Pradeep00000

SQL Diabetes Prediction Dataset

```sql
create schema diabet;
select * from diabet.diabetes_data;

#1. Count the number of users in the dataset.
select count(*) from diabet.diabetes_data;

#2. Retrieve all records where the blood glucose level is above 120.
select * from diabet.diabetes_data
where blood_glucose > 120 ;

#3. Get the average BMI across all users.
select avg(bmi) from diabet.diabetes_data;

```

```sql
#4. List the distinct stress levels recorded.
select distinct(stress_level)
from diabet.diabetes_data;

#5. Find the minimum and maximum risk scores.
select min(risk_score) as minimum , max(risk_score) as max
from diabet.diabetes_data;

#6. Find users who had physical activity greater than
#   30 minutes and blood glucose levels below 100.
select user_id , physical_activity , blood_glucose
from diabet.diabetes_data
where physical_activity > 30 and blood_glucose < 100;
```

```sql
#7. Calculate the average sleep hours grouped by stress level.
select avg(sleep_hours) , stress_level
from diabet.diabetes_data
group by stress_level;

#8. Get the count of users adhering to their diet and medication.
select count(user_id) , diet , medication_adherence
from diabet.diabetes_data
where diet = 1 and medication_adherence = 1;

#9. Rank users based on their risk score in descending order.
select user_id , risk_score, rank() over (order by risk_score desc) as 'rank'
from diabet.diabetes_data;
```

```sql
#10. Retrieve the top 5 users with the highest physical activity.
select user_id , physical_activity
from diabet.diabetes_data
order by physical_activity desc
limit 5;

#11. Calculate the total physical activity per user.
select user_id , sum(physical_activity)
from diabet.diabetes_data
group by user_id;

```

```sql
#12. Filter users with a BMI above 25 and below 30.
select user_id , bmi
from diabet.diabetes_data
where bmi between 25 and 30;

#13. Determine the percentage of users with blood
#    glucose levels below 100.
select (count(case when blood_glucose < 100 then 1 end)
* 100.0 / count(*)) as percentage_below_100
from diabet.diabetes_data;

#14. Retrieve users with the highest risk scores for each hydration level.
select  hydration_level , max(risk_score)
from diabet.diabetes_data
group by hydration_level;
```

```sql
#15. Find the user with the lowest risk score who is
#    not adhering to their medication.
select * from diabet.diabetes_data
where medication_adherence = 0
order by risk_score asc
limit 1;
```

```sql
#16. Calculate the BMI category distribution
#    (e.g., Underweight, Normal, Overweight, Obese).
select case
        when bmi < 18.5  then 'Underweight'
        when bmi between 18.5 and 24.9 then 'Normal'
        when bmi between 25 and 29.9 then 'Overweight'
        else 'Obese'
    end as bmi_category ,
    count(*) as category_count
from diabet.diabetes_data
group by bmi_category;
```

```sql
#17. Find anomalies where users with high physical activity
#    (>60 mins) still have high blood glucose (>140).
select user_id , physical_activity , blood_glucose
from diabet.diabetes_data
where physical_activity > 60 and blood_glucose > 140;

#18. Generate a report on the relationship between sleep
#    hours and risk scores, segmented by stress levels.
select stress_level , avg(sleep_hours) , avg(risk_score)
from diabet.diabetes_data
group by stress_level;
```