-- DATA ANALYSIS USING SQL ON HealthCare Database

-- This dataset is not based on real facts, please don't consider the result sets to be actual and utilize it for any purpose.

-- Creating Database named Healthcare.

```
Create Database Hospital1;
```

-- Selecting Healthcare database to query.

```
USE Hospital1;
```

-- Viewing Data on Database

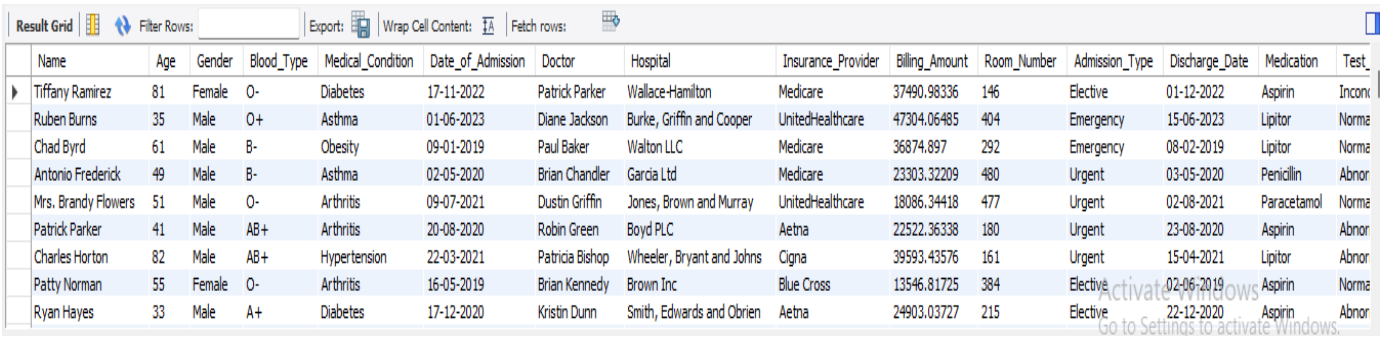
```
SELECT * FROM hospital1
```

-- Describing characteristics of table.

```
DESC hospital1
```

-- 1. Counting Total Record in Database

```
select count(*) hospital1
```

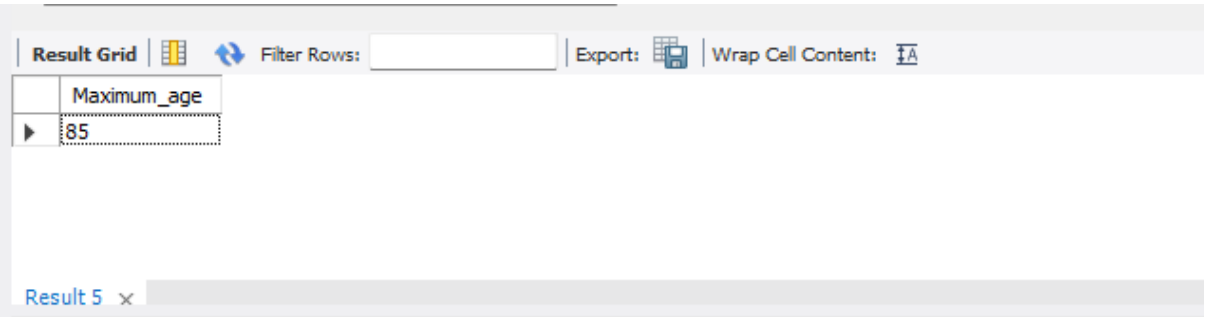


Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission	Doctor	Hospital	Insurance_Provider	Billing_Amount	Room_Number	Admission_Type	Discharge_Date	Medication	Test_
Tiffany Ramirez	81	Female	O-	Diabetes	17-11-2022	Patrick Parker	Wallace-Hamilton	Medicare	37490.98336	146	Elective	01-12-2022	Aspirin	Incon
Ruben Burns	35	Male	O+	Asthma	01-06-2023	Diane Jackson	Burke, Griffin and Cooper	UnitedHealthcare	47304.06485	404	Emergency	15-06-2023	Lipitor	Norma
Chad Byrd	61	Male	B-	Obesity	09-01-2019	Paul Baker	Walton LLC	Medicare	36874.897	292	Emergency	08-02-2019	Lipitor	Norma
Antonio Frederick	49	Male	B-	Asthma	02-05-2020	Brian Chandler	Garcia Ltd	Medicare	23303.32209	480	Urgent	03-05-2020	Penicillin	Abnor
Mrs. Brandy Flowers	51	Male	O-	Arthritis	09-07-2021	Dustin Griffin	Jones, Brown and Murray	UnitedHealthcare	18086.34418	477	Urgent	02-08-2021	Paracetamol	Norma
Patrick Parker	41	Male	AB+	Arthritis	20-08-2020	Robin Green	Boyd PLC	Aetna	22522.36338	180	Urgent	23-08-2020	Aspirin	Abnor
Charles Horton	82	Male	AB+	Hypertension	22-03-2021	Patricia Bishop	Wheeler, Bryant and Johns	Cigna	39593.43576	161	Urgent	15-04-2021	Lipitor	Abnor
Patty Norman	55	Female	O-	Arthritis	16-05-2019	Brian Kennedy	Brown Inc	Blue Cross	13546.81725	384	Elective	02-06-2019	Aspirin	Norma
Ryan Hayes	33	Male	A+	Diabetes	17-12-2020	Kristin Dunn	Smith, Edwards and Obrien	Aetna	24903.03727	215	Elective	22-12-2020	Aspirin	Abnor

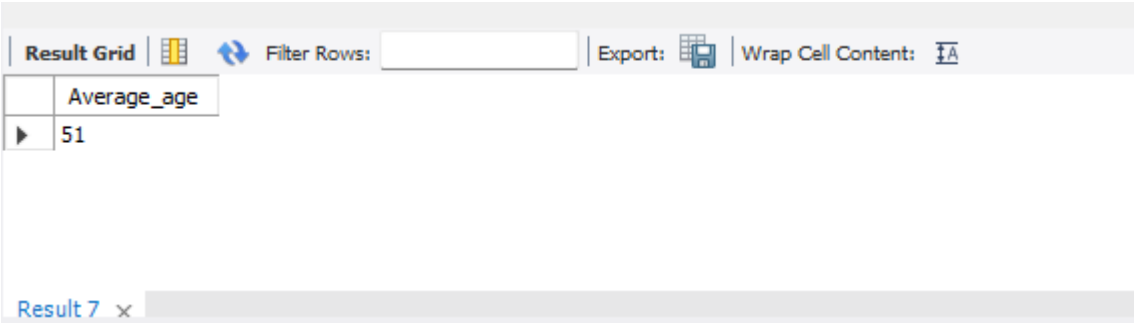
-- 2. Finding maximum age of patient admitted.

```
select max(age) as Maximum_age from hospital1
```



-- 3. Finding Average age of hospitalized patients.

select round(avg(age),0) as Average_age from hospital1



-- 4. Calculating Patients Hospitalized Age-wise from Maximum to Minimum

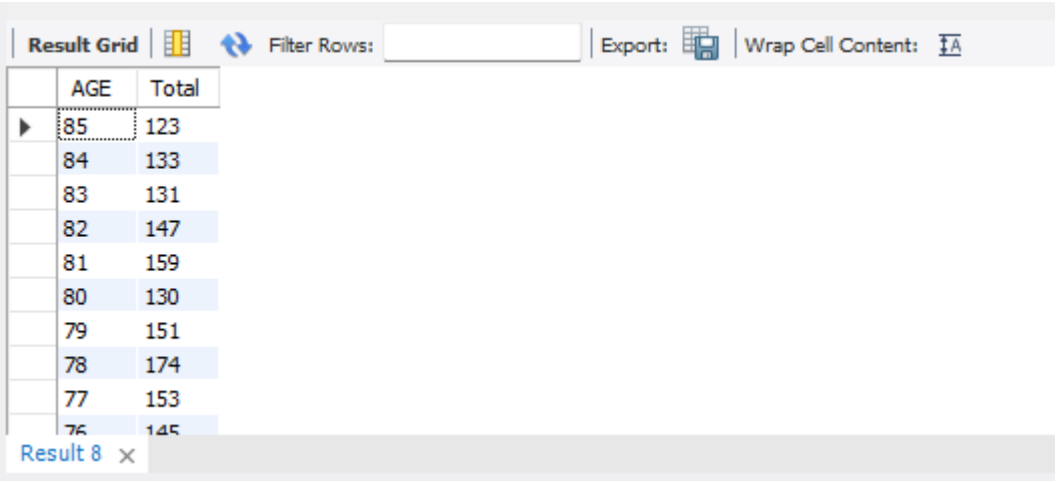
SELECT AGE, COUNT(AGE) AS Total

FROM Hospital1

GROUP BY age

ORDER BY AGE DESC;

-- Findings : The output will display a list of unique ages present in the "Healthcare" table along with the count of occurrences for each age, sorted from oldest to youngest.



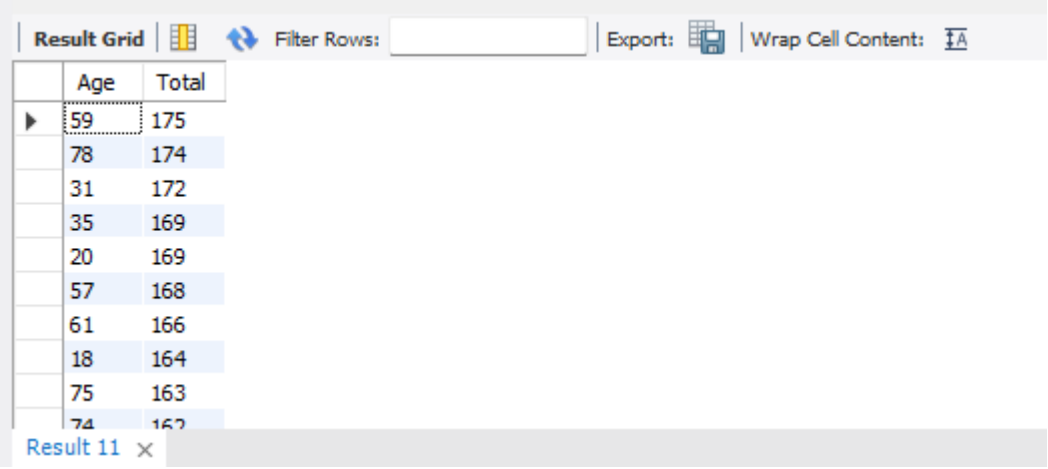
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

AGE	Total
85	123
84	133
83	131
82	147
81	159
80	130
79	151
78	174
77	153
76	145

Result 8

-- 5. Calculating Maximum Count of patients on basis of total patients hospitalized with respect to age.

```
select Age, count(age) as Total

from Hospital1

group by age

order by Total desc,age desc;
```



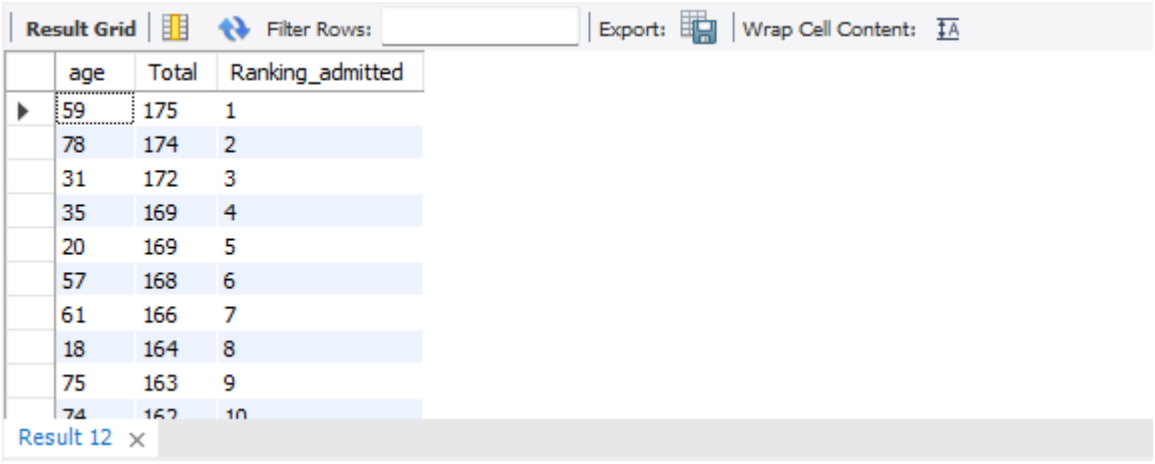
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Age	Total
59	175
78	174
31	172
35	169
20	169
57	168
61	166
18	164
75	163
74	162

Result 11

-- 6. Ranking Age on the number of patients Hospitalized

```
select age, count(age) as Total, dense_rank() OVER(ORDER BY COUNT(AGE) DESC, age DESC) as Ranking_admitted

from hospital1

group by age

Having Total > Avg(age);
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

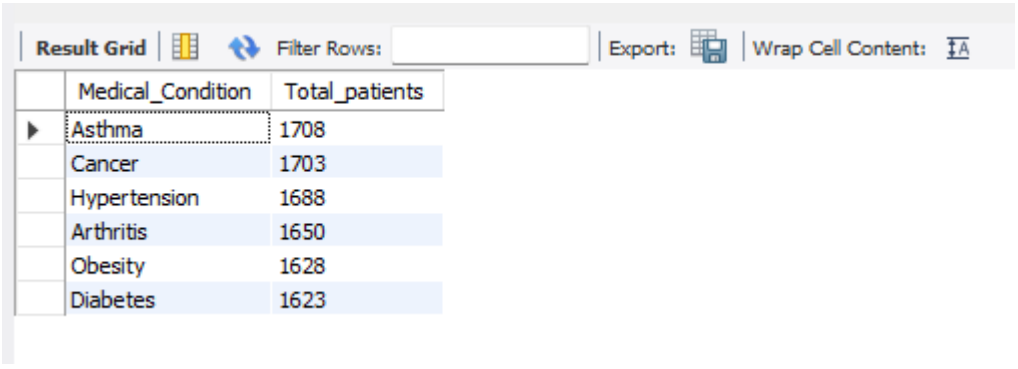
age	Total	Ranking_admitted
59	175	1
78	174	2
31	172	3
35	169	4
20	169	5
57	168	6
61	166	7
18	164	8
75	163	9
74	162	10

Result 12

-- 7. Finding Count of Medical Condition of patients and lisitng it by maximum no of patients.

```
SELECT Medical_Condition, count(Medical_Condition) as Total_patients

FROM hospital1

GROUP BY medical_condition

Order by Total_patients Desc;
```

-- Findings : This query retrieves a breakdown of medical conditions recorded in a healthcare dataset along with the total number of patients diagnosed with each condition. It groups the data by distinct medical conditions, counting the occurrences of each condition across the dataset. The result is presented in descending order based on the total number of patients affected by each medical condition, providing an insight into the prevalence or frequency of various health issues within the dataset



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Medical_Condition	Total_patients
Asthma	1708
Cancer	1703
Hypertension	1688
Arthritis	1650
Obesity	1628
Diabetes	1623

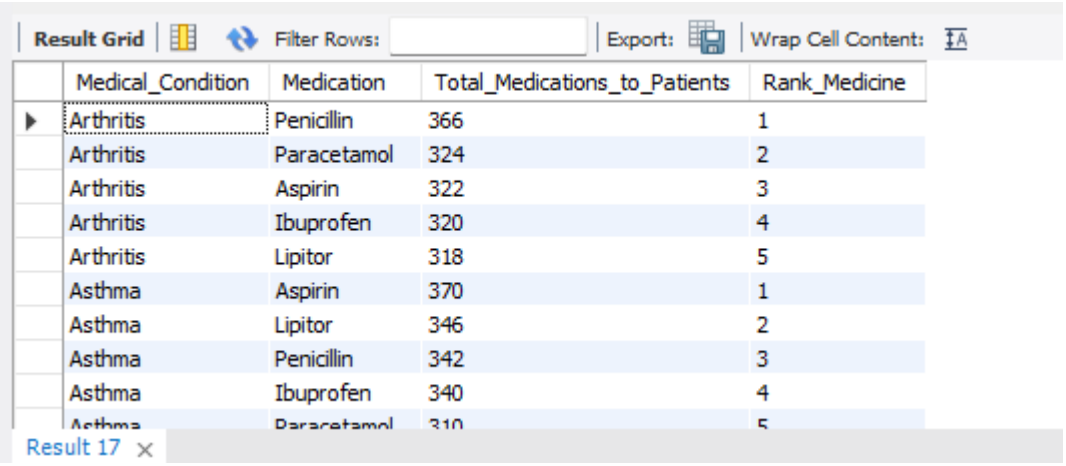
-- 8. Finding Rank & Maximum number of medicines recommended to patients based on Medical Condition pertaining to them.

```
SELECT Medical_Condition, Medication, COUNT(medication) as
Total_Medications_to_Patients, RANK() OVER(PARTITION BY Medical_Condition ORDER BY
COUNT(medication) DESC) as Rank_Medicine

FROM hospital1
```

GROUP BY 1,2

ORDER BY 1;

-- Finding : The output provides insight into the most common medications used for various medical conditions, assigning a rank to each medication based on how frequently its prescribed within its corresponding condition.



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Medical_Condition	Medication	Total_Medications_to_Patients	Rank_Medicine
Arthritis	Penicillin	366	1
Arthritis	Paracetamol	324	2
Arthritis	Aspirin	322	3
Arthritis	Ibuprofen	320	4
Arthritis	Lipitor	318	5
Asthma	Aspirin	370	1
Asthma	Lipitor	346	2
Asthma	Penicillin	342	3
Asthma	Ibuprofen	340	4
Asthma	Paracetamol	310	5

Result 17

-- 9. Most preffered Insurance Provide by Patients Hospatilized

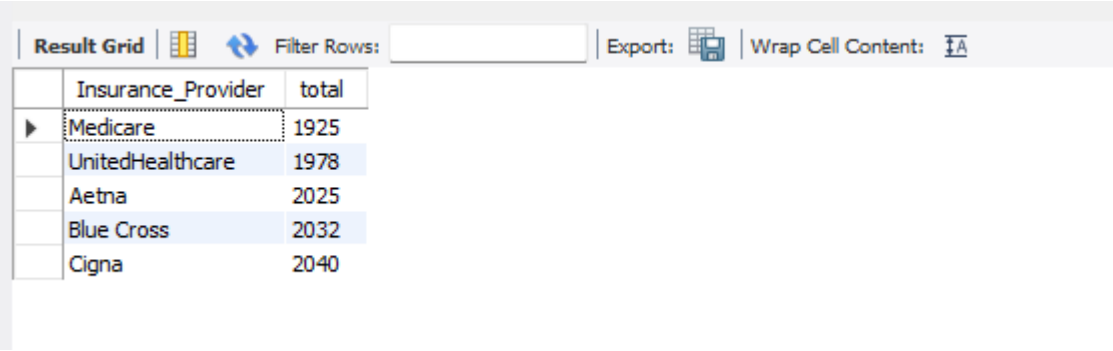
select Insurance_Provider, count(insurance_provider) as total

From Hospital1

group by insurance_provider

order by total

-- Findings : This information helps identify the most prevalent insurance providers among the patient population, offering valuable data for resource allocation, understanding coverage preferences, and potentially indicating trends in healthcare accessibility based on insurance networks



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Insurance_Provider	total
Medicare	1925
UnitedHealthcare	1978
Aetna	2025
Blue Cross	2032
Cigna	2040

-- 10. Finding out most preffered Hospital

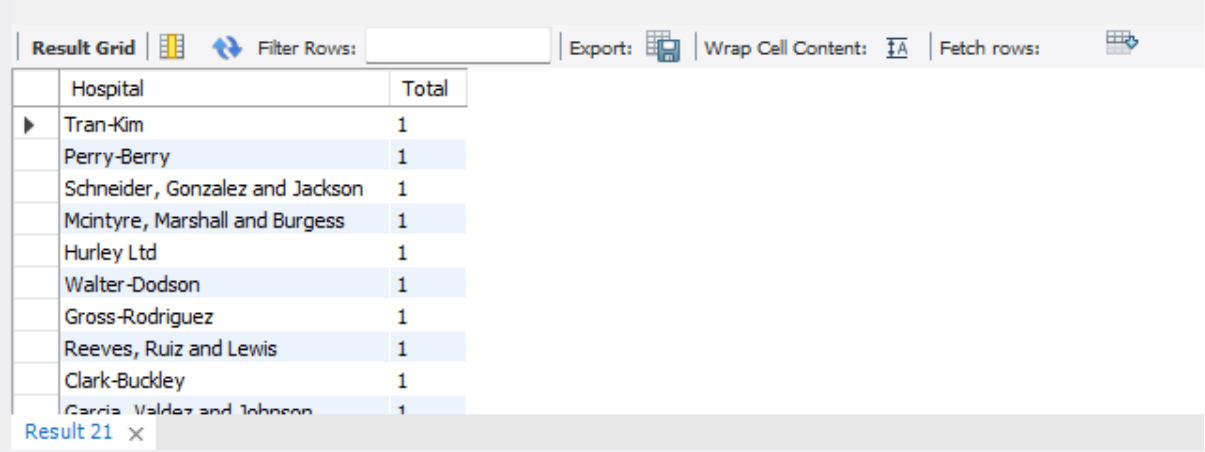
Select Hospital, Count(hospital) as Total

from hospital1

group by Hospital

order by Total;

-- Findings : It provides insight into which hospitals have the highest frequency of records within the healthcare dataset. The resulting list showcases hospitals based on their patient count or the number of entries related to each hospital, allowing for an understanding of the distribution or prominence of healthcare services among different medical facilities.


Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Hospital	Total
Tran-Kim	1
Perry-Berry	1
Schneider, Gonzalez and Jackson	1
Mcintyre, Marshall and Burgess	1
Hurley Ltd	1
Walter-Dodson	1
Gross-Rodriguez	1
Reeves, Ruiz and Lewis	1
Clark-Buckley	1
Garcia, Valdez and Johnson	1

Result 21

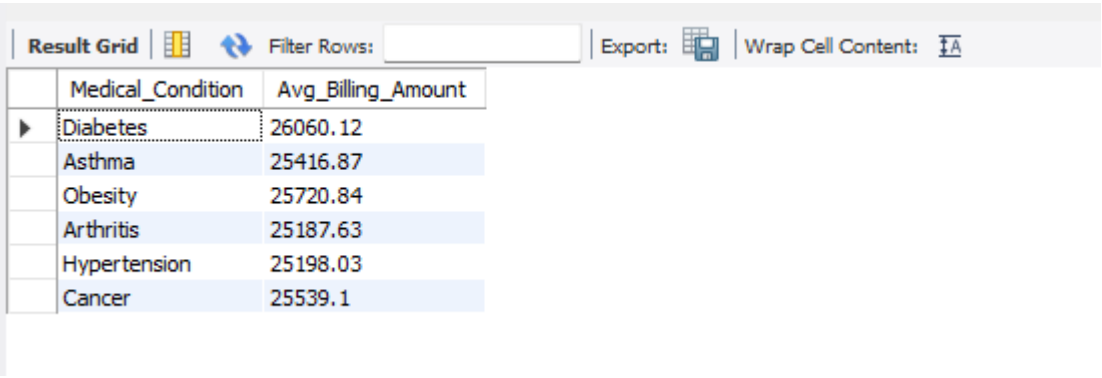
-- 11. Identifying Average Billing Amount by Medical Condition.

select Medical_Condition, ROUND(avg(Billing_Amount),2) AS Avg_Billing_Amount

from Hospital1

group by Medical_condition

-- Findings : It offers insights into the typical costs associated with various medical conditions. This information can be valuable for analyzing the financial impact of different health issues, identifying expensive conditions, or assisting in resource allocation within healthcare facilities.


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Medical_Condition	Avg_Billing_Amount
Diabetes	26060.12
Asthma	25416.87
Obesity	25720.84
Arthritis	25187.63
Hypertension	25198.03
Cancer	25539.1

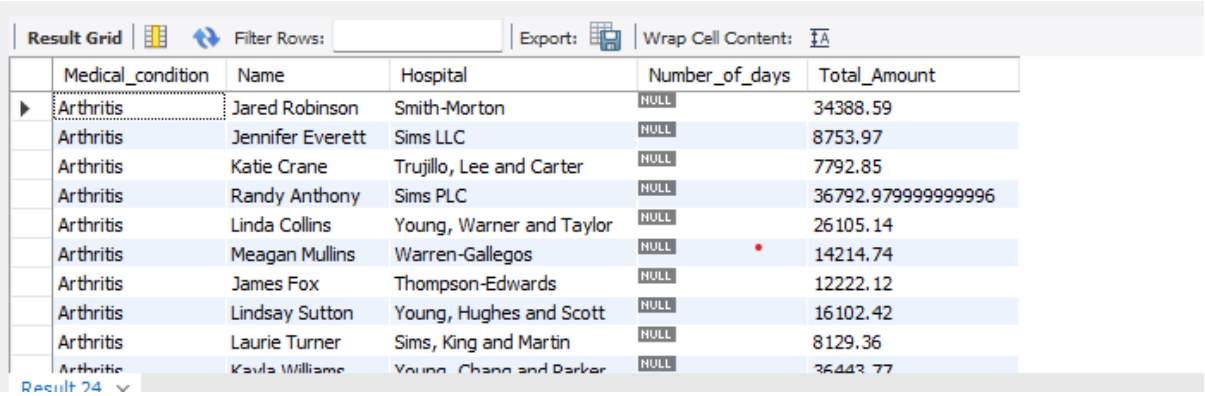
-- 12. Finding Billing Amount of patients admitted and number of days spent in respective hospital.

Select Medical_condition, Name, Hospital, DATEDIFF(discharge_date,Date_of_Admission) as Number_of_days,

SUM(ROUND(Billing_Amount,2)) Over(partition by Hospital ORDER BY hospital DESC) AS Total_Amount

from Hospital1

Order by Medical_condition



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Medical_condition	Name	Hospital	Number_of_days	Total_Amount
Arthritis	Jared Robinson	Smith-Morton	NULL	34388.59
Arthritis	Jennifer Everett	Sims LLC	NULL	8753.97
Arthritis	Katie Crane	Trujillo, Lee and Carter	NULL	7792.85
Arthritis	Randy Anthony	Sims PLC	NULL	36792.979999999996
Arthritis	Linda Collins	Young, Warner and Taylor	NULL	26105.14
Arthritis	Meagan Mullins	Warren-Gallegos	NULL	14214.74
Arthritis	James Fox	Thompson-Edwards	NULL	12222.12
Arthritis	Lindsay Sutton	Young, Hughes and Scott	NULL	16102.42
Arthritis	Laurie Turner	Sims, King and Martin	NULL	8129.36
Arthritis	Kayla Williams	Young, Chang and Parker	NULL	36443.77

Result 24

-- 13. Finding Total number of days spent by patient in an hospital for given medical condition

select Name, Medical_condition, Round(Billing_Amount,2) as Billing_Amount, Hospital, DATEDIFF(discharge_date, Date_of_Admission) as Total_Hospitaized_days

from Hospital1;

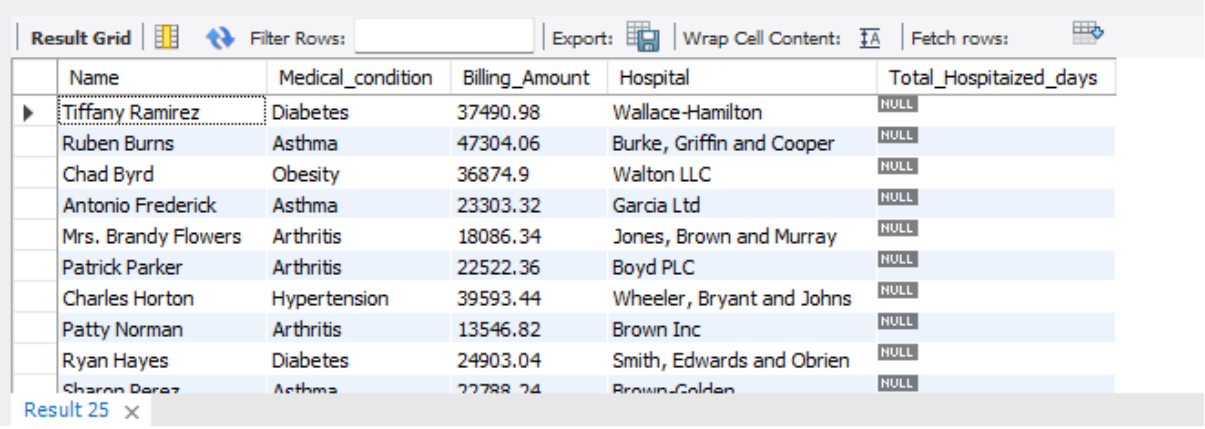
-- Findings : This query retrieves a dataset showing the names of patients, their respective medical conditions, billed amounts (rounded to two decimal places), the hospitals they visited, and the duration of their hospital stay in days. Insights gleaned include:

-- Individual Patient Details: It presents a comprehensive view of patients, their medical conditions, billed amounts, and hospitals involved, aiding in understanding the scope of medical services availed by patients.

-- Financial Overview: The rounded billing amounts provide an overview of the costs incurred by patients for their treatments, assisting in analyzing the financial aspect of healthcare services.

-- Hospital Performance: By knowing the length of hospital stays, an evaluation of the efficiency of hospitals in managing patient care and treatment duration is possible.

-- Potential Patterns: Patterns in medical conditions, billed amounts, and duration of hospitalization may emerge, offering insights into prevalent health issues and associated costs in the healthcare dataset.



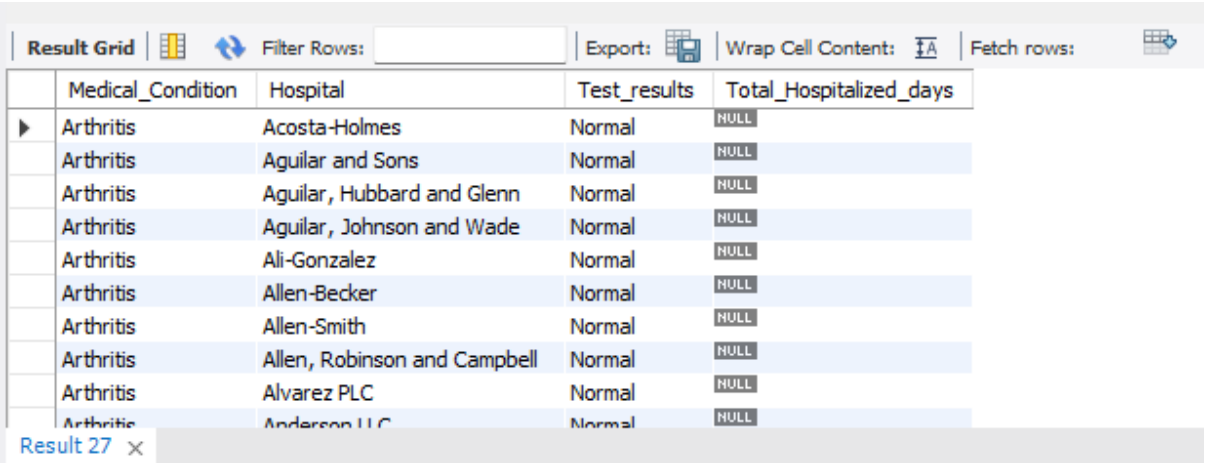
Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Name	Medical_condition	Billing_Amount	Hospital	Total_Hospitaized_days
Tiffany Ramirez	Diabetes	37490.98	Wallace-Hamilton	NULL
Ruben Burns	Asthma	47304.06	Burke, Griffin and Cooper	NULL
Chad Byrd	Obesity	36874.9	Walton LLC	NULL
Antonio Frederick	Asthma	23303.32	Garcia Ltd	NULL
Mrs. Brandy Flowers	Arthritis	18086.34	Jones, Brown and Murray	NULL
Patrick Parker	Arthritis	22522.36	Boyd PLC	NULL
Charles Horton	Hypertension	39593.44	Wheeler, Bryant and Johns	NULL
Patty Norman	Arthritis	13546.82	Brown Inc	NULL
Ryan Hayes	Diabetes	24903.04	Smith, Edwards and Obrien	NULL
Sharon Perez	Asthma	22788.24	Brown-Golden	NULL

Result 25

-- 14. Finding Hospitals which were successful in discharging patients after having test results as 'Normal' with count of days taken to get results to Normal

```
SELECT Medical_Condition, Hospital, Test_results, DATEDIFF(Discharge_Date, Date_of_Admission) as Total_Hospitalized_days

FROM Hospital1

WHERE Test_results LIKE 'Normal'

ORDER BY Medical_Condition, Hospital;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Medical_Condition	Hospital	Test_results	Total_Hospitalized_days
Arthritis	Acosta-Holmes	Normal	NULL
Arthritis	Aguilar and Sons	Normal	NULL
Arthritis	Aguilar, Hubbard and Glenn	Normal	NULL
Arthritis	Aguilar, Johnson and Wade	Normal	NULL
Arthritis	Ali-Gonzalez	Normal	NULL
Arthritis	Allen-Becker	Normal	NULL
Arthritis	Allen-Smith	Normal	NULL
Arthritis	Allen, Robinson and Campbell	Normal	NULL
Arthritis	Alvarez PLC	Normal	NULL
Arthritis	Anderson LLC	Normal	NULL

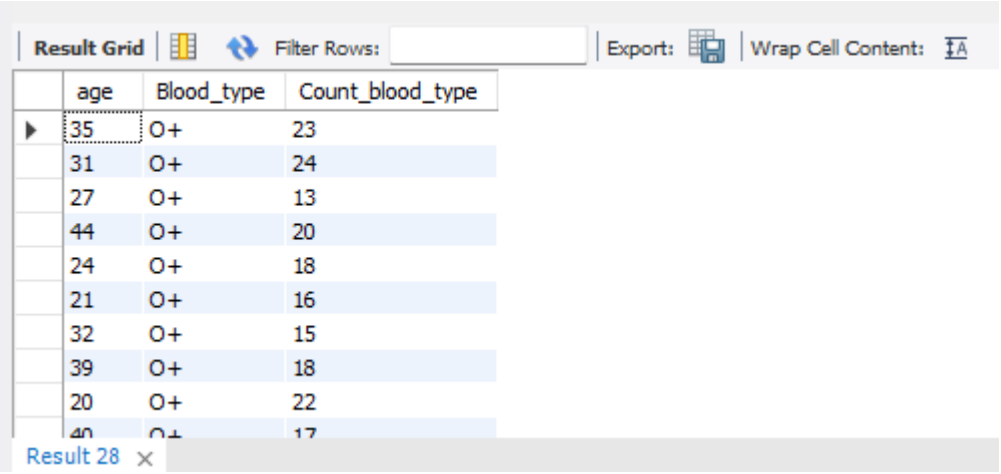
Result 27

-- 15. Calculate number of blood types of patients which lies betwwen age 20 to 45

```
select age, Blood_type, count(Blood_type) as Count_blood_type

from Hospital1

Where age Between 20 AND 45

group by 1, 2

order by Blood_type DESC;
```

-- Findings: This query filters healthcare data for individuals aged between 20 and 45, grouping them by their age and blood type. It then counts the occurrences of each blood

type within this age range. The output provides a breakdown of blood type distribution among individuals aged 20 to 45, revealing the prevalence of different blood types within this specific age bracket. The results may offer insights into any potential correlations between age groups and blood type occurrences within the dataset.


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

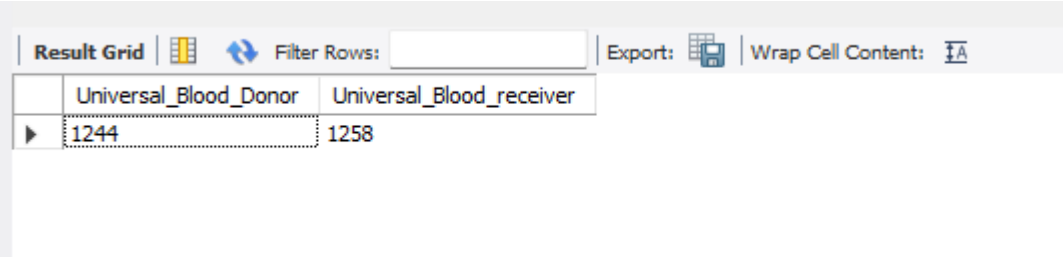
age	Blood_type	Count_blood_type
35	O+	23
31	O+	24
27	O+	13
44	O+	20
24	O+	18
21	O+	16
32	O+	15
39	O+	18
20	O+	22
40	O+	17

Result 28

-- 16. Find how many of patient are Universal Blood Donor and Universal Blood receiver

```
select distinct(select count(blood_Type) From Hospital1
Where Blood_type IN ('o-')) as Universal_Blood_Donor,
(select count(blood_type) from Hospital1
where Blood_type in ('AB+')) as Universal_Blood_receiver
```

-- Findings : This query extracts specific counts of individuals with particular blood types ('O-' and 'AB+') from the healthcare dataset. It compares the count of 'O-' blood type individuals (considered universal donors) against the count of 'AB+' blood type individuals (considered universal recipients). The result showcases the stark contrast in the prevalence of these two blood types within the dataset, highlighting the potential availability of universal donors compared to universal recipients.


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Universal_Blood_Donor	Universal_Blood_receiver
1244	1258

-- 17. Create a procedure to find Universal Blood Donor to an Universal Blood Reciever, with priority to same hospital and afterwards other hospitals

```
DELIMITER $$
```

```
CREATE PROCEDURE Blood_Matcher(IN Name_of_patient VARCHAR(200))
BEGIN
   SELECT
      D.Name AS Donor_name,
      D.Age AS Donor_Age,
      D.Blood_Type AS Donors_Blood_type,
      D.Hospital AS Donors_Hospital,
      R.Name AS Receiver_name,
      R.Age AS Receiver_Age,
      R.Blood_Type AS Receivers_Blood_type,
      R.Hospital AS Receivers_hospital
   FROM Hospital1 D
   INNER JOIN Hospital1 R
      ON (D.Blood_type = 'O-' AND R.Blood_type = 'AB+')
     AND ((D.Hospital = R.Hospital) OR (D.Hospital != R.Hospital))
   WHERE (R.Name REGEXP Name_of_patient)
    AND (D.Age BETWEEN 20 AND 40);
END $$

DELIMITER ;

CALL Blood_Matcher('Matthew Cruz');-- Enter the Name of patient as Argument
```

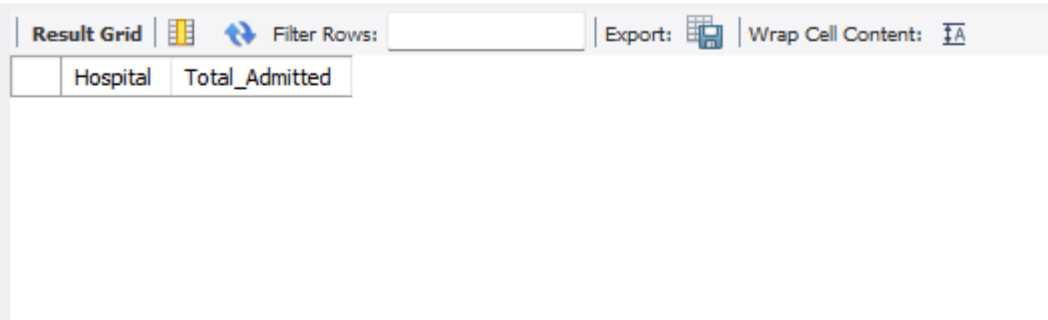
-- Findings : This stored procedure named `Blood_Matcher` is designed to identify potential donors and recipients based on specific blood types ('O-' and 'AB+') within a certain age range (20 to 40 years old). It retrieves the names, ages, blood types, and hospitals of potential donors and recipients from the Healthcare database. The condition checks for a match between the blood types and hospitals of donors and recipients, or if they are from

different hospitals. Additionally, it filters recipient names matching the input provided in the procedure call using regular expression. Overall, this procedure aims to find potential matches for blood donation between donors and recipients meeting specific criteria of blood type, age, and hospital affiliation or non-affiliation.

-- 18. Provide a list of hospitals along with the count of patients admitted in the year 2024 AND 2025?

```
select distinct hospital, count(*) as Total_admitted

from Hospital1;

WHERE YEAR(Date_of_Admission) IN (2024, 2025)

GROUP BY 1

ORDER BY Total_admitted Desc;
```

-- Findings : This query provides insights into the total admissions in different hospitals for the years 2024 and 2025. It retrieves the count of distinct admissions per hospital within the specified timeframe. The results are ordered in descending order based on the total number of admissions, highlighting hospitals with the highest influx of patients during these years. This data can aid in identifying healthcare facilities experiencing higher patient volumes across the specified period, aiding in resource allocation or further analysis of healthcare demand.

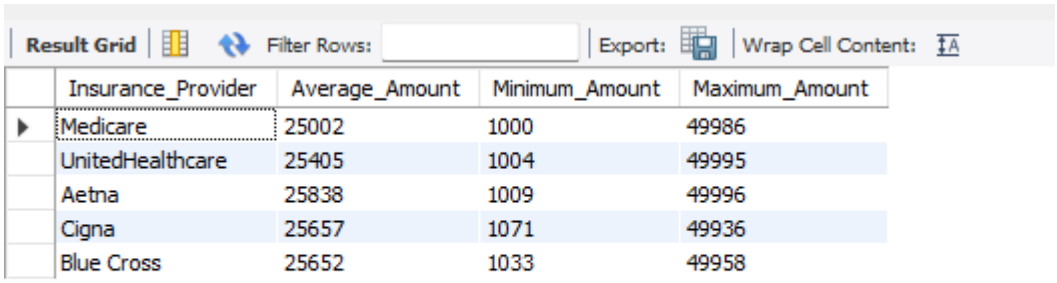


-- 19. Find the average, minimum and maximum billing amount for each insurance provider?

```
SELECT Insurance_Provider, ROUND(AVG(Billing_Amount),0) as Average_Amount,
ROUND(Min(Billing_Amount),0) as Minimum_Amount, ROUND(Max(Billing_Amount),0) as
Maximum_Amount

FROM Hospital1

GROUP BY 1;
```

-- Findings : This query provides insights into billing amounts across different insurance providers in the healthcare dataset. It calculates the average, minimum, and maximum billing amounts per insurance provider. By examining these metrics, we can understand the typical billing amount range associated with each insurance provider. This

information helps identify patterns in healthcare expenses linked to specific insurance companies, highlighting variations in billing practices or potential cost disparities among providers.



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Insurance_Provider	Average_Amount	Minimum_Amount	Maximum_Amount
Medicare	25002	1000	49986
UnitedHealthcare	25405	1004	49995
Aetna	25838	1009	49996
Cigna	25657	1071	49936
Blue Cross	25652	1033	49958

-- 20. Create a new column that categorizes patients as high, medium, or low risk based on their medical condition.

```
SELECT Name, Medical_Condition, Test_Results,
CASE
        WHEN Test_Results = 'Inconclusive' THEN 'Need More Checks / CANNOT be Discharged'
    WHEN Test_Results = 'Normal' THEN 'Can take discharge, But need to follow Prescribed medications timely'
    WHEN Test_Results =  'Abnormal' THEN 'Needs more attention and more tests'
    END AS 'Status', Hospital, Doctor
FROM Hospital1;
```

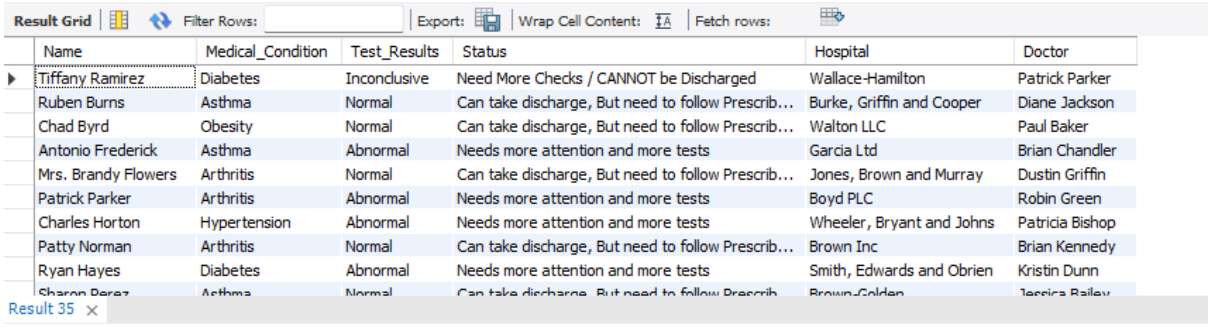
-- Findings : This query provides a summary of patients status based on their test results for various medical conditions.

--It categorizes patients into distinct statuses: those requiring additional checks and unable to be discharged due to inconclusive results,

-- individuals fit for discharge but needing strict adherence to prescribed medications for normal results, and those needing more

--attention and further tests for abnormal findings. It also displays associated details like the patient's name, hospital,

--and attending doctor, offering an overview of patient conditions, discharge possibilities, and necessary follow-up actions.



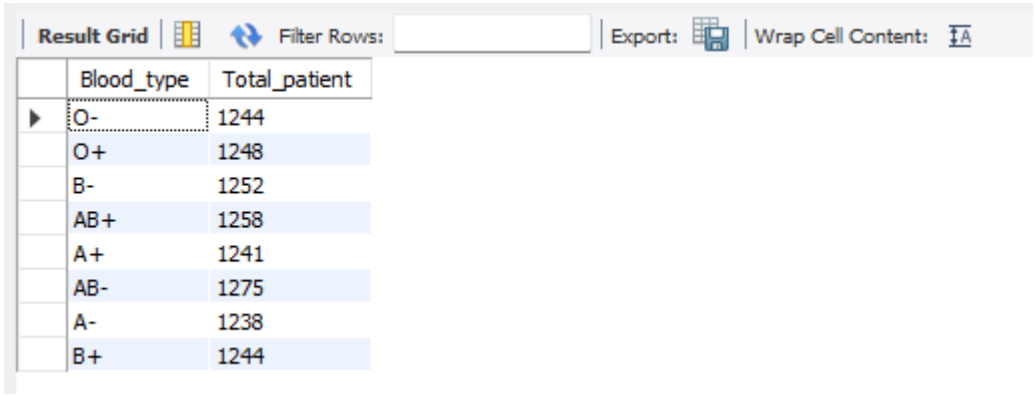
Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Name	Medical_Condition	Test_Results	Status	Hospital	Doctor
Tiffany Ramirez	Diabetes	Inconclusive	Need More Checks / CANNOT be Discharged	Wallace-Hamilton	Patrick Parker
Ruben Burns	Asthma	Normal	Can take discharge, But need to follow Prescrib...	Burke, Griffin and Cooper	Diane Jackson
Chad Byrd	Obesity	Normal	Can take discharge, But need to follow Prescrib...	Walton LLC	Paul Baker
Antonio Frederick	Asthma	Abnormal	Needs more attention and more tests	Garcia Ltd	Brian Chandler
Mrs. Brandy Flowers	Arthritis	Normal	Can take discharge, But need to follow Prescrib...	Jones, Brown and Murray	Dustin Griffin
Patrick Parker	Arthritis	Abnormal	Needs more attention and more tests	Boyd PLC	Robin Green
Charles Horton	Hypertension	Abnormal	Needs more attention and more tests	Wheeler, Bryant and Johns	Patricia Bishop
Patty Norman	Arthritis	Normal	Can take discharge, But need to follow Prescrib...	Brown Inc	Brian Kennedy
Ryan Hayes	Diabetes	Abnormal	Needs more attention and more tests	Smith, Edwards and Obrien	Kristin Dunn
Sharon Perez	Asthma	Normal	Can take discharge, But need to follow Prescrib...	Brown-Golden	Jessica Bailey

Result 35

--21. Find the total patient of each blood group

```
select Blood_type, Count(Blood_type) as Total_patient
from hospital1
group by Blood_type
```

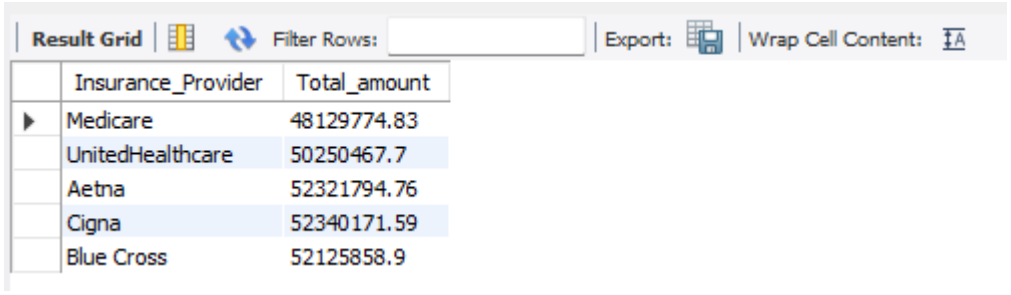


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Blood_type	Total_patient
O-	1244
O+	1248
B-	1252
AB+	1258
A+	1241
AB-	1275
A-	1238
B+	1244

--22. Total amount by the insurance provider

```
select Insurance_Provider, round(sum(Billing_Amount),2) as Total_amount
from hospital1
group by Insurance_Provider
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Insurance_Provider	Total_amount
Medicare	48129774.83
UnitedHealthcare	50250467.7
Aetna	52321794.76
Cigna	52340171.59
Blue Cross	52125858.9