HEALTHCARE DATA ANALYTICS PROJECT USING SQL

Insights from Patient Admissions, Billing, and Hospital Operations etc.

Tools used - MySQL (Database) and MySQL Workbench (Interface/Query Editor)

Project by: Manish Rajpurohit

INTRODUCTION

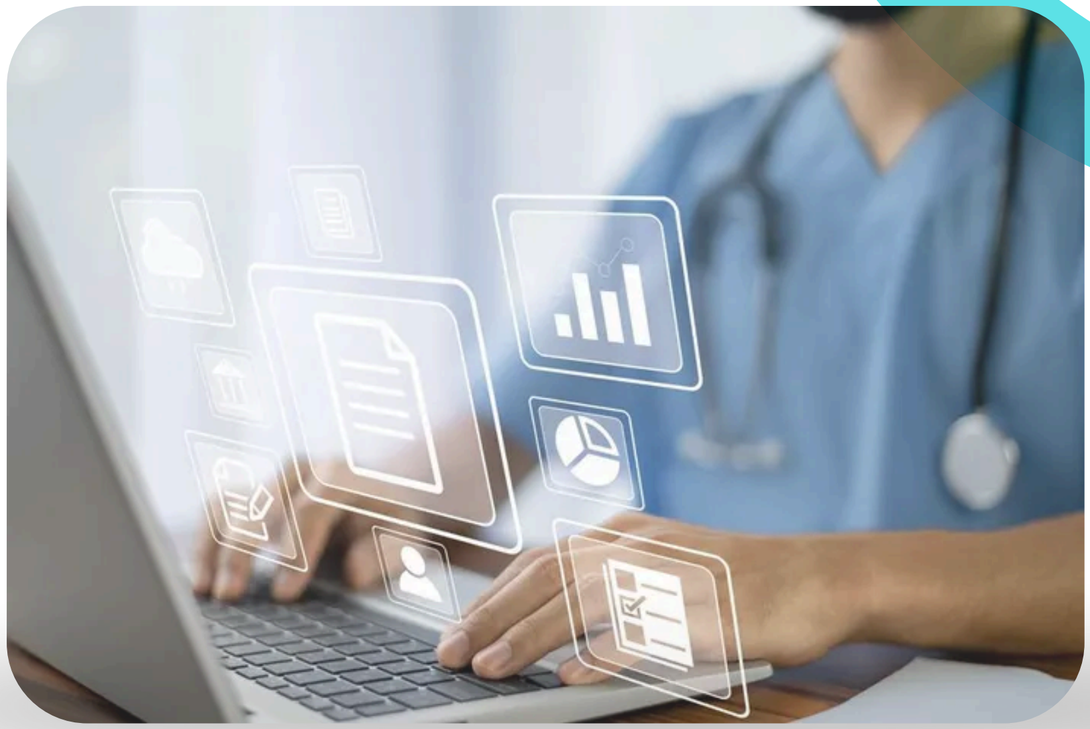
- This project focuses on analyzing a healthcare dataset to generate insights into patient admissions, medical conditions, billing patterns, and hospital operations.

- The dataset contains records of patients including demographics, medical conditions, medications, test results, admission and discharge details, doctors, hospitals, and billing information.

- Objective: To use SQL queries to extract meaningful insights that can help hospitals and healthcare organizations make informed decisions.

DATASET OVERVIEW

- Columns in the dataset:
  - Name – Patient name
  - Age – Patient age at admission
  - Gender – Male / Female
  - Blood Type – A+, O-, etc.
  - Medical Condition – Primary diagnosis
  - Date of Admission
  - Doctor – Attending doctor
  - Hospital – Name of the hospital
  - Insurance Provider – Aetna, Cigna, Blue Cross, etc.
  - Billing Amount – Cost billed to patient/insurance
  - Room Number
  - Admission Type – Emergency, Elective, Urgent
  - Discharge Date
  - Medication – Prescribed medications
  - Test Results – Normal, Abnormal, Inconclusive
- Original dataset had 55,000+ rows, but for this project it was reduced to 500 rows for easier loading and query execution in MySQL.
- Source: Kaggle – Hospital / Healthcare Dataset



Note:

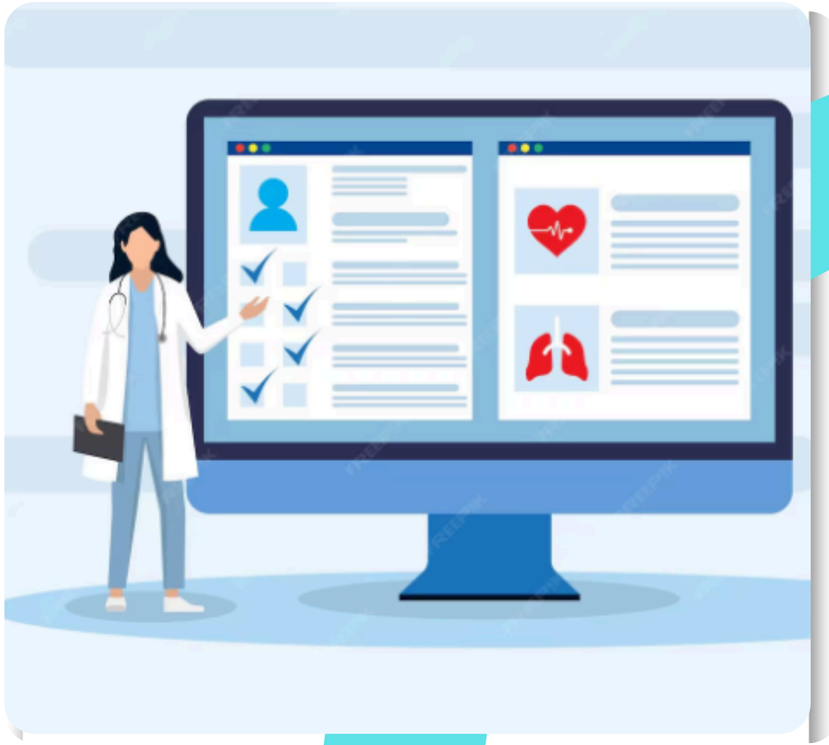
A primary key column “recordId” was created in MySQL to uniquely distinguish each patient record

OBJECTIVE / PURPOSE

- Analyze patient demographics: Gender, age, blood type distribution.
- Understand hospital operations: Admission types, length of stay, doctor workloads.
- Identify financial insights: Billing patterns, high-cost conditions, insurance contribution.
- Trend & condition analysis: Most common diseases, medication usage, test results patterns.
- Generate actionable insights: Help hospitals optimize operations, resource allocation, and patient care.

METHODOLOGY

- Loaded the dataset into MySQL environment.
- Performed basic data cleaning (check nulls, correct data types, format dates).
- Written SQL queries to answer each analytical question.
- Extracted insights using aggregations, grouping, filtering, ranking, and comparisons.
- Captured query outputs as screenshots for reporting in the project.
- Presented results in a structured PPT format (Question → Query → Result).



UNDERSTANDING THE QUERY SLIDES

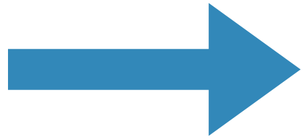
Each of the upcoming slides will present:

- Analytical Question(Total 30) – the insight we want to get or the problem we want to solve
- SQL Query – the exact SQL used to find the answer.
- Result Screenshot – the output from the query.

Each step showcases how SQL queries help uncover patterns, trends, and key insights from healthcare data.

1. What is the total number of patients admitted to the hospital?

```
SELECT
    COUNT(recordId) AS Total_Patients
FROM
    healthcare;
```



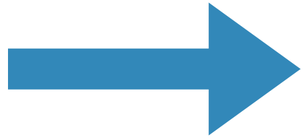
Result Grid

	Total_Patients
▶	500

---

2. How many patients are male vs female?

```
SELECT
    gender, COUNT(recordId) AS Total_Patients
FROM
    healthcare
GROUP BY gender;
```

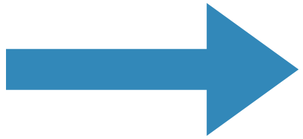


Result Grid | Filter Rows

	gender	Total_Patients
▶	Male	256
	Female	244

3. List the count of patients by admission type.

```
SELECT
    admissionType, COUNT(recordId) AS Total_Patients
FROM
    healthcare
GROUP BY admissionType;
```

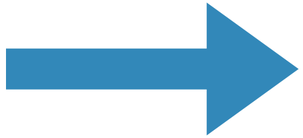


Result Grid | Filter Rows:

admissionType	Total_Patients
Urgent	159
Emergency	140
Elective	201

4. Find the top 4 most common medical conditions among admitted patients.

```
SELECT
    medicalCondition, COUNT(recordId) AS Total_Patients
FROM
    healthcare
GROUP BY medicalCondition
ORDER BY COUNT(recordId) DESC
LIMIT 4;
```



Result Grid | Filter Rows:

medicalCondition	Total_Patients
Diabetes	91
Cancer	85
Hypertension	85
Asthma	83

5. Which insurance provider has the maximum number of patients covered?

```
SELECT
    insuranceProvider, COUNT(recordId)
FROM
    healthcare
GROUP BY insuranceProvider
ORDER BY COUNT(recordId) DESC
LIMIT 1;
```

Result Grid | Filter Rows:

insuranceProvider	Total_Patients_Covered
UnitedHealthcare	105

6. What is the average age of patients admitted with each medical condition?

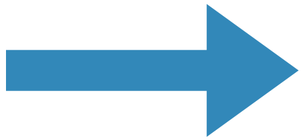
```
SELECT
    medicalCondition, ROUND(AVG(age), 2) AS
Average_Age
FROM
    healthcare
GROUP BY medicalCondition;
```

Result Grid | Filter Rows:

medicalCondition	Average_Age
Cancer	49.55
Obesity	50.50
Diabetes	49.47
Asthma	50.46
Hypertension	47.86
Arthritis	52.24

7. Calculate the average billing amount for each type of admission.

```
SELECT
    insuranceProvider, COUNT(recordId)
FROM
    healthcare
GROUP BY insuranceProvider
ORDER BY COUNT(recordId) DESC
LIMIT 1;
```

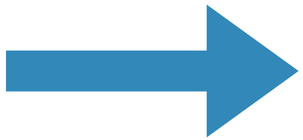


Result Grid | Filter Rows:

admissionType	Average_Bill
Urgent	23866.85
Emergency	23217.99
Elective	27186.43

8. Which hospital generated the highest total billing amount?

```
SELECT
    hospital, SUM(billingAmount) AS Total_Bill
FROM
    healthcare
GROUP BY hospital
ORDER BY Total_Bill DESC
LIMIT 1;
```

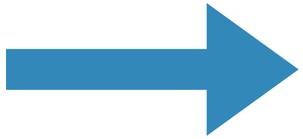


Result Grid | Filter Rows:

hospital	Total_Bill
Group Thompson	67696.34

9. Find the average length of stay for patients across different admission types.

```
SELECT
    admissionType,
    ROUND(AVG(DATEDIFF(dischargeDate,
dateOfAdmission)),2)
    AS Average_Stay_Days
FROM
    healthcare
GROUP BY admissionType;
```

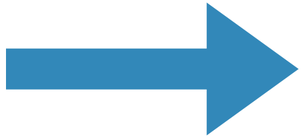


Result Grid | Filter Rows:

admissionType	Average_Stay_Days
Urgent	15.63
Emergency	14.68
Elective	16.91

10. Identify the doctor with the highest number of patients treated.

```
SELECT
    doctor, COUNT(recordId) AS
Total_Patients_Treated
FROM
    healthcare
GROUP BY doctor
ORDER BY Total_Patients_Treated DESC
LIMIT 2;
```

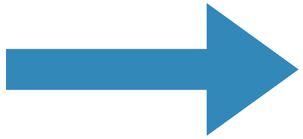


Result Grid | Filter Rows:

doctor	Total_Patients_Treated
John Smith	2
John Hansen	2

11. Which medication is prescribed most frequently, and for which condition is it most common?

```
SELECT
    medication,
    medicalCondition,
    COUNT(recordId) AS Prescription_Count
FROM
    healthcare
GROUP BY medication , medicalCondition
ORDER BY Prescription_Count DESC
LIMIT 1;
```

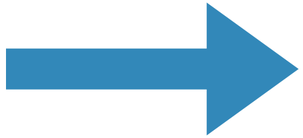


Result Grid | Filter Rows:

	medication	medicalCondition	Prescription_Count
▶	Ibuprofen	Asthma	27

12. Compare the average billing amount by gender.

```
SELECT
    gender, ROUND(AVG(billingAmount), 2) AS Average_Bill
FROM
    healthcare
GROUP BY gender;
```

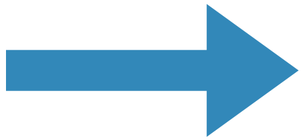


Result Grid | Filter Rows:

	gender	Average_Bill
▶	Male	25651.86
	Female	24356.34

13. Show the monthly patient admission trend (patients admitted per month).

```
SELECT
    DATE_FORMAT(dateOfAdmission, '%M %Y')
    AS Admission_Month,
    COUNT(recordId) AS Total_Patients
FROM
    healthcare
GROUP BY Admission_Month
ORDER BY MIN(dateOfAdmission);
```



Result Grid | Filter Rows:

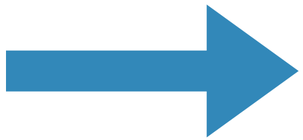
Admission_Month	Total_Patients
May 2019	4
June 2019	7
July 2019	9
August 2019	11
September 2019	7
October 2019	7

These are just few records; actual table is bigger

---

14. Which blood type patients are most commonly associated with abnormal test results?

```
SELECT
    bloodType, COUNT(*) Abnormal_Count
FROM
    healthcare
WHERE
    testResults = 'Abnormal'
GROUP BY bloodType
ORDER BY Abnormal_Count DESC
LIMIT 1;
```

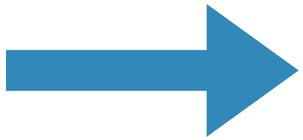


Result Grid | Filter Rows:

bloodType	Abnormal_Count
O+	33

15. Rank hospitals based on their average billing amount per patient.

```
SELECT
    hospital, ROUND(AVG(billingAmount),2)
    AS Average_Bill,
    RANK() OVER (ORDER BY AVG(billingAmount) DESC)
    AS `RANK`
FROM healthcare
GROUP BY hospital
ORDER BY `RANK`;
```



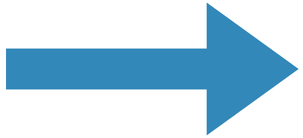
Result Grid | Filter Rows:

hospital	Average_Bill	RANK
Baker, Weber Patton and	51587.94	1
Foster-Thomas	50332.93	2
Robinson-Reyes	50254.2	3
James-Smith	50142.03	4
Sons Rich and	50119.22	5
PLC Young	49943.28	6

These are just few records;
actual table is bigger

16. Which insurance provider contributes the most in terms of total billing amount?

```
SELECT
    insuranceProvider,
    ROUND(SUM(billingAmount), 2) AS Total_Bill
FROM
    healthcare
GROUP BY insuranceProvider
ORDER BY Total_Bill
LIMIT 1;
```

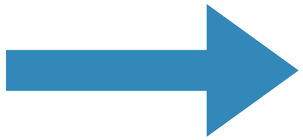


Result Grid | Filter Rows:

insuranceProvider	Total_Bill
Aetna	2253026.89

17. Find the most expensive medical condition.

```
SELECT
    medicalCondition,
    ROUND(AVG(billingAmount), 2) AS Average_Cost
FROM
    healthcare
GROUP BY medicalCondition
ORDER BY Average_Cost DESC
LIMIT 1;
```



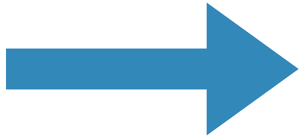
Result Grid | Filter Rows:

	medicalCondition	Average_Cost
▶	Asthma	27170.68

---

18. Identify the doctor–hospital pair that has generated the maximum revenue.

```
SELECT
    doctor, hospital, SUM(billingAmount) AS Revenue
FROM
    healthcare
GROUP BY doctor , hospital
ORDER BY Revenue DESC
LIMIT 1;
```

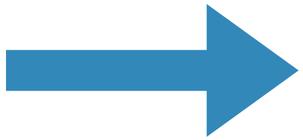


Result Grid | Filter Rows: | Expo

	doctor	hospital	Revenue
▶	David Gonzalez	Baker, Weber Patton and	51587.94

19. For each medical condition, calculate the average length of stay and see which one leads to the longest hospitalizations.

```
SELECT
    medicalCondition,
    ROUND(AVG(DATEDIFF(dischargeDate, dateOfAdmission)),
          2) AS Total_Stay_Days
FROM
    healthcare
GROUP BY medicalCondition
ORDER BY Total_Stay_Days DESC;
```



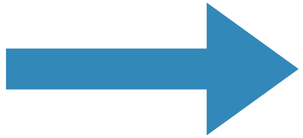
Result Grid | Filter Rows:

medicalCondition	Average_Stay_Days
Asthma	18.47
Diabetes	16.42
Hypertension	15.82
Cancer	15.13
Obesity	14.70
Arthritis	14.54

---

20. Compare the average billing difference between Emergency and Elective admissions.

```
SELECT
    admissionType,
    ROUND(AVG(billingAmount), 2) AS Average_Bill_Amount
FROM
    healthcare
WHERE
    admissionType IN ('Emergency' , 'Elective')
GROUP BY admissionType;
```



Result Grid | Filter Rows:

admissionType	Average_Bill_Amount
Emergency	23217.99
Elective	27186.43

21. Find the average length of stay per medical condition, and rank them from longest to shortest.

```
SELECT medicalCondition,
    ROUND(AVG(DATEDIFF(dischargeDate, dateOfAdmission)),2)
      AS Avg_Stay_Days,
    RANK() OVER(ORDER BY ROUND(AVG(DATEDIFF(dischargeDate,
dateOfAdmission)),2) DESC)
      AS `Rank`
FROM healthcare
GROUP BY medicalCondition;
```

Result Grid | Filter Rows:

medicalCondition	Avg_Stay_Days	Rank
Asthma	18.47	1
Diabetes	16.42	2
Hypertension	15.82	3
Cancer	15.13	4
Obesity	14.70	5
Arthritis	14.54	6

---

22. Compare the average billing amount for patients with Normal vs Abnormal test results across all hospitals.

```
SELECT
    hospital,
    testResults,
    ROUND(AVG(billingAmount), 2) AS Average_Bill
FROM
    healthcare
WHERE
    testResults IN ('Normal' , 'Abnormal')
GROUP BY hospital , testResults
ORDER BY hospital , testResults;
```

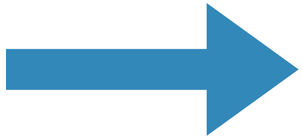
Result Grid | Filter Rows: | Export:

hospital	testResults	Average_Bill
Adams-Molina	Abnormal	21576.5
Alexander and Jensen Andrews,	Abnormal	3730
Alvarez-Cruz	Normal	42956.13
and Anderson Smith Sanchez,	Normal	40598.42
and Boyd Powell, Sims	Abnormal	49700.32
and Brown Oneal, Shah	Normal	23067.67
and Brown White, Jones	Abnormal	19065.06

These are just few records;
actual table is bigger

23. Calculate the yearly trend of admissions and identify the peak year.

```
SELECT
    YEAR(dateOfAdmission) AS `Year`,
    COUNT(recordId) AS Total_Patients
FROM
    healthcare
GROUP BY YEAR(dateOfAdmission)
ORDER BY Total_Patients DESC;
```



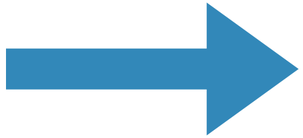
Result Grid | Filter Rows:

Year	Total_Patients
2020	117
2021	101
2022	93
2023	84
2019	62
2024	43

---

24. Find the most common medication prescribed for each medical condition.

```
SELECT medicalCondition, medication, Total_Prescription_Count
FROM(
 SELECT medicalCondition, medication, COUNT(recordId) AS
Total_Prescription_Count,
    RANK() OVER(PARTITION BY medicalCondition ORDER BY
COUNT(recordId) DESC) AS `Rank`
    FROM healthcare
    GROUP BY medicalCondition, medication
    ) AS Ranked
WHERE `Rank` = 1
ORDER BY medicalCondition;
```

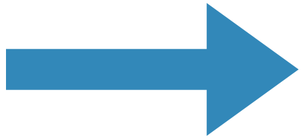


Result Grid | Filter Rows: | Export:

medicalCondition	medication	Total_Prescription_Count
Arthritis	Paracetamol	17
Arthritis	Penicillin	17
Arthritis	Ibuprofen	17
Asthma	Ibuprofen	27
Cancer	Paracetamol	24
Cancer	Ibuprofen	24
Diabetes	Ibuprofen	24
Hypertension	Paracetamol	22
Obesity	Aspirin	24

25. Determine which admission type generates the highest revenue per patient.

```
SELECT
    admissionType, ROUND(SUM(billingAmount), 2)
    AS Total_Revenue
FROM
    healthcare
GROUP BY admissionType
ORDER BY Total_Revenue DESC
LIMIT 1;
```



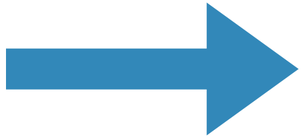
Result Grid | Filter Rows:

admissionType	Total_Revenue
Elective	5464473.39

---

26. Identify the top 3 insurance providers by number of patients, and calculate their average billing.

```
SELECT
    insuranceProvider,
    COUNT(recordId) AS Total_Patients_Covered,
    ROUND(AVG(billingAmount), 2) AS Average_Bill
FROM
    healthcare
GROUP BY insuranceProvider
ORDER BY Total_Patients_Covered DESC
LIMIT 3;
```

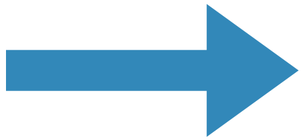


Result Grid | Filter Rows: | Export:

insuranceProvider	Total_Patients_Covered	Average_Bill
UnitedHealthcare	105	25224.57
Medicare	103	25092.06
Aetna	101	22307.2

27. For each hospital, find the gender-wise distribution of patients and their corresponding average billing.

```
SELECT
    hospital, gender, AVG(billingAmount) AS Average_Bill
FROM
    healthcare
GROUP BY hospital , gender
ORDER BY hospital , gender;
```



Result Grid | Filter Rows: | Export:

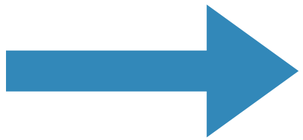
hospital	gender	Average_Bill
Adams-Molina	Female	21576.5
Alexander and Jensen Andrews,	Female	3730
Alvarado-Deleon	Female	27385.74
Alvarez-Cruz	Female	42956.13
and Anderson Smith Sanchez,	Female	40598.42
and Black, Henson Rhodes	Female	22480.97
and Boyd Powell, Sims	Female	49700.32
and Brown Oneal, Shah	Female	23067.67
and Brown White, Jones	Female	19065.06

These are just few records;
actual table is bigger

---

28. Find the percentage of patients over 60 years old admitted for each medical condition.

```
SELECT medicalCondition,
      ROUND(((SUM(CASE WHEN age < 60 THEN 1 ELSE 0 END)
      /
      COUNT(recordId)) * 100),2)
      AS `%Patients_Over60Age`
FROM healthcare
GROUP BY medicalCondition;
```

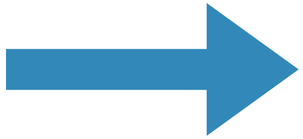


Result Grid | Filter Rows:

medicalCondition	%Patients_Over60Age
Cancer	67.06
Obesity	65.00
Diabetes	67.03
Asthma	63.86
Hypertension	67.06
Arthritis	57.89

29. Identify the doctors who treated more than 1 medical condition and calculate their total billing share.

```
SELECT
    doctor, SUM(billingAmount) AS Total_Bill
FROM
    healthcare
GROUP BY doctor
HAVING COUNT(DISTINCT (medicalCondition)) > 1;
```



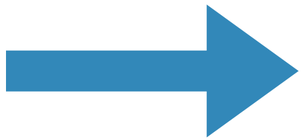
Result Grid | Filter Rows:

doctor	Total_Bill
John Hansen	65540.27
John Smith	36473.42

---

30. Segment patients into age groups (0–20, 21–40, 41–60, 61+) and calculate the average billing amount for each group to identify which age group contributes most to revenue. Show all age groups.

```
SELECT
    CASE
        WHEN age BETWEEN 0 AND 20 THEN '0-20'
        WHEN age BETWEEN 21 AND 40 THEN '21-40'
        WHEN age BETWEEN 41 AND 60 THEN '41-60'
        ELSE '61+'
    END AS Age_Group,
    ROUND(AVG(billingAmount), 2) AS Average_Bill
FROM
    healthcare
GROUP BY Age_Group
ORDER BY Average_Bill DESC;
```



Result Grid | Filter Rows

Age_Group	Average_Bill
0-20	29085.84
61+	25333.21
21-40	25314.42
41-60	23316.07

KEY INSIGHTS & FINDINGS

- **Gender** distribution is **fairly balanced** across patients.
- Patients above **60 years old** contribute significantly to hospital revenue.
- **Emergency admissions** are more frequent and associated with higher billing amounts.
- **Average length of stay** varies by admission type and medical condition, with chronic illnesses leading to longer hospitalizations.
- A few **doctors and hospitals** are responsible for a large share of patients and revenue.
- **Top medical conditions** such as Diabetes, Hypertension, Cancer, and Asthma dominate patient admissions.
- **Abnormal test results** are more common in certain blood types and medical conditions.
- Some **medications** (e.g., Paracetamol, Aspirin) are prescribed frequently for common illnesses.
- **Billing amounts** vary widely across admission types and medical conditions.
- **Insurance providers** differ in their role – some cover more patients, while others handle higher-cost cases.
- Certain hospitals stand out for having a **higher average billing per patient**, not just higher patient volumes.

CONCLUSION

- This project demonstrated how SQL can be applied to healthcare data to uncover meaningful patterns and insights.
- Through structured queries, we analyzed patient demographics, admissions, medical conditions, billing trends, and insurance coverage.
- The analysis highlighted operational and financial insights that hospitals and insurers can use for better decision-making.
- Overall, this project reinforced the value of SQL as a powerful tool for data analytics and business intelligence.

THANK YOU