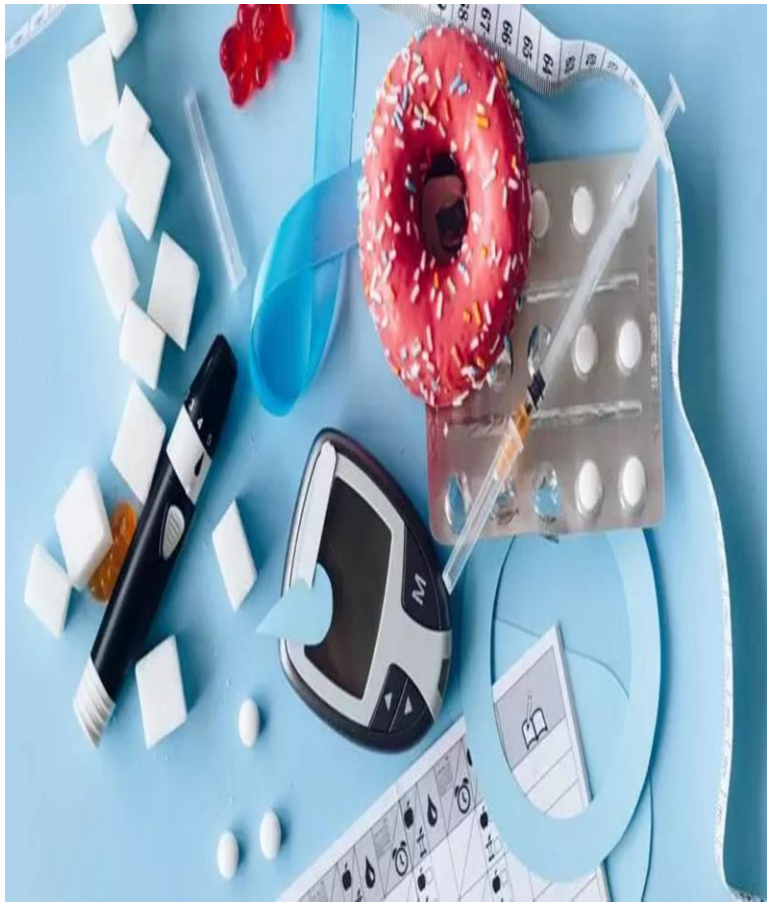
Diabetes Data Analysis

Project Overview :

- Diabetes is a chronic (long-lasting) health condition that affects how your body turns food into energy. Your body breaks down most of the food you eat into sugar (glucose) and releases it into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin. Diabetes is rapidly spreading day by day and doctors are not able to find the actual reason . Several factors like Age ,Hypertension , Smoking issue, Heart disease and many more can be the key factor of high diabetes. Most of them have multiple problems including diabetes so doctors need an actual clear view so that they can treat accordingly.

Analysis Goal :



- Diabetes data analysis involves using statistical and computational techniques to examine patterns and trends in diabetes. The process includes data cleaning, exploratory analysis, and visualization to understand key metrics such as blood glucose levels, Hypertension and treatment outcomes. Advanced analytics can identify risk factors, predict complications, and evaluate the effectiveness of interventions. The insights gained from this analysis can inform healthcare strategies, improve patient management, and guide policy decisions to enhance overall diabetes care.

Task:

- Calculate the average BMI of patients.
- List patients in descending order of blood glucose levels.
- Find patients who have hypertension and diabetes.
- Determine the number of patients with heart disease.
- Group patients by smoking history and count how many smokers and nonsmokers there are.
- Retrieve the Patient_ids of patients who have a BMI greater than the average BMI.
- Find the patient with the highest HbA1c level and the patient with the lowest HbA1clevel.
- Rank patients by blood glucose level within each gender group.
- Insert a new patient into the database with sample data.
- Delete all patients with heart disease from the database.
- Find patients who have hypertension but not diabetes using the EXCEPT operator.
- Define a unique constraint on the "patient_id" column to ensure its values are unique.

Visualization:

- **Question** : Calculate the average BMI of patients
- **Explain** : Here I have calculated average BMI(Body Mass Index) which is ‘27.36’ where normal range varies between 18.5-24.9. A BMI of less than 18.5 is considered underweight, while a BMI of 25–29.9 is considered overweight, and a BMI of 30 or higher is considered obese.
- **Query** : select round(avg(bmi),2) from dia_predict;
- **Result** :

	round(avg(bmi),2)
▶	27.36

Question : List patients in descending order of blood glucose levels.

- **Explain** : Here I have listed the patients according to descending blood glucose levels. The expected values for normal fasting blood glucose concentration are between 70 mg/dL (3.9 mmol/L) and 100 mg/dL (5.6 mmol/L). When fasting blood glucose is between 100 to 125 mg/dL (5.6 to 6.9 mmol/L) changes in lifestyle and monitoring glycemia are recommended . But here average level is 127.82

- **Query** : select patient_id,blood_glucose_level from dia_predict order by blood_glucose_level desc;

- **Result** :

patient_id	blood_glucose_level
PT18089	300
PT18711	300
PT16657	300
PT18122	300
PT17634	300

patient_id	blood_glucose_level
PT12471	280
PT11628	280
PT14195	280
PT10480	280
PT11076	280

patient_id	blood_glucose_level
PT18426	240
PT18404	240
PT15674	240
PT17549	240
PT17054	240

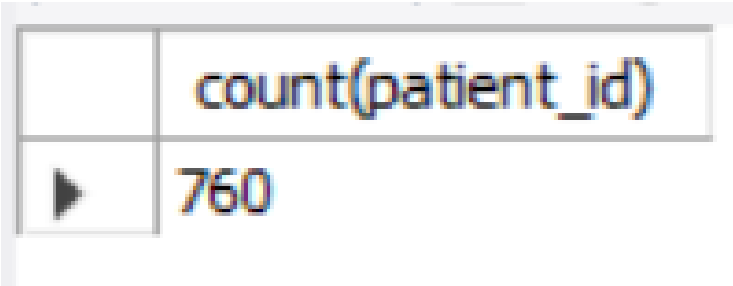
Question : Find patients who have hypertension and diabetes.

- **Explain** : Here I calculated the total number of patients who have both hypertension and diabetes which is 409. Here ‘1’ defines that have both the disease.
- **Query** : select count(*) from dia_predict where hypertension ='1' and diabetes='1';
- **Result** :

	count(*)
▶	409

Question : Determine the number of patients with heart disease

- **Explain :** Here I calculated the total number of patients who have heart disease which is 760. Here ‘1’ defines that have heart disease.
- **Query :** select count(patient_id) from dia_predict where heart_disease='1';
- **Result :**



	count(patient_id)
▶	760

Question : Group patients by smoking history and count how many smokers and nonsmokers there are.

- **Explain :** Here I divided Smokers into 4 category like ‘Former Smoker’ , ‘Current Smoker’, ‘Non-Smoker’ and ‘Others’ using CASE statement and calculated total no of patients for each group.

  No of ‘Former Smokers’ are 2587

  No of ‘Current Smokers’ are 1836

  No of ‘Non-Smokers’ are 13826

  No of ‘Others’ are 1230

➢ **Query :**

```
SELECT
 CASE
        WHEN smoking_history IN ('never', 'no info') THEN 'Non-smoker'
        WHEN smoking_history IN ('ever', 'former') THEN 'Former smoker'
        WHEN smoking_history IN ('no current', 'format') THEN 'Unknown'
        WHEN smoking_history = 'current' THEN 'Current smoker'
 ELSE 'Other'
 END AS smoking_status,
 COUNT(*) AS patients
  FROM
    dia_predict
GROUP BY
  CASE
      WHEN smoking_history IN ('never', 'no info') THEN 'Non-smoker'
      WHEN smoking_history IN ('ever', 'former') THEN 'Former smoker'
      WHEN smoking_history IN ('no current', 'format') THEN 'Unknown'
            WHEN smoking_history = 'current' THEN 'Current smoker'
      ELSE 'Other'
  END;
```

➢ **Output :**

	smoking_status	patients
▶	Non-smoker	13826
	Current smoker	1836
	Former smoker	2587
	Other	1230

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

- **Explain :** Here I have calculated 6592 patients who have BMI grater than average BMI and showing them in the form of Patient_id.
- **Query :** select patient_id,bmi from dia_predict where bmi > (select avg(bmi) from dia_predict);
- **Result :**

patient_id	bmi
PT109	33.64
PT112	54.7
PT113	36.05
PT117	30.36
PT121	36.38

Question : Find the patient with the highest HbA1c level and the patient with the lowest HbA1clevel.

- **Explain :** Here I have calculated patients who have highest HbA1c level(9) and lowest HbA1c level(3.5) by using MAX() and MIN() function.
- **Query :**

```
select EmployeeName,HbA1c_level
from dia_predict
where
HbA1c_level = (select max(HbA1c_level)
from dia_predict)
or
HbA1c_level = (select min(HbA1c_level)
from dia_predict) ;
```

- **Result :**

EmployeeName	HbA1c_level
ELLEN MOFFATT	3.5
JOHN TURSI	3.5
MICHAEL THOMPSON	9
SHARON MCCOLE WICHER	3.5
KEVIN CASHMAN	9

Question : Rank patients by blood glucose level within each gender group.

- **Explain :** Here I have calculated ranking of each patient by blood glucose level and gender group using RANK() function
- **Query :** select patient_id, gender, blood_glucose_level, rank() over (partition by gender order by blood_glucose_level) as glucose_level_rank from dia_predict;
- **Result :**

Patient_id	Gender	Blood_glucose_level	glucose_level_rank
PT18605	Female	80	1
PT18881	Female	80	1
PT18466	Female	80	1
PT18449	Female	80	1
PT18547	Female	80	1

Patient_id	Gender	Blood_glucose_level	glucose_level_rank
PT7235	Female	85	866
PT5594	Female	85	866
PT6487	Female	85	866
PT5566	Female	85	866
PT5995	Female	85	866

Patient_id	Gender	Blood_glucose_level	glucose_level_rank
PT8095	Female	140	5033
PT5567	Female	140	5033
PT4738	Female	140	5033
PT8121	Female	140	5033
PT4736	Female	140	5033

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

- **Explain :** Here I have calculated patients who have hypertension but not diabetes using EXCEPT operator
- **Query :**

```
select Patient_id from dia_predict where hypertension = 1
except
select Patient_id from dia_predict where diabetes = 1;
```

- **Result :**

Patient_id
PT105
PT129
PT143
PT155
PT161

Project Link:

- https://github.com/Chandan65171/Diabetes_data_analysis