SQL-Based Healthcare Data Analysis

Author: Amit Singh
Email: ikarannegi4919@gmail.com
LinkedIn: www.linkedin.com/in/amit-singh2003
Dataset Source: Kaggle - Healthcare Dataset
Dataset Size: 55,500 records

Project Overview

This project performs SQL-based analysis on a large healthcare dataset. The dataset includes patient-level admission records across multiple hospitals. The purpose of this analysis is to extract key business insights that can aid hospital management in decision-making.

Tools Used:

- **Excel:** Data Cleaning
- **ERD:** Designed from raw structure for clarity in MySQL Workbench
- **MySQL:** Querying & Analysis

Excel Data Cleaning Process:-

Before importing the dataset into MySQL, we performed extensive cleaning and preparation using Excel:

- **Removed titles** ("Mr.", "Mrs.", "Dr.") from patient and doctor names.
- **Trimmed spaces** and removed duplicate rows for hospitals, doctors, and patients etc.
- **Generated unique IDs** for:
  - Patients (e.g., P0001, P0002)
  - Doctors (e.g., D0001, D0002)
  - Hospitals (e.g., H0001, H0002)
  - Records (e.g., R0001, R0002)
- Ensured **date formats** were standardized (e.g., YYYY-MM-DD).

  After cleaning, the structured data was saved as a CSV and imported into MySQL Workbench.

Database Design & ER Diagram

After data cleaning, the dataset was structured into a relational database.

The raw flat data was broken into logical entities:

-Patients Table: Contain details about Patients like Patient_id, Name , Gender etc.

- Doctor Table: Contain details about Doctor like Doctor_id, Doctor_Name etc.

- Hospitals Table: Contains Hospitals details like Hospital_id, Hospital_Name etc .

-Insurance Table: Contains Insurance details like Insurance_id, Insurance_Provider etc .

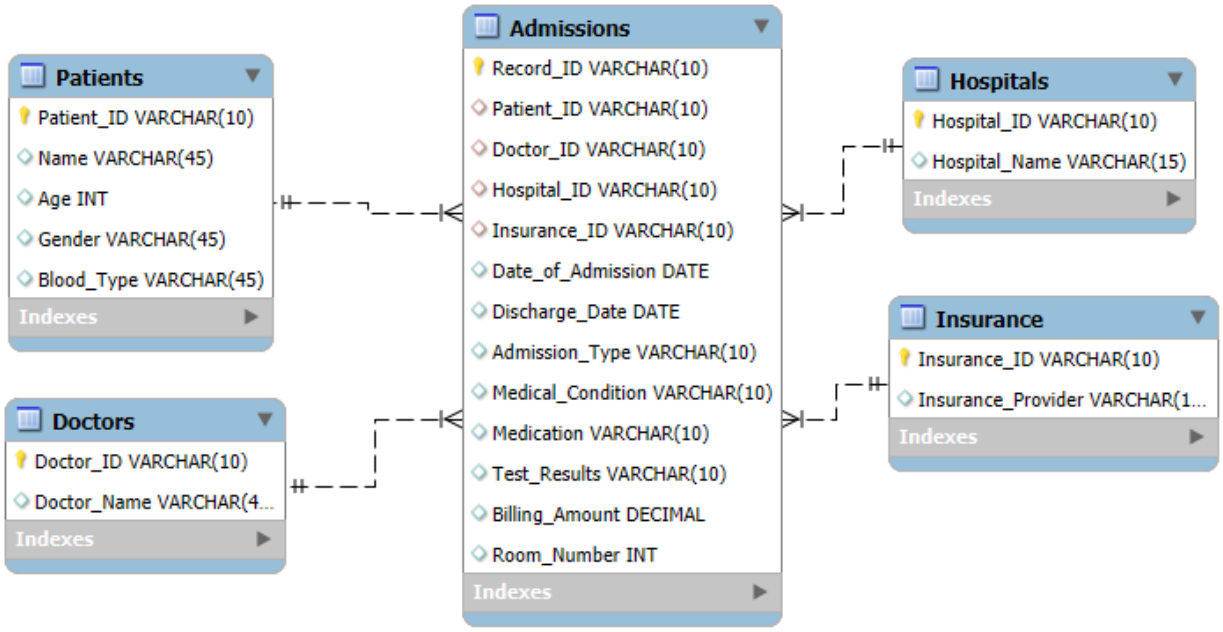
- Admissions table: Contains Healthcare Details and links other tables.

Primary & Foreign Keys :

- Patients_id → primary key in Patients, foreign key in Admissions.
- Doctor_id → primary key in Doctors, foreign key in Admissions.
- Hospital_id → primary key in Hospitals, foreign key in Admissions.
- Insurance_id → primary key in Insurance, foreign key in Admissions.

ER – DIAGRAM :

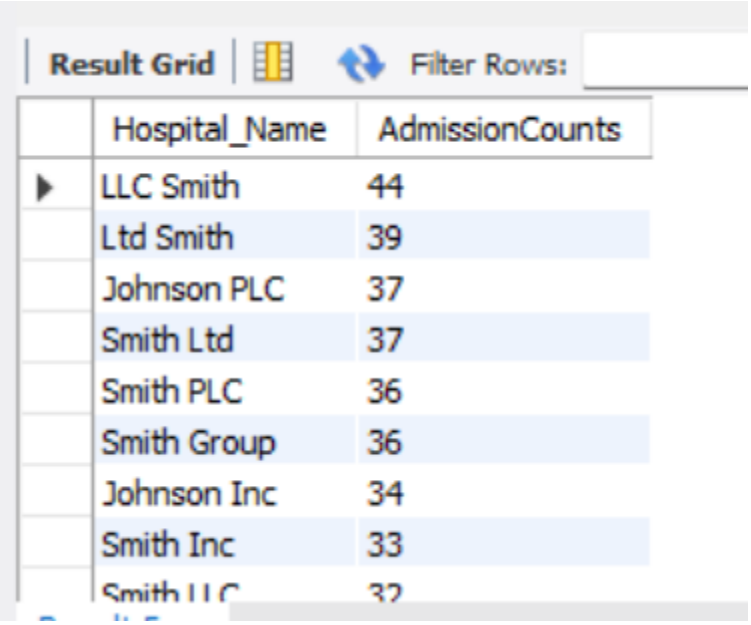
The Entity Relationship Diagram (ERD) was created using MySQL Workbench. It visually represents the relationships between tables.



Business Problems and SQL Insights

1. **Most Visited Hospitals**

```sql
SELECT h.Hospital_Name, COUNT(a.Patient_ID) AS AdmissionCounts
FROM admissions a
JOIN hospitals h ON a.Hospital_ID = h.Hospital_ID
GROUP BY Hospital_Name
ORDER BY AdmissionCounts DESC;
```



Hospital_Name	AdmissionCounts
LLC Smith	44
Ltd Smith	39
Johnson PLC	37
Smith Ltd	37
Smith PLC	36
Smith Group	36
Johnson Inc	34
Smith Inc	33
Smith LLC	32

Insight:

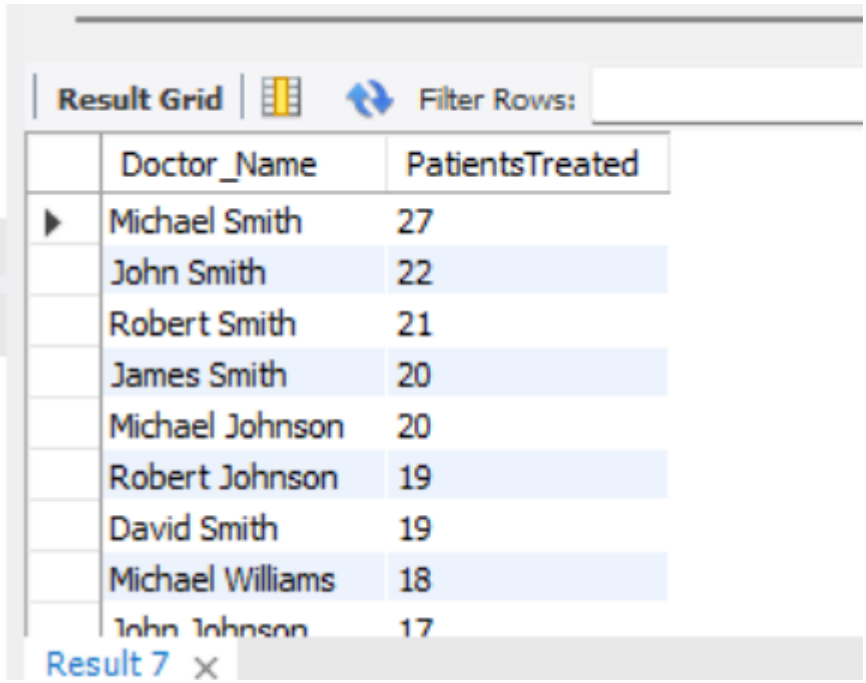
- Identifies hospitals with the highest patient traffic.
- Useful for resource allocation, funding, and infrastructure planning.

---

2. **Doctors with Highest Patient Count**

```sql
SELECT d.Doctor_Name, COUNT(a.Patient_ID) AS PatientsTreated
FROM admissions a
JOIN doctors d ON a.Doctor_ID = d.Doctor_ID
```

```
GROUP BY Doctor_Name
ORDER BY PatientsTreated DESC;
```



Result Grid | Filter Rows:

Doctor_Name	PatientsTreated
Michael Smith	27
John Smith	22
Robert Smith	21
James Smith	20
Michael Johnson	20
Robert Johnson	19
David Smith	19
Michael Williams	18
John Johnson	17

Result 7

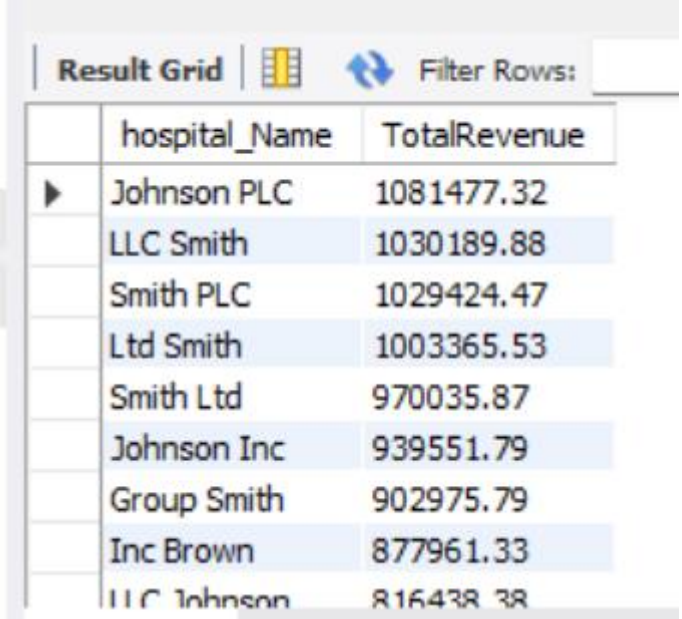
Insight:

- Reveals the top-performing doctors by patient volume.
- Can help in recognizing overburdened or popular medical professionals.

---

3. **Total Revenue by Hospital**

```
SELECT h.Hospital_Name, SUM(Billing_Amount) AS TotalRevenue
FROM admissions a
JOIN hospitals h ON a.Hospital_ID = h.Hospital_ID
GROUP BY Hospital_Name
ORDER BY TotalRevenue DESC;
```



Result Grid | Filter Rows:

hospital_Name	TotalRevenue
Johnson PLC	1081477.32
LLC Smith	1030189.88
Smith PLC	1029424.47
Ltd Smith	1003365.53
Smith Ltd	970035.87
Johnson Inc	939551.79
Group Smith	902975.79
Inc Brown	877961.33
LLC Johnson	816438.38

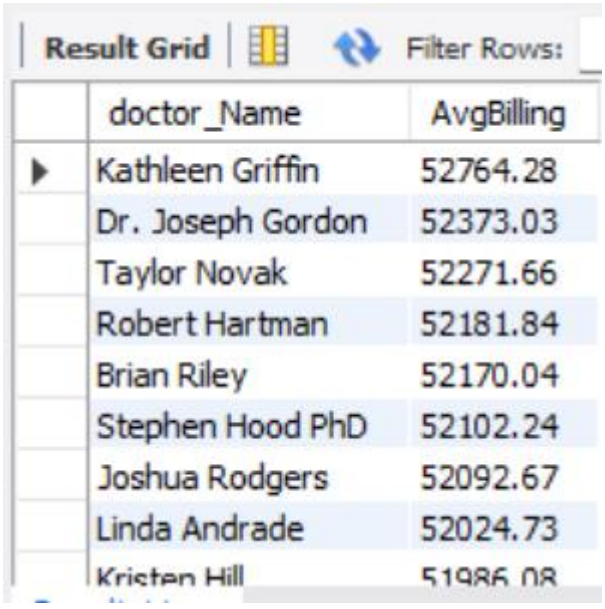
Insight:

- Shows which hospitals generate the most billing revenue.

---

4. **Average Billing per Doctor**

```
SELECT d.Doctor_Name, ROUND(AVG(Billing_Amount), 2) AS AvgBilling
FROM admissions a
JOIN doctors d ON a.Doctor_ID = d.Doctor_ID
GROUP BY Doctor_Name
ORDER BY AvgBilling DESC;
```



Result Grid | Filter Rows:

doctor_Name	AvgBilling
Kathleen Griffin	52764.28
Dr. Joseph Gordon	52373.03
Taylor Novak	52271.66
Robert Hartman	52181.84
Brian Riley	52170.04
Stephen Hood PhD	52102.24
Joshua Rodgers	52092.67
Linda Andrade	52024.73
Kristen Hill	51986.08

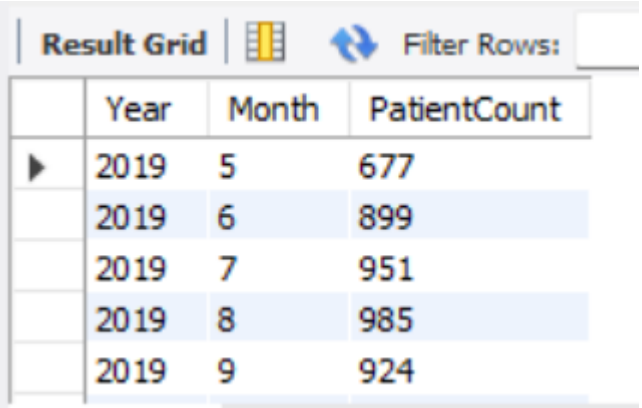
Insight:

- Highlights doctors associated with the highest treatment costs.

---

5. **Monthly Admission Trend (with Year)**

```
SELECT
  YEAR(Date_of_Admission) AS Year,
  MONTH(Date_of_Admission) AS Month,
  COUNT(Patient_ID) AS PatientCount
FROM admissions
GROUP BY YEAR(Date_of_Admission), MONTH(Date_of_Admission)
ORDER BY Year, Month;
```



Year	Month	PatientCount
2019	5	677
2019	6	899
2019	7	951
2019	8	985
2019	9	924

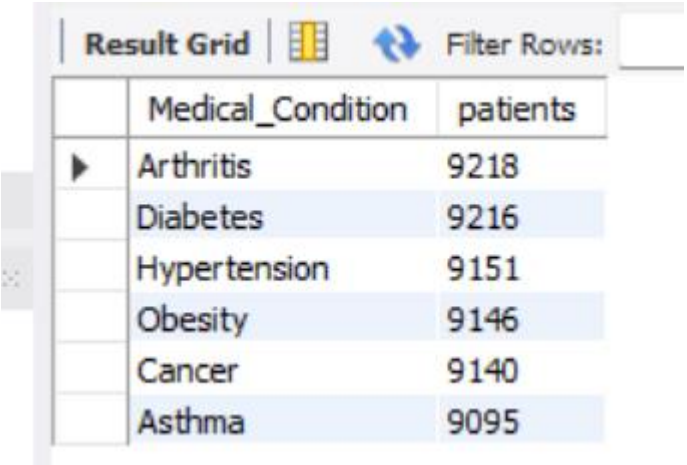
Insight:

- Tracks seasonality in patient visits.

---

6. **Most Common Medical Conditions**

```
SELECT Medical_Condition, COUNT(Patient_ID) AS Patients
FROM admissions
GROUP BY Medical_Condition
ORDER BY Patients DESC;
```



Medical_Condition	patients
Arthritis	9218
Diabetes	9216
Hypertension	9151
Obesity	9146
Cancer	9140
Asthma	9095

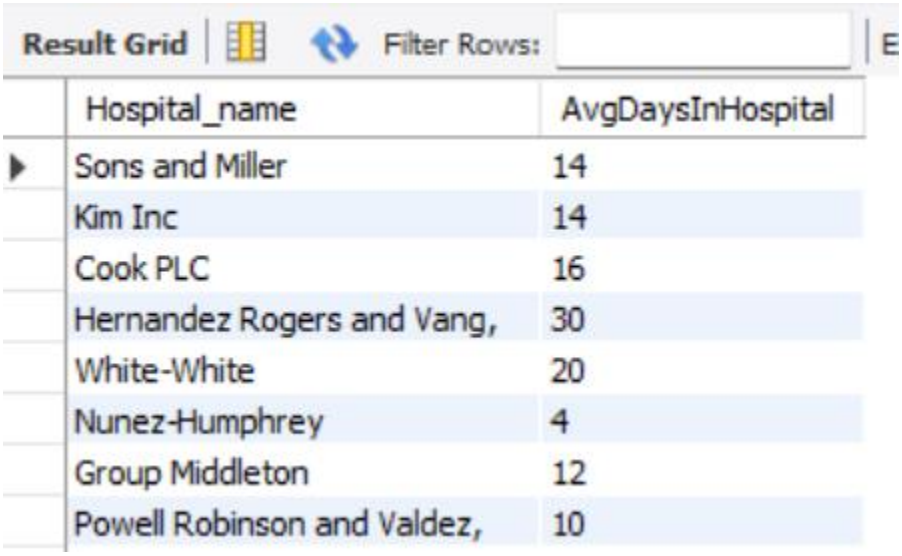
Insight:

- Identifies top health concerns among the population.
- Useful for public health campaigns and department staffing.

---

7. **Average Stay Duration per Hospital**

```
SELECT h.Hospital_Name, CEIL(AVG(DATEDIFF(Discharge_Date,
Date_of_Admission))) AS AvgDaysInHospital
FROM admissions a
JOIN hospitals h ON a.Hospital_ID = h.Hospital_ID
GROUP BY Hospital_Name;
```



Result Grid | Filter Rows:

Hospital_name	AvgDaysInHospital
Sons and Miller	14
Kim Inc	14
Cook PLC	16
Hernandez Rogers and Vang,	30
White-White	20
Nunez-Humphrey	4
Group Middleton	12
Powell Robinson and Valdez,	10

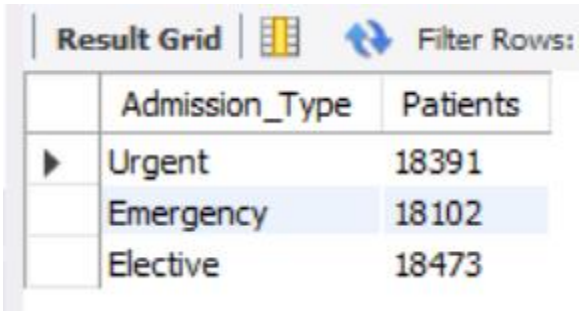
Insight:

- Measures hospital efficiency and patient recovery time.

---

8. **Most Common Admission Types**

```
SELECT Admission_Type, COUNT(Patient_ID) AS Patients
FROM admissions
GROUP BY Admission_Type;
```



Result Grid | Filter Rows:

Admission_Type	Patients
Urgent	18391
Emergency	18102
Elective	18473

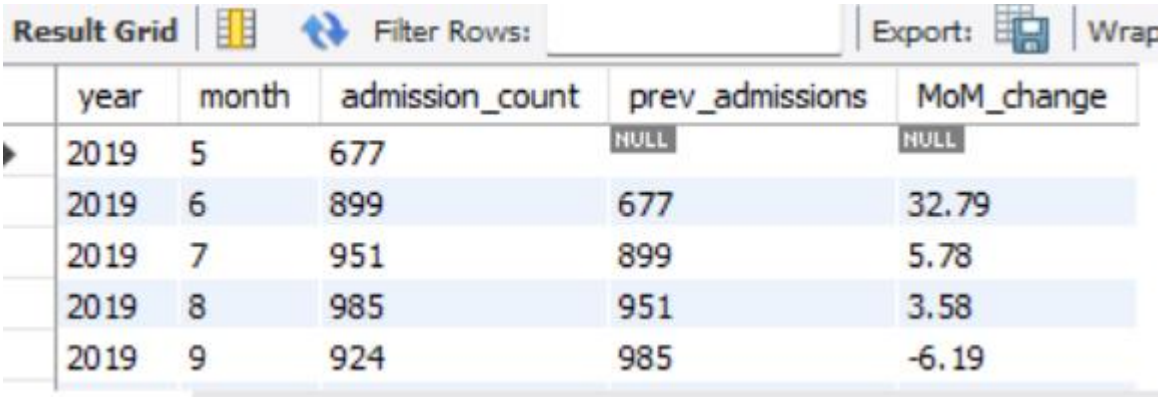
Insight:

- Segregates patient entries by type: Emergency, Urgent, Elective.

---

9. Month-over-Month Growth in Admissions

```
SELECT *, ROUND((admission_count - prev_admissions)/prev_admissions * 100, 2)
AS MoM_Change
FROM (
  SELECT
    YEAR(Date_of_Admission) AS Year,
    MONTH(Date_of_Admission) AS Month,
    COUNT(*) AS admission_count,
    LAG(COUNT(*)) OVER (ORDER BY YEAR(Date_of_Admission),
MONTH(Date_of_Admission)) AS prev_admissions
  FROM admissions
  GROUP BY YEAR(Date_of_Admission), MONTH(Date_of_Admission)
) t;
```



year	month	admission_count	prev_admissions	MoM_change
2019	5	677	NULL	NULL
2019	6	899	677	32.79
2019	7	951	899	5.78
2019	8	985	951	3.58
2019	9	924	985	-6.19

Insight:

- Identifies growth trends in admissions month over month.
- Accounts for different years, ensuring accurate comparisons.

---

10. Top 3 Medical Conditions Per Hospital

```
SELECT * FROM (
  SELECT h.Hospital_Name, Medical_Condition, COUNT(*) AS Condition_Count,
         ROW_NUMBER() OVER(PARTITION BY h.Hospital_Name ORDER BY COUNT(*)
DESC) AS Rank
  FROM admissions a
  JOIN hospitals h ON a.Hospital_ID = h.Hospital_ID
  GROUP BY h.Hospital_Name, Medical_Condition
) t
WHERE Rank <= 3;
```

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

Hospital_Name	Medical_Condition	condition_count	ranks
Acosta Group	Arthritis	3	1
Acosta Inc	Diabetes	1	1
Acosta Inc	Obesity	1	2
Acosta LLC	Diabetes	1	1
Acosta Ltd	Hypertension	1	1

Insight:

- Gives a breakdown of frequent diseases per hospital.

---

11. Hospitals with Above Average Revenue

```
WITH totalsales AS (
  SELECT h.Hospital_Name, SUM(a.Billing_Amount) AS TotalSales
  FROM admissions a
  JOIN hospitals h ON a.Hospital_ID = h.Hospital_ID
  GROUP BY h.Hospital_Name
),
avg_sales AS (
  SELECT AVG(TotalSales) AS avg_sales FROM totalsales
)
SELECT t.*
FROM totalsales t
JOIN avg_sales a ON t.TotalSales > a.avg_sales;
```

Result Grid | Filter Rows:

Hospital_Name	TotalSales
Sons and Miller	143444.93
Kim Inc	145680.00
Cook PLC	103509.11
Hernandez Rogers and Vang,	37909.78
Nunez-Humphrey	48145.11

Insight:

- Highlights hospitals whose revenue exceeds the average.

---

12. **Hospital with Fastest Discharge Rate**

```
SELECT h.Hospital_Name, ROUND(AVG(DATEDIFF(a.Discharge_Date,
a.Date_of_Admission)), 2) AS AvgStay
FROM admissions a
JOIN hospitals h ON a.Hospital_ID = h.Hospital_ID
GROUP BY h.Hospital_Name
ORDER BY AvgStay ASC
LIMIT 1;
```

Result Grid | Filter Rows:

Hospital_Name	Avg_Stay
Allen Watson and Robinson,	1.00
Davenport-Simpson	1.00
Pena-Barr	1.00
Davidson-Smith	1.00
Williams, and Pittman Alexander	1.00

Insight:

- Identifies the hospital with the most efficient discharge rate.
- Can indicate faster recovery.

---

13. **Month-over-Month Revenue Growth**

```
SELECT *, ROUND((revenue - prev_revenue)/prev_revenue * 100, 2) AS
RevenueGrowth
FROM (
  SELECT
    YEAR(Date_of_Admission) AS Year,
    MONTH(Date_of_Admission) AS Month,
    SUM(Billing_Amount) AS revenue,
    LAG(SUM(Billing_Amount)) OVER (ORDER BY YEAR(Date_of_Admission),
MONTH(Date_of_Admission)) AS prev_revenue
  FROM admissions
  GROUP BY YEAR(Date_of_Admission), MONTH(Date_of_Admission)
) t;
```

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

Month	Year	CurrentRevenue	PreviousRevenue	MoM_Revenue_Growth
5	2019	17406252.58	NULL	NULL
6	2019	23614582.03	17406252.58	35.67
7	2019	24632828.25	23614582.03	4.31
8	2019	25000186.37	24632828.25	1.49
9	2019	23021967.16	25000186.37	-7.91

Insight:

- Tracks monthly changes in total revenue.
- Ensures accuracy by separating months across years.

---

14. **Revenue per Medical Condition**

```
SELECT Medical_Condition, ROUND(SUM(Billing_Amount), 2) AS TotalRevenue
FROM admissions
GROUP BY Medical_Condition
ORDER BY TotalRevenue DESC;
```

Result Grid | Filter Rows:

Medical_Condition	totalamount
Diabetes	236502347.78
Obesity	236027423.08
Arthritis	235179312.38
Hypertension	233395396.39
Asthma	233156145.12
Cancer	229914237.50

Insight:

- Reveals which conditions generate the most revenue.
- Helps identify high-cost diseases.

Conclusion

This analysis provided data-backed insights for healthcare operations, including:

- Hospital traffic and revenue trends
- Doctor performance

- Disease prevalence
- Admission trends and types

These queries are essential for administrators, policymakers, and healthcare strategists aiming to improve operational efficiency and patient care.

For further collaboration or academic reference, feel free to reach out to me at **ikarannegi4919@gmail.com** or connect via **LinkedIn**.