

Title of the project 2: “Diabetes Patients”

ABOUT DATA SET:

This dataset is originally from **The National Institute of Diabetes and Digestive and Kidney Diseases**.

OBJECTIVE: The objective of the dataset is to diagnostically predict whether a patient has diabetes based on certain diagnostic measurements included in the dataset.

STEP 1-IMPORT EXCEL CSV FILE TO POSTGRESQL

CREATE TABLE DiabetesData (

Pregnancies **INT**,

Glucose **INT**,

BloodPressure **INT**,

SkinThickness **INT**,

Insulin **INT**,

BMI **FLOAT**,

DiabetesPedigree **FLOAT**,

Age **INT**,

Outcome **INT**

);

STEP-2

COPY DiabetesData(

Pregnancies,Glucose,BloodPressure,SkinThickness,Insulin,BMI,DiabetesPedigree,Age,Outcome)

FROM 'D:\My Projects\diabetes.csv'

DELIMITER ','

CSV HEADER;

STEP-3 : QUERIES

QUERIES TO FIND THE PATIENT HAS DIABETES ON CERTAIN MEASUREMENTS

1.Count of Diabetic and Non-Diabetic Patients:

SELECT Outcome, **COUNT(*) AS** Count

FROM DiabetesData

GROUP BY Outcome;

	outcome integer	count bigint
1	0	500
2	1	268

2. Average Glucose Level for Diabetic Patients:

SELECT ROUND(AVG(Glucose),2) AS Average_Glucose_Diabetic

FROM DiabetesData

WHERE Outcome = 1;

	average_glucose_diabetic numeric
1	141.26

3.Maximum and Minimum BMI among Non-Diabetic Patients and diabetic patients

SELECT

MAX(CASE WHEN Outcome = 0 **AND** BMI > 0 **THEN** BMI **ELSE** NULL **END) AS**
Max_BMI_Non_Diabetic,

MIN(CASE WHEN Outcome = 0 **AND** BMI > 0 **THEN** BMI **ELSE** NULL **END) AS**
Min_BMI_Non_Diabetic,

MAX(CASE WHEN Outcome = 1 **AND** BMI > 0 **THEN** BMI **ELSE** NULL **END) AS**
Max_BMI_Diabetic,

MIN(CASE WHEN Outcome = 1 **AND** BMI > 0 **THEN** BMI **ELSE** NULL **END) AS**
Min_BMI_Diabetic **FROM** DiabetesData;

	max_bmi_non_diabetic double precision	min_bmi_non_diabetic double precision	max_bmi_diabetic double precision	min_bmi_diabetic double precision
1	57.3	18.2	67.1	22.9

4. Average Age for Patients with More Than 4 Pregnancies:

```
SELECT ROUND(AVG(age),2) AS Average_Age_More_Pregnancies FROM DiabetesData
WHERE pregnancies > 4;
```

	average_age_more_pregnancies
	numeric
1	41.89

5. Determine the distribution of diabetic patients in various BMI categories (underweight, healthy, overweight, and obesity) to assess the correlation between BMI and diabetes in the dataset

```
SELECT
BMI_Status,
COUNT(*) AS Count
FROM (
SELECT *,
CASE
WHEN BMI < 18.5 THEN 'underweight'
WHEN BMI >= 18.5 AND BMI < 25 THEN 'healthy'
WHEN BMI >= 25 AND BMI < 30 THEN 'overweight'
WHEN BMI >= 30 THEN 'obesity'
ELSE 'unknown'
END AS BMI_Status
FROM DiabetesData
WHERE Outcome = 1
) AS DiabeticPatients
GROUP BY BMI_Status;
```

	bmi_status	count
	text	bigint
1	obesity	219
2	overweight	40
3	healthy	7
4	underweight	2

6. Diabetic Patients with a Diabetes Pedigree Score Above 0.5:

```
SELECT COUNT(diabetespedigree) AS diabetespedigreecount FROM
DiabetesData
```

```
WHERE Outcome = 1 AND DiabetesPedigree > 0.5;
```

	diabetespedigreecount
	bigint
1	126

7. Average SkinThickness for Patients with Low Glucose (under 90 mg/dL):

```
SELECT ROUND(AVG(SkinThickness),2) AS  
Average_SkinThickness_Low_Glucose  
  
FROM DiabetesData  
  
WHERE Glucose < 90;
```

	average_skinthickness_low_glucose numeric	
1	21.01	

8. Oldest Patient with Diabetes

```
SELECT MAX(Age) AS Oldest_Age_Diabetic  
  
FROM DiabetesData  
  
WHERE Outcome = 1;
```

	oldest_age_diabetic integer	
1	70	

9. Age range Distribution of Patients with Diabetes:

```
SELECT  
  '20-30' AS Age_Range,  
  COUNT(*) AS Count  
FROM DiabetesData  
WHERE Outcome = 1  
  AND Age >= 20  
  AND Age <= 30  
UNION ALL  
SELECT  
  '30-60' AS Age_Range,  
  COUNT(*) AS Count  
FROM DiabetesData  
WHERE Outcome = 1  
  AND Age > 30  
  AND Age <= 60;
```

	age_range text	count bigint
1	20-30	90
2	30-60	171

10. Patients with the Highest Glucose Level by Age Group:

```
SELECT Age_Group, MAX(Glucose) AS Max_Glucose
FROM (
  SELECT
    CASE
      WHEN Age < 30 THEN 'Under 30'
      WHEN Age >= 30 AND Age < 50 THEN '30-49'
      ELSE '50 and Over'
    END AS Age_Group, Glucose
  FROM DiabetesData
) AS DiabetesData
GROUP BY Age_Group;
```

	age_group text	max_glucose integer
1	50 and Over	197
2	30-49	197
3	Under 30	199

11. Average BMI for Patients with Diabetes by Number of Pregnancies:

```
SELECT Pregnancies, AVG(BMI) AS Avg_BMI_Diabetic
FROM DiabetesData
WHERE Outcome = 1
GROUP BY Pregnancies
ORDER BY Pregnancies DESC
LIMIT 5;
```

	pregnancies integer	avg_bmi_diabetic double precision
1	17	40.9
2	15	37.1
3	14	35.1
4	13	36.71999999999999
5	12	34.575

12. Patients with the Highest Glucose Levels and Blood Pressure above the Average:

```
SELECT *  
FROM DiabetesData  
WHERE Glucose = (SELECT MAX(Glucose) FROM DiabetesData) AND  
      BloodPressure > (SELECT AVG(BloodPressure) FROM DiabetesData);
```

pregnancies integer	glucose integer	bloodpressure integer	skinthickness integer	insulin integer	bmi double precision	diabetespedigree double precision	age integer	outcome integer
1	199	76	43	0	42.9	1.394	22	1

13. Diabetic and Non Diabetic patients with average and maximum insulin level

```
SELECT Outcome, ROUND(AVG(Insulin),2) AS Avg_Insulin_Level,  
MAX(Insulin) AS Max_Insulin_Level  
FROM DiabetesData  
GROUP BY Outcome;
```

	outcome integer	avg_insulin_level numeric	max_insulin_level integer
1	0	68.79	744
2	1	100.34	846

KEY INSIGHTS

1. **Diabetes Prevalence:** The dataset consists of both nondiabetic and diabetic patients, which is **500** and **268** respectively.
 2. **Glucose Levels:** Diabetic patients have an average glucose level of **141.26**, which is a key diagnostic marker in diabetes management.
 3. **BMI Variation:** By observing the dataset, it becomes evident that diabetic patients exhibit a **maximum BMI of 67.1** and a **minimum BMI of 22.9**, while **non-diabetic patients range from 18.2 to 57.3 in BMI**. These variations signal the presence of both **obesity and underweight** cases among patients.
 4. **Age and Pregnancy:** Patients with more than **four pregnancies** may be at risk for gestational diabetes. Their **average age is 41.89** provides insights into the age group at higher risk.
 5. **BMI Categories:** Categorizing diabetic patients into BMI categories helps uncover potential relationships between BMI and diabetes. Where patients with diabetes has **obesity , Overweight, healthy, and underweight** are **219,40,7,2** respectively.
 6. **Genetic Predisposition:** Diabetic patients with a **Diabetes Pedigree Score above 0.5, which is 126** may indicate a genetic influence on diabetes.
 7. **Skin Thickness and Glucose:** **Average skin thickness** for patients with low glucose levels (under 90 mg/dL) is **21.01** ,could reveal the impact of glucose levels on skin health.
 8. **Glucose and Blood Pressure:** Patients with the **highest glucose levels** is **199** and average blood pressure is **76** might require special attention for managing multiple risk factors.
 9. **Insulin level:** Max and average insulin of **non diabetic** patient has **68.79** and **744** and **100.34** and **846** for diabetic patient.
-