## **Project: Diabetes Data Analysis with SQL & Power BI**

### **Overview**

* **Objective:** Analyze patterns in diabetes-related medical diagnostics using SQL for querying and Power BI for visualization.
* **Dataset:** Pima Indians Diabetes Dataset
* **Tools:** MySQL, Power BI Desktop

SELECT \* FROM diabetes;

– CLEANED DATA BY REPLACING NULL VALUES WITH MEDIAN

SELECT \* FROM diabetes

WHERE Glucose = 0 OR Insulin = 0 OR BMI=0;

-- Step 1: Replace invalid 0s with NULLs

UPDATE diabetes SET glucose = NULL WHERE glucose = 0;

UPDATE diabetes SET bloodpressure = NULL WHERE bloodpressure = 0;

UPDATE diabetes SET skinthickness = NULL WHERE skinthickness = 0;

UPDATE diabetes SET insulin = NULL WHERE insulin = 0;

UPDATE diabetes SET bmi = NULL WHERE bmi = 0;

-- Step 2: Fill NULLs with median values

UPDATE diabetes SET glucose = 104 WHERE glucose IS NULL;

UPDATE diabetes SET bloodpressure = 75.5 WHERE bloodpressure IS NULL;

UPDATE diabetes SET skinthickness = 31.5 WHERE skinthickness IS NULL;

UPDATE diabetes SET insulin = 130 WHERE insulin IS NULL;

UPDATE diabetes SET bmi = 30.5 WHERE bmi IS NULL;

-- KPIS

SELECT

outcome,

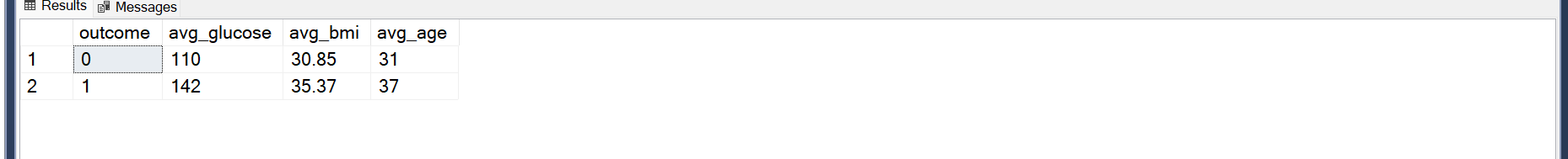
AVG(glucose) AS avg\_glucose,

ROUND(AVG(bmi),2) AS avg\_bmi,

AVG(age) AS avg\_age

FROM diabetes

GROUP BY outcome;



--PERCENTAGE DIABETIC

SELECT

CONCAT(CAST((Diabetic\*100/Total\_count) as DECIMAL (2)),'%') as percent\_diabetic

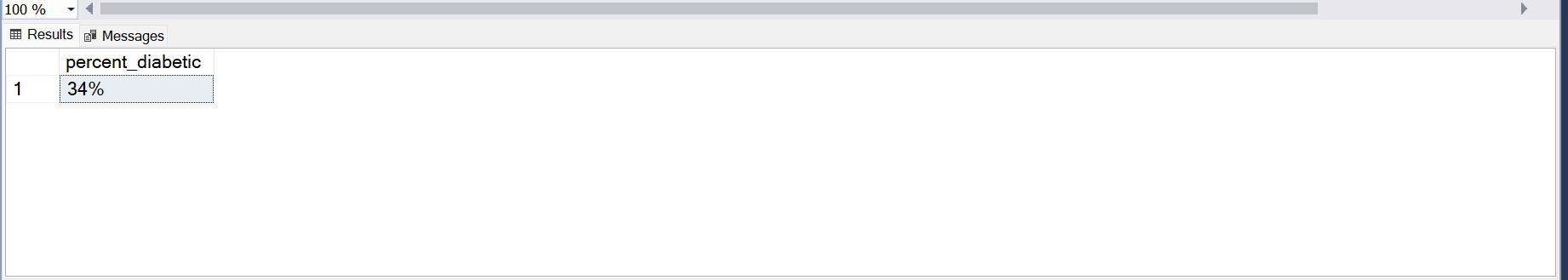
FROM

(SELECT

SUM(Outcome) as Diabetic,

COUNT(\*) as Total\_count

FROM diabetes)t



-- AGE GROUP OF PATIENTS COUNT WITH DIABETES

WITH CTE AS(

SELECT

CASE

WHEN age <= 30 THEN '20–30'

WHEN age <= 40 THEN '31–40'

WHEN age <= 50 THEN '41–50'

WHEN age <= 60 THEN '51–60'

WHEN age <= 70 THEN '61–70'

ELSE '71+'

END AS age\_group,

COUNT(\*) AS count,

sum(Outcome) as Outcome

FROM diabetes

GROUP BY

CASE

WHEN age <= 30 THEN '20–30'

WHEN age <= 40 THEN '31–40'

WHEN age <= 50 THEN '41–50'

WHEN age <= 60 THEN '51–60'

WHEN age <= 70 THEN '61–70'

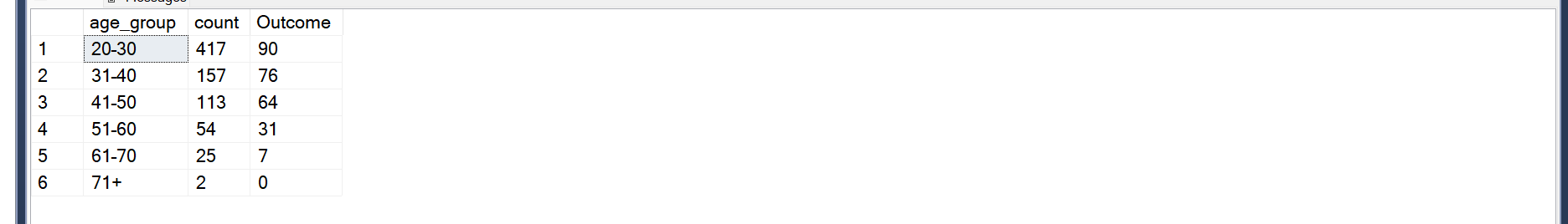
ELSE '71+'

END

)

SELECT \* FROM CTE

ORDER BY age\_group



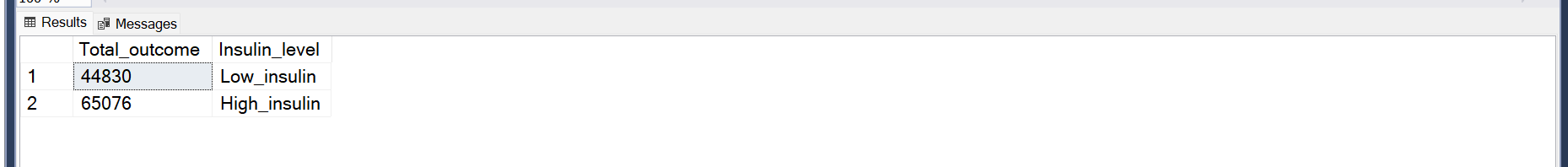
– INSULIN BY OUTCOME

SELECT sum(Insulin) as Total\_outcome,'Low\_insulin' as Insulin\_level from diabetes

where Outcome=1

union

SELECT sum(Insulin) as Total\_outcome, 'High\_insulin' as Insulin\_level FROM diabetes where Outcome = 0

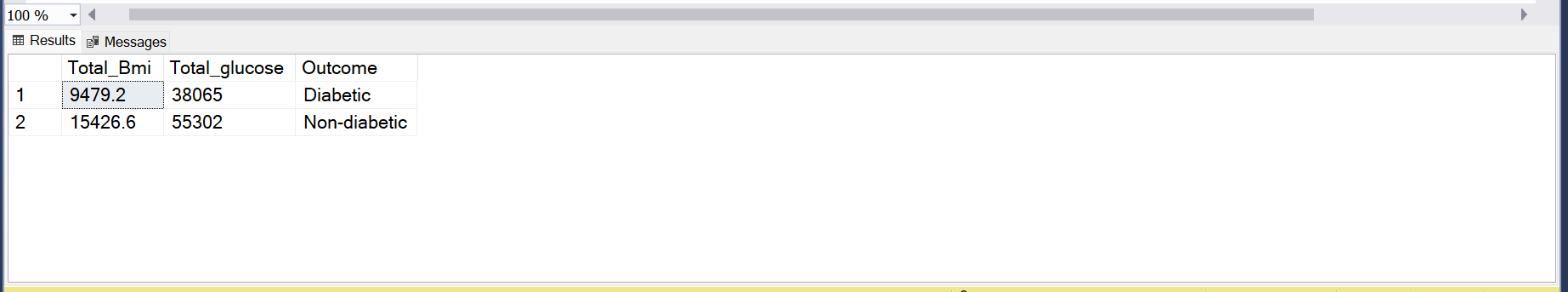


– BMI and GLUCOSE BY OUTCOME

SELECT round(sum(BMI),2) as Total\_Bmi,sum(Glucose) as Total\_glucose, 'Diabetic' as Outcome from diabetes where Outcome = 1

union

SELECT round(sum(BMI),2) as Total\_Bmi,sum(Glucose) as Total\_glucose , 'Non-diabetic' as Outcome from diabetes where Outcome = 0



### **Insights:**

* Do people with higher glucose or BMI levels tend to be diabetic?

Answer: The higher the BMI and GLUCOSE, the less likely the person is to get diabetic

* Which age group has the highest percentage of diabetics?

Answer: The age group of 20 - 30 has the highest number of 90 with 417 case

* How does insulin level vary with outcome?

Answer: The lower the insulin level (44830), the more diabetic a person is likely to be. The higher the insulin level(65076) the less diabetic.