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;

Data Output   Messages   Notifications

	outcome integer	patients bigint
1	0	11
2	1	18

Total rows: 2 of 2   Query complete 00:00:00.077

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

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

Data Output   Messages   Notifications

	avg numeric	outcome integer	patients bigint
1	111.2727272727272727	0	11
2	142.2777777777777778	1	18

Total rows: 2 of 2   Query complete 00:00:00.053

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

SELECT max(bmi) AS Maximum_Bmi from diabetes;

	maximum_bmi 🔒 numeric
1	45.8

Total rows: 1 of 1   Query complete 00:00:00.041

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

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

	num_of_pregnancies 🔒 bigint
1	132

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 ;

	pregnancies integer	glucose integer	bloodpressure integer	skinthickness integer	insulin integer	bmi numeric	diabetespedigreefunction numeric	age integer	outcome integer
1	2	197	70	45	543	30.5	0.158	53	1
2	10	168	74	0	0	38	0.537	34	1
3	1	189	60	23	846	30.1	0.398	59	1
4	7	196	90	0	0	39.8	0.451	41	1

Total rows: 4 of 4   Query complete 00:00:00.035   Ln 66, Col 1

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

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

	avg numeric
1	39.8181818181818182

Total rows: 1 of 1   Query complete 00:00:00.051

Find the youngest and oldest individual in the dataset.

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

	youngest_age integer	maximum_age integer
1	21	59

Total rows: 1 of 1   Query complete 00:00:00.042

Identify individuals with missing values for SkinThickness or Insulin.

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

Data Output   Messages   Notifications

	pregnancies integer	glucose integer	bloodpressure integer	skinthickness integer	insulin integer	bmi numeric	diabetespedigreefunction numeric	age integer	outcome integer
1	8	183	64	0	0	23.3	0.672	32	1
2	5	116	74	0	0	25.6	0.201	30	0
3	10	115	0	0	0	35.3	0.134	29	0
4	8	125	96	0	0	0	0.232	54	1
5	4	110	92	0	0	37.6	0.191	30	0
6	10	168	74	0	0	38	0.537	34	1
7	10	139	80	0	0	27.1	1.441	57	0
8	7	100	0	0	0	30	0.484	32	1
9	7	107	74	0	0	29.6	0.254	31	1
10	8	99	84	0	0	35.4	0.388	50	0
11	7	196	90	0	0	39.8	0.451	41	1
12	7	147	76	0	0	39.4	0.257	43	1

Total rows: 12 of 12   Query complete 00:00:00.042   Ln 71,

What is the distribution of blood pressure among diabetic and non-diabetic individuals?

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

Data Output   Messages   Notifications

	avg numeric	outcome integer	count bigint
1	66.1818181818181818	0	11
2	68.6666666666666667	1	18

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;

Data Output   Messages   Notifications

	max integer	outcome integer	count bigint
1	92	0	11
2	96	1	18

Total rows: 2 of 2   Query complete 00:00:00.043

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;

Data Output   Messages   Notifications

	correlation double precision
1	-0.2537053961573817

Total rows: 1 of 1   Query complete 00:00:00.045

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   Messages   Notifications

	patients integer	average_age numeric	maximum_age integer	minimum_age integer	total_count bigint
1	0	35.1818181818181818	57	21	11
2	1	40.1666666666666667	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;

avg numeric	patients integer	number_of_patients bigint
0.40836363636363636364	0	11
0.49972222222222222222	1	18

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?

average_insulin_level numeric	outcome integer
60.1818181818181818	0
133.7222222222222222	1

Total rows: 2 of 2   Query complete 00:00:00.066

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;

Data Output   Messages   Notifications

	outcome integer	total_count bigint
1	0	5
2	1	12

Total rows: 2 of 2   Query complete 00:00:00.063

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);

Data Output   Messages   Notifications

	average_age_for_max_glucose numeric
1	53.0000000000000000

Total rows: 1 of 1   Query complete 00:00:00.050

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

	average_age_for_min_glucose 🔒 numeric
1	26.0000000000000000

Data Output   Messages   Notifications

Total rows: 1 of 1   Query complete 00:00:00.048

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;

	pregnancies integer	total_count bigint	avg_insulin numeric	min_insulin integer	max_insulin integer
1	0	2	199.0000000000000000	168	230
2	1	6	209.8333333333333333	0	846
3	2	1	543.0000000000000000	543	543
4	3	2	161.5000000000000000	88	235
5	4	1	0.00000000000000000000	0	0
6	5	2	87.5000000000000000	0	175
7	6	1	0.00000000000000000000	0	0
8	7	4	0.00000000000000000000	0	0
9	8	3	0.00000000000000000000	0	0
10	9	1	0.00000000000000000000	0	0
11	10	4	28.7500000000000000	0	115
12	11	1	146.0000000000000000	146	146
13	13	1	110.0000000000000000	110	110

Total rows: 13 of 13   Query complete 00:00:00.053

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;

	outcome integer	count bigint	average_age numeric
1	0	5	33.8000000000000000
2	1	12	41.1666666666666667

Total rows: 2 of 2   Query complete 00:00:00.057

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;

Data Output   Messages   Notifications

	outcome integer	total_count bigint	avg_skin_thickness numeric	min_skin_thickness integer	max_skin_thickness integer
1	0	11	15.0000000000000000	0	41
2	1	18	20.0000000000000000	0	47

Total rows: 2 of 2   Query complete 00:00:00.057