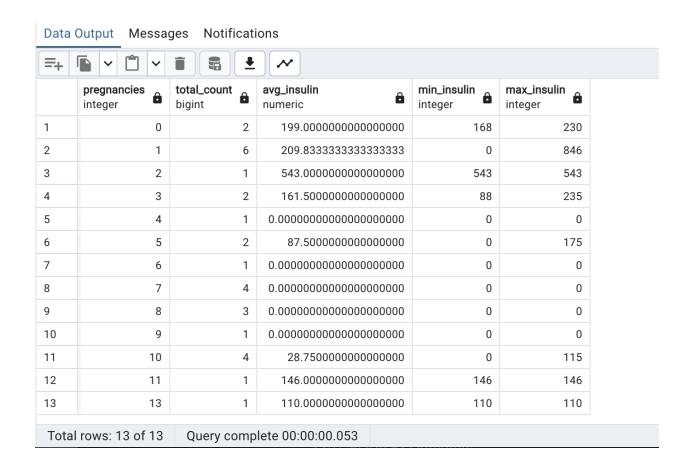
diabetes);

SELECT AVG(Age) FROM diabetes WHERE Glucose = (SELECT MIN(Glucose) FROM diabetes);



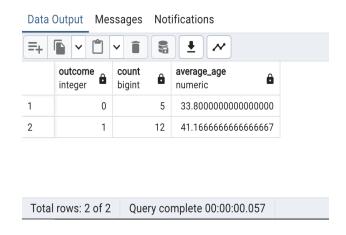
Are there any trends in insulin levels among individuals with different numbers of pregnancies?

SELECT Pregnancies, COUNT(*) AS total_count, AVG(Insulin) AS avg_insulin, MIN(Insulin) AS min_insulin, MAX(Insulin) AS max_insulin FROM diabetes GROUP BY Pregnancies ORDER BY Pregnancies;



What is the average age of individuals with a BMI over 30, grouped by diabetic status?

SELECT Outcome, Count(Outcome), AVG(age) AS average_age FROM diabetes WHERE BMI > 30 GROUP BY Outcome;



How does the distribution of SkinThickness vary between diabetic and non-diabetic individuals?

SELECT AVG(SkinThickness), MIN(SkinThickness), MAX(SkinThickness), Outcome, Count(*) AS total_count FROM diabetes GROUP BY Outcome;

