Healthcare Data

Summary:

This project involved cleaning and normalizing a healthcare dataset for efficient storage and SQL querying. After addressing inconsistencies and duplicates, the data was reorganized into multiple related tables following 3NF principles, including Patients, Doctors, Hospitals, Admissions, Admission-Medications, and Admission-Testes.

Normalization Report

1st Normal Form (1NF):

- Ensured each cell contains atomic (single) values
- Split multivalued fields (e.g., medications, test results) into separate records

2nd Normal Form (2NF):

 Moved data into separate tables so that all non-key columns depend on the full primary key

3rd Normal Form (3NF):

- Removed transitive dependencies
- For example, doctors are linked to hospitals through hospital_id instead of repeating hospital names

1. Patients

Column	Description
patient_id	Primary Key
name	Patient's name
age	Patient's age
gender	Male / Female
blood_type	Blood group
insurance_provider	(Optional)

2. Doctors

Column	Description		
doctor_id	Primary Key		
name	Doctor's name		
Hospital ID	Foreign Key → Hospitals		

3. Hospitals

Column	Description		
Hospital ID	Primary Key		
hospital_name	Hospital's name		

4. Admissions

Column	Description		
Admisions ID	Primary Key		
Patient-ID	Foreign Key → Patients		
doctor_id	Foreign Key → Doctors		
Hospital ID	Foreign Key → Hospitals		
Room Number	Room assigned		
Admission Type	e.g., Emergency, Urgent		
Medical Condition	Diagnosis, if available		
Date of Admission	Date of admission		
Discharge Date	Date of discharge		
Billing Amount	Total billing for the stay		

5. Admission-Medications

Column	Description		
admission_id	Foreign Key → Admissions		
medication_name	Name of medication used		

6.Admission-Testes.

Column	Description		
test_id	Primary Key		
admission_id	Foreign Key → Admissions		
result	Result: Normal / Abnormal		

⊘ Table Relationships Overview

- Patients → Admissions via patient_id
- Doctors → Admissions via doctor_id
- Hospitals → both Doctors and Admissions via hospital_id
- Admissions → Admission_Medications and Test_Results via admission_id

Data Modeling

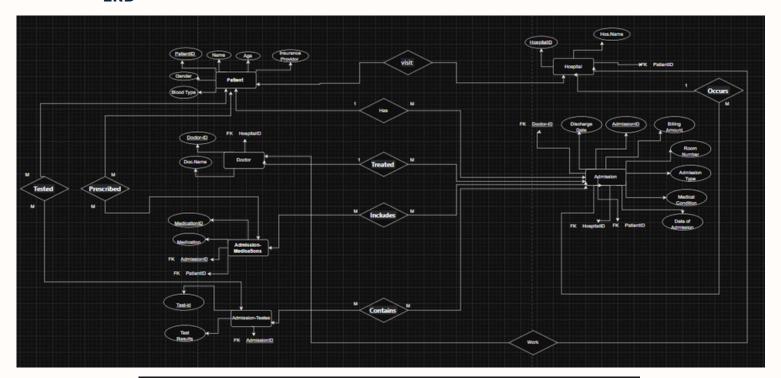
Entity-Relationship Model

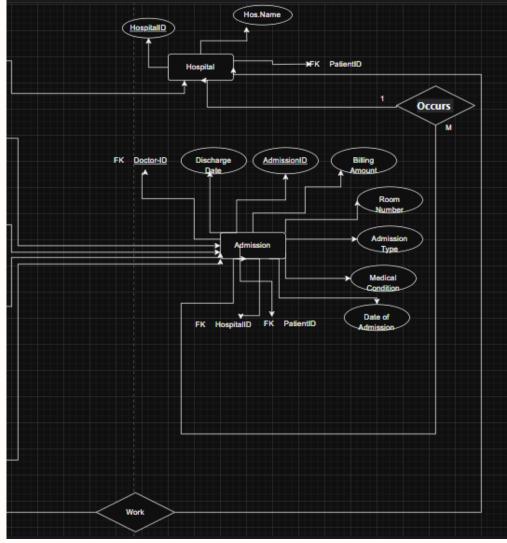
- Entities and Attributes
 - 1. Patient: PatientID (PK) Name Age Gender BloodType Insurance Provider.
 - 2. Hospital: HospitalID (PK), HospitalName
 - 3. **Doctor:** DoctorID (PK), DoctorName, HospitalID (FK)
 - 4. **Admission:** AdmissionID (PK), PatientID (FK), HospitalID (FK), DoctorID (FK), DateOfAdmission, AdmissionType, DischargeDate, RoomNumber, BillingAmount
 - 5. Admission-Medications: Medication ID (PK), PatientID (FK), AdmissionID (FK), Medication
 - 6.Admission-Testes: PatientTestResultID (PK), PatientID (FK), MedicationID (FK), AdmissionID (FK).

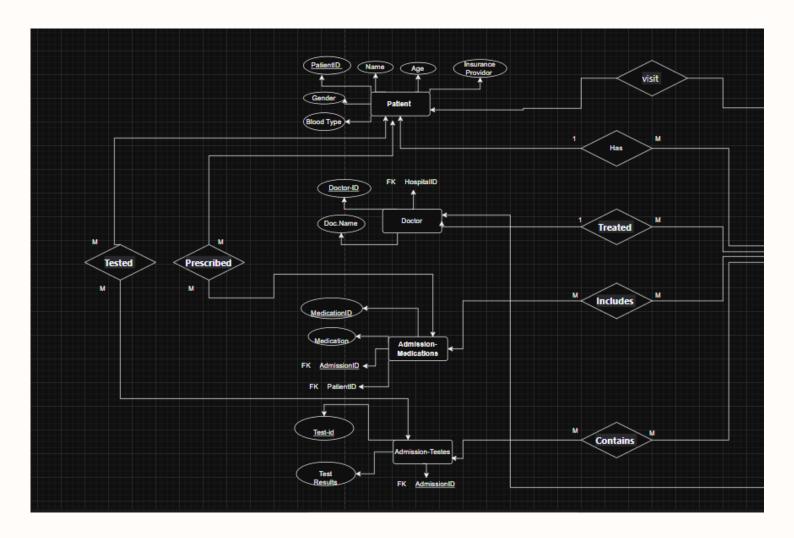
Relationships

- 1. Patient → Admission (One-to-Many)
- 2. Hospital → Admission (One-to-Many)
- 3. Doctor → Admission (One-to-Many)
- 4.InsuranceProvider → Admission (One-to-Many)
- 5. Patient ↔ Medication (Many-to-Many)
- 6. Patient ↔ TestResult (Many-to-Many)

• ERD

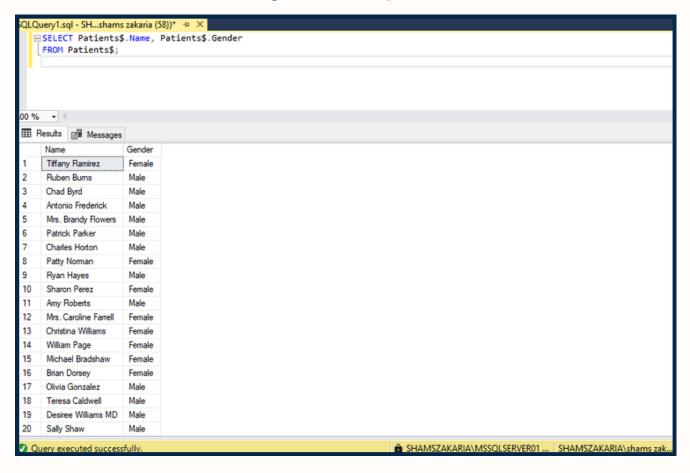




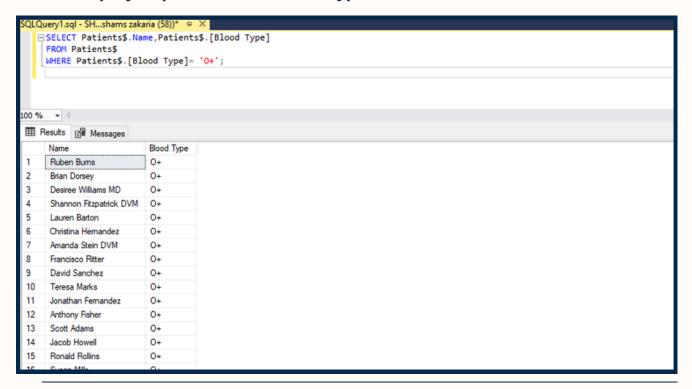


SQL Queries

1- Retrieve the names and genders of all patients.

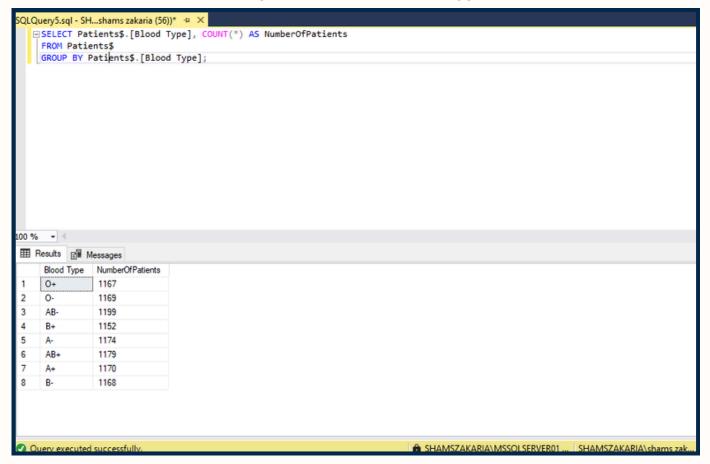


2- Display all patients with blood type O+.

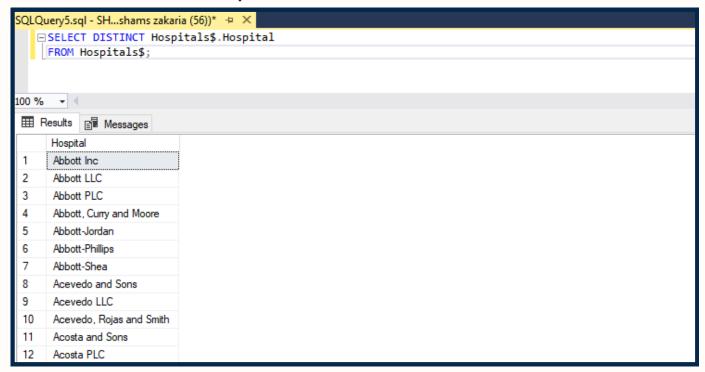


SQL Queries

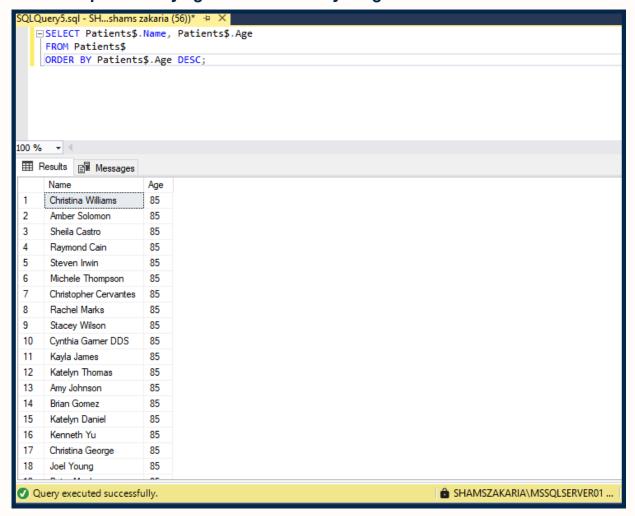
3- Count the number of patients in each blood type.



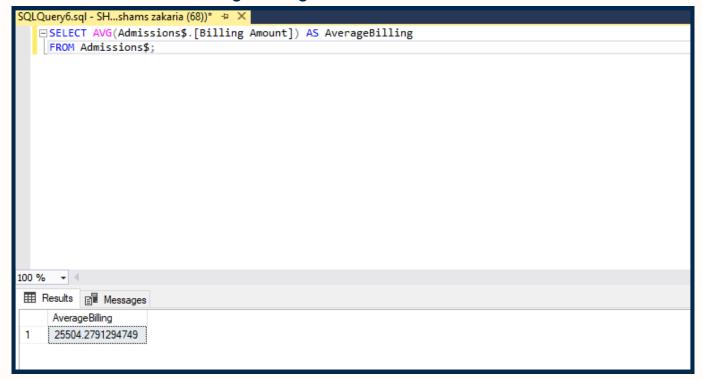
4-List all distinct hospital names.



5- Order patients by age from oldest to youngest.

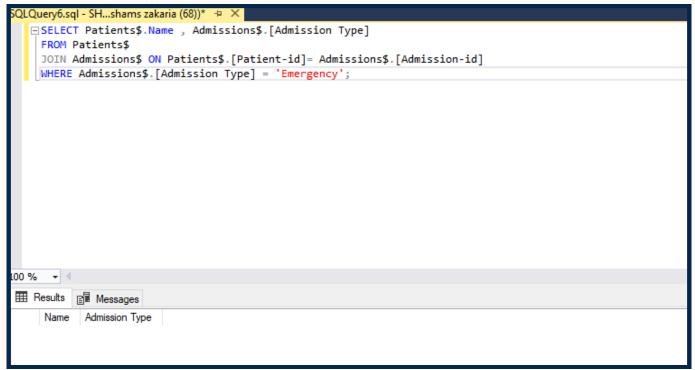


6- Calculate the average billing amount for all admissions.

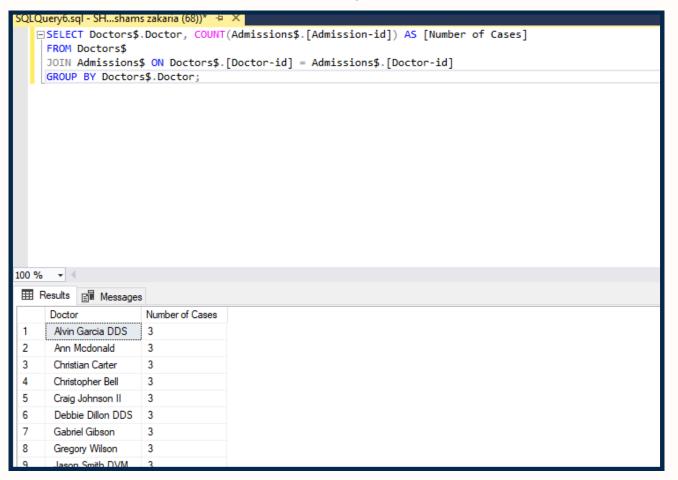


7- Display patients admitted through 'Emergency'.

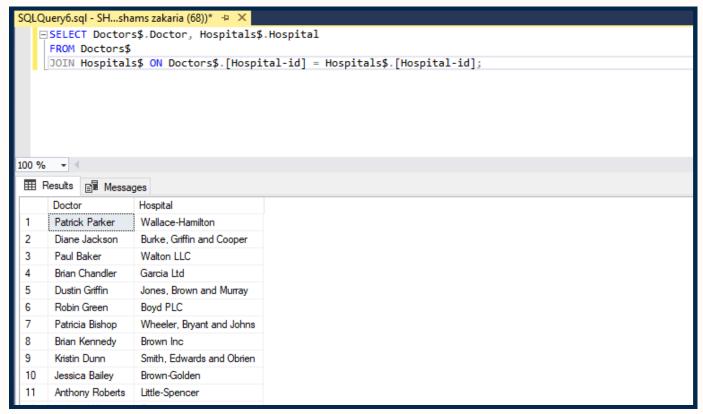
Note: The query is correct and runs without errors, but it returns no results because there are no records in the database that match the condition



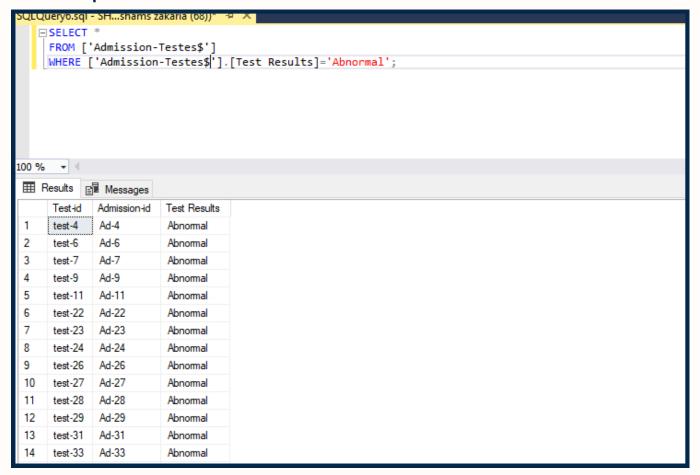
8- Count the number of cases handled by each doctor.



9- Show the doctor name and the hospital they work in.



10- List patients whose test results are 'Abnormal'.

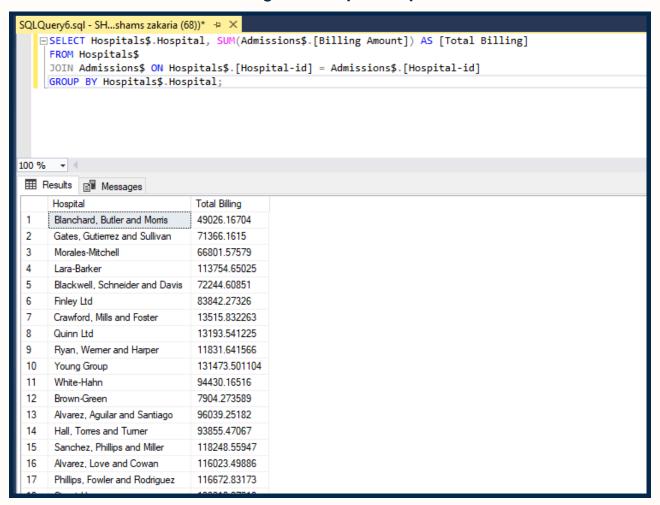


11- Display all patients treated in a hospital named 'Cairo Hospital' using JOIN.

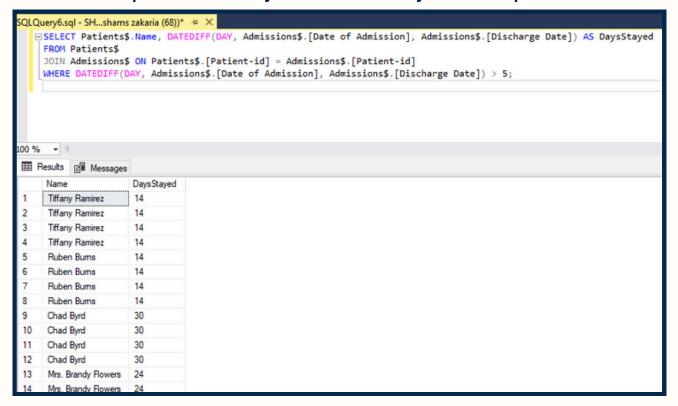
Last Update: April 2025

Note: The query is correct and runs without errors, but it returns no results because there are no records in the database that match the condition

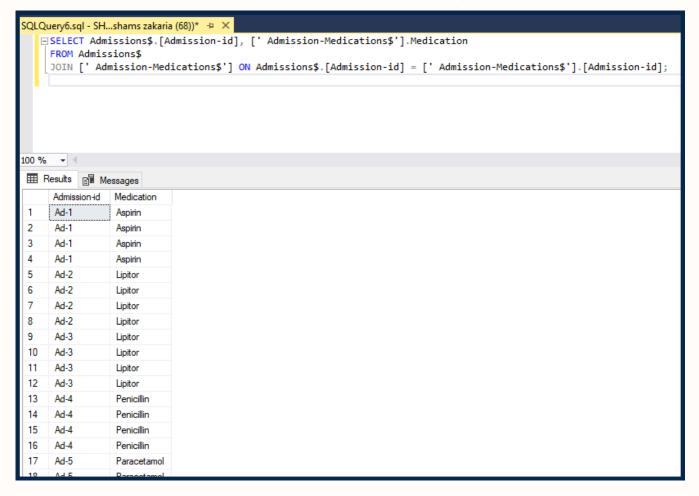
12- Calculate the total billing amount per hospital.



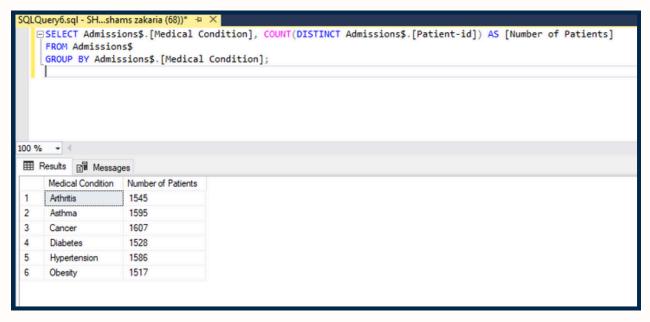
13- List the patients who stayed more than 5 days in the hospital.



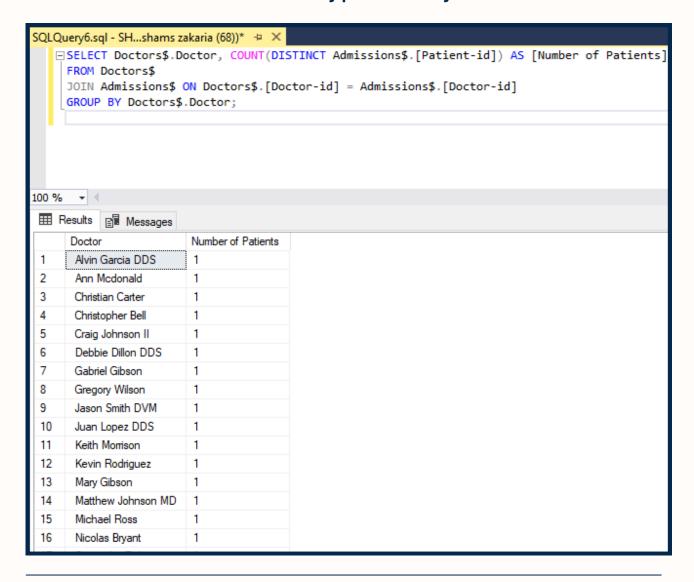
14- Show all medications used per admission using JOIN.



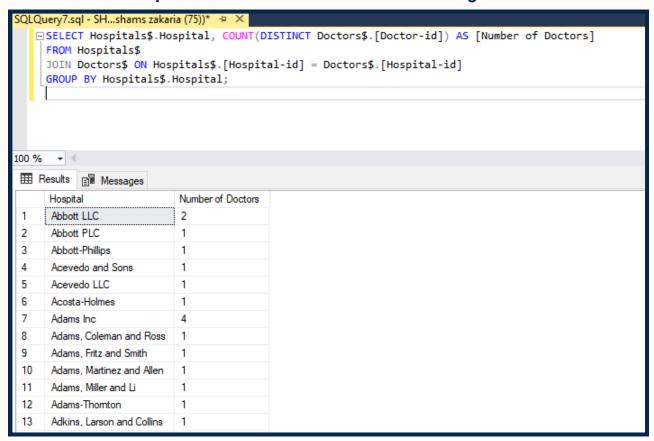
15- Count the number of patients per medical condition.



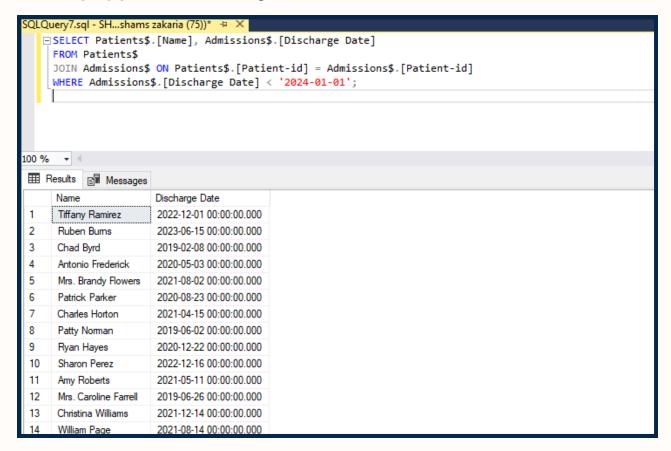
16- Show each doctor and how many patients they treated.



17- List each hospital and the number of doctors working there.



18- Display patients discharged before '2024-01-01'.



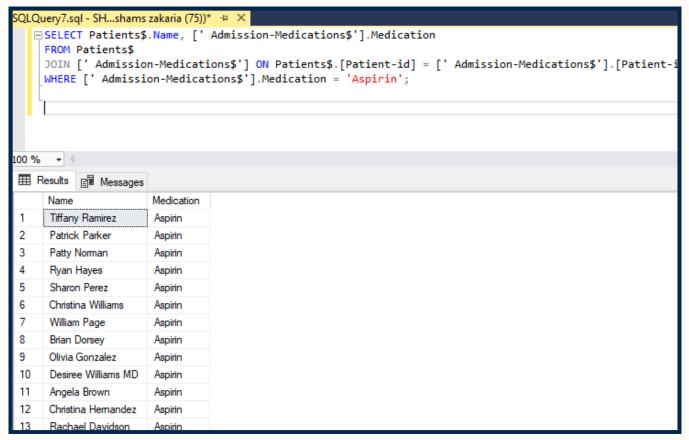
19- Show patients with billing amounts above the average (use subquery).

```
SQLQuery7.sql - SH...shams zakaria (75))* 😕 🗶

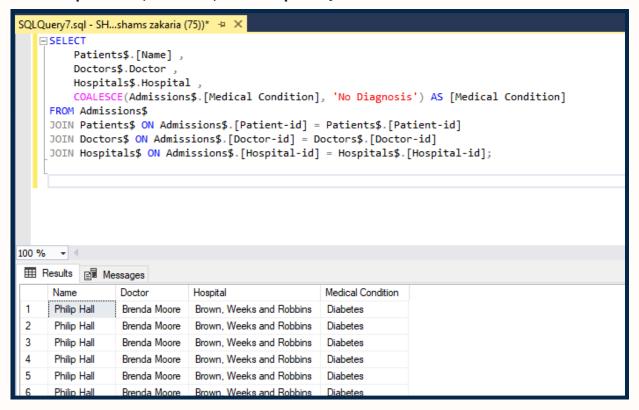
☐SELECT Patients$.[Name], Admissions$.[Billing Amount]

     FROM Patients$
     JOIN Admissions$ ON Patients$.[Patient-id] = Admissions$.[Patient-id]
     WHERE Admissions $. [Billing Amount] > (
          SELECT AVG([Billing Amount])
          FROM Admissions$
100 %
Results 📳 Messages
      Name
                       Billing Amount
     Tiffany Ramirez
                       37490.98336
 2
      Ruben Bums
                       47304.06485
 3
      Chad Byrd
                       36874.897
      Charles Horton
                       39593.43576
 5
      Amy Roberts
                       40325.07139
 6
      Brian Dorsey
                       27174.94291
      Teresa Caldwell
                       45213.53763
 8
      William Johnson
                       32263.62216
 9
      Steven Bennett
                       42610.70456
 10
      Beverly Miller
                       41319.50032
 11
      Daniel Dickson
                       37766 52124
                        35834 3566
```

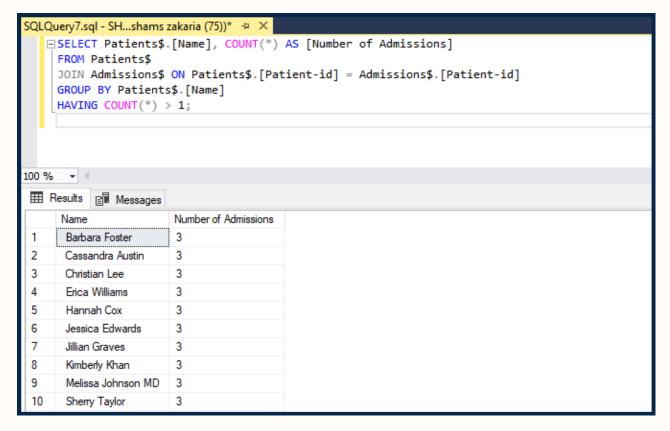
20- List patients who were given the medication 'Aspirin'.



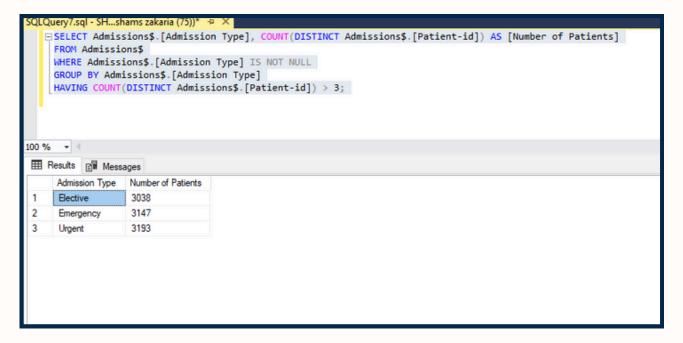
21- Show each patient's name, their doctor's name, and hospital name. If the medical condition is NULL, show 'No Diagnosis' (use COALESCE with JOINs across patients, doctors, and hospitals).



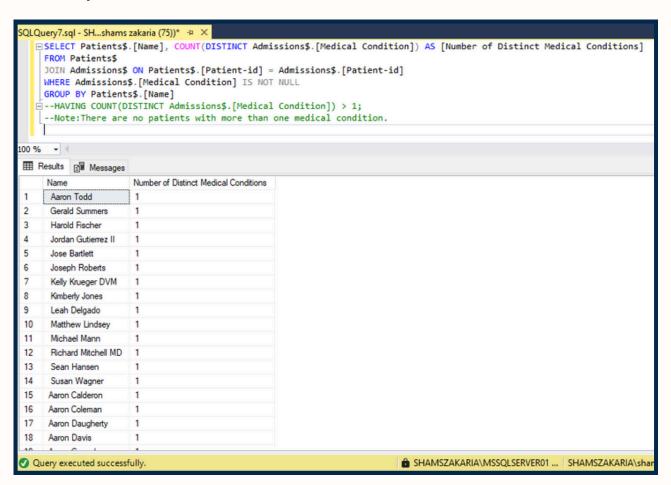
22- List patients who were admitted more than once, showing their name and number of admissions (use GROUP BY and HAVING COUNT(*) > 1).



23- Count the number of patients per admission_type, only showing those with more than 3 and where admission_type is not NULL.



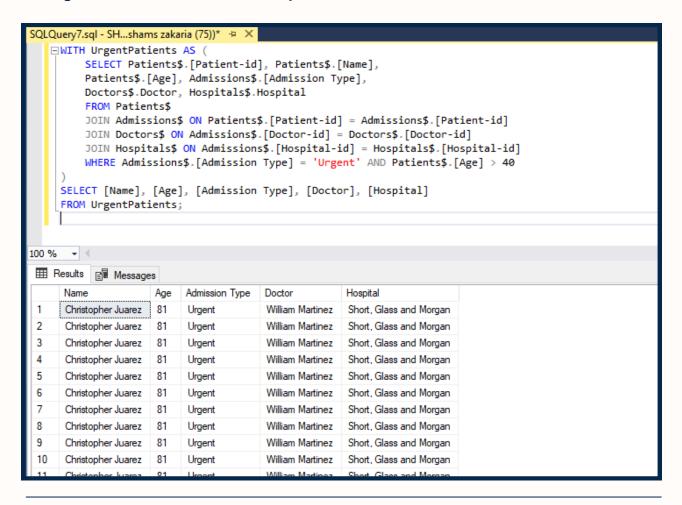
24- List patients with more than one distinct medical condition using JOINs across patients, admissions, and medical conditions.



25- Show hospitals where the total billing is more than 50,000 EGP and display the total billing amount (use GROUP BY + SUM).

SQLC	SQLQuery7.sql - SHshams zakaria (75))* → ×				
	□ SELECT Hospitals\$.Hospital, SUM(Admissions\$.[Billing Amount]) AS [Total Billing] FROM Hospitals\$ JOIN Admissions\$ ON Hospitals\$.[Hospital-id] = Admissions\$.[Hospital-id]				
	GROUP BY Hospitals\$.Hos		.ai-id] = Admissions\$.[nospicai-id]		
	HAVING SUM(Admissions\$.[Billing Amount]) > 50000;				
100 %	6 +				
▦	Results Messages				
	Hospital	Total Billing			
1	Silva, Adams and Price	60691.33653			
2	Moreno PLC	93753.0642			
3	Cochran-King	145987.9086			
4	Weiss Inc 105501.92799				
5	Brooks, Wright and Olson	57958.22283			
6	Wheeler-Cowan	50316.21429			
7	Burke, Mora and Shannon	135622.27347			
Q	Rivera-Knight	128686 93146			

26- Using a CTE, show patients over age 40 who were admitted as 'Urgent', along with their doctor and hospital name.



27- Use a subquery to find all patients treated by the same doctor who treated a patient named 'Ahmed'.

```
SQLQuery7.sql - SH...shams zakaria (75)* ** X

SELECT Patients$. [Name], Admissions$. [Doctor-id]

FROM Patients$

JOIN Admissions$ ON Patients$. [Patient-id] = Admissions$. [Patient-id]

WHERE Admissions$. [Doctor-id]

FROM Patients$

JOIN Admissions$ ON Patients$. [Patient-id] = Admissions$. [Patient-id]

WHERE Patients$. [Name] = 'Ahmed'

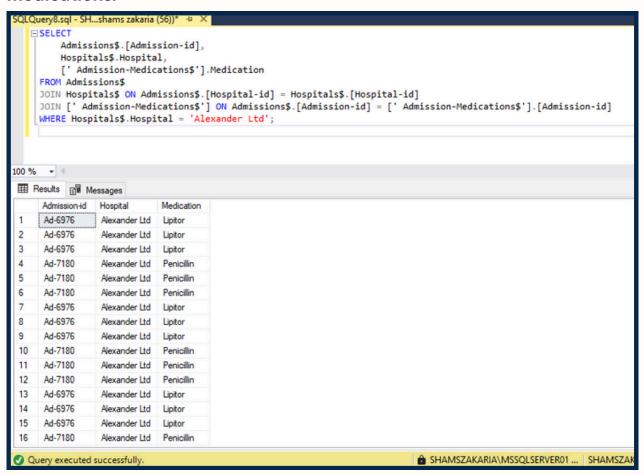
)

100 % **

THE Regulte SH Manager

Admissions$ ON Patients$ ON Patients$ ON Patients$ ON Patients$ ON Patients$ ON Patients$ ON Patient ON
```

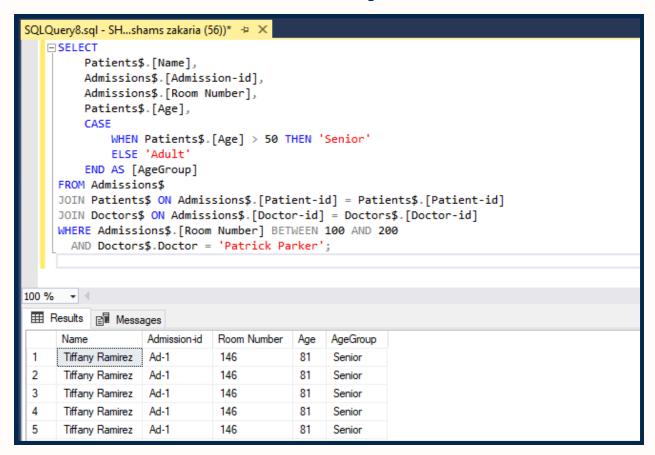
28- Show all admissions at 'Alex Medical Center' with the medication names and quantities using joins across admissions, hospitals, and medications.



29- List patients who were discharged 10 or more days after admission (use DATEDIFF or date arithmetic).

SQLQı	QLQuery8.sql - SHshams zakaria (56))* → ×					
SQLQuery8.sql - SHshams zakaria (56))* → × SELECT Patients\$.[Name], Admissions\$.[Admission-id], Admissions\$.[Date of Admission], Admissions\$.[Discharge Date], DATEDIFF(DAY, Admissions\$.[Date of Admission], Admissions\$.[Discharge Date]) AS [DaysStayed] FROM Admissions\$						
JOIN Patients\$ ON Admissions\$.[Patient-id] = Patients\$.[Patient-id] WHERE DATEDIFF(DAY, Admissions\$.[Date of Admission], Admissions\$.[Discharge Date]) >= 10; 100 % Results Results Results Results Results						
	Name	Admission-id	Date of Admission	Discharge Date	DaysStayed	
1	Tiffany Ramirez	Ad-1	2022-11-17 00:00:00.000	2022-12-01 00:00:00.000	14	
2	Tiffany Ramirez	Ad-1	2022-11-17 00:00:00.000	2022-12-01 00:00:00.000	14	
3	Tiffany Ramirez	Ad-1	2022-11-17 00:00:00.000	2022-12-01 00:00:00.000	14	
4	Tiffany Ramirez	Ad-1	2022-11-17 00:00:00.000	2022-12-01 00:00:00.000	14	
5	Mark Moss	Ad-1000	2022-06-14 00:00:00.000	2022-06-27 00:00:00.000	13	
6	Mark Moss	Ad-1000	2022-06-14 00:00:00.000	2022-06-27 00:00:00.000	13	
7	Mark Moss	Ad-1000	2022-06-14 00:00:00.000	2022-06-27 00:00:00.000	13	
0	Disilia Liali	Ad 1000	2021 02 01 00:00:00 000	2021 02 10 00:00:00 000	15	

30- Show patients admitted to rooms between 100 and 200, treated by 'Dr. Mona', and include a column that displays 'Senior' if the patient is older than 50, else 'Adult' (use CASE, JOIN, and filtering).



31- Show the number of admissions per insurance provider.

```
SQLQuery8.sql - SH...shams zakaria (56))* 垣 🗙
   □ SELECT
         Patients$.[Insurance Provider],
         COUNT(Admissions$.[Admission-id]) AS [NumberOfAdmissions]
    FROM Admissions$
    JOIN Patients$ ON Admissions$.[Patient-id] = Patients$.[Patient-id]
    GROUP BY Patients$.[Insurance Provider];
100 % 🕶 🔻
Results 📳 Messages
     Insurance Provider NumberOfAdmissions
     Cigna
                      5783
2
     Medicare
                      5422
     Blue Cross
                      5716
     UnitedHealthcare
                      5524
     Aetna
                      5789
```