```
-- ----------- HealthCare Analytics -----------------
-- --------------------------------------------------
/* Business Problem Statement
PROBLEM: A healthcare organization wants to analyze patient data to:
    1. Understand patient demographics and medical conditions
    2. Optimize resource allocation (rooms, doctors)
    3. Analyze financial aspects (billing, insurance)
    4. Improve patient care and operational efficiency

Key Business Questions
    1. Demographic Analysis
        What is the age and gender distribution of patients?
        How are blood types distributed among patients?

    2. Medical Analysis
        What are the most common medical conditions?
        Which medications are most frequently prescribed?
        How do test results correlate with medical conditions?

    3. Operational Analysis
        What are the most common admission types?
        Which hospitals and doctors handle the most patients?
        How long do patients typically stay in the hospital?

    4. Financial Analysis
        What is the average billing amount by medical condition?
        Which insurance providers cover the most patients?
        How does billing amount correlate with length of stay?
*/

SELECT * FROM healthcare_data;


-- ------------ Demographic Analysis -------------------
```

```sql
-- Age distribution
SELECT age_group, COUNT(age_group) as PatientCount
FROM
    (SELECT age,
            CASE
                WHEN age <=25 THEN '18-25'
                WHEN age <=35 THEN '26-35'
                WHEN age <=45 THEN '36-45'
                WHEN age <=55 THEN '46-55'
                WHEN age <=65 THEN '56-65'
                WHEN age <=75 THEN '66-75'
                ELSE 'Above 75'
            END AS age_group
    FROM healthcare_data) AS age_group
GROUP BY age_group
ORDER BY PatientCount DESC

-- Gender distribution
SELECT gender, COUNT(gender) as PatientCount
FROM healthcare_data
GROUP BY gender

-- Age and Gender Distribution
SELECT
    gender,
    COUNT(gender) as patient_count,
    ROUND(AVG(age),2) as avg_age,
    MIN(age) as min_age,
    MAX(age) as max_age
FROM healthcare_data
GROUP BY gender
ORDER BY patient_count DESC;

-- Blood types distribution among patients?
SELECT Blood_type, COUNT(blood_type) AS PatientCount
```

```sql
FROM healthcare_data
GROUP BY blood_type
ORDER BY PatientCount DESC


-- ------------------------------------------------
-- ------------ Medical Analysis --------------------
-- Most Common Medical Conditions
SELECT
    medical_condition,
    COUNT(medical_condition) AS cases,
    CONCAT(
        LEFT(
            ROUND(
                COUNT(medical_condition) * 100.0 / (SELECT COUNT(*) FROM healthcare_data),
            2),
        5),
    '%') AS percCount
FROM healthcare_data
GROUP BY medical_condition
ORDER BY percCount DESC;


-- Which medications are most frequently prescribed?
SELECT
    medication,
    COUNT(medication) AS cases,
    CONCAT(
        LEFT(
            ROUND(
                COUNT(medication) * 100.0 / (SELECT COUNT(*) FROM healthcare_data),
            2),
        5),
    '%') AS percCount
FROM healthcare_data
```

```sql
GROUP BY medication
ORDER BY percCount DESC;


-- How do test results correlate with medical conditions?
WITH corr AS
    (SELECT
        medical_condition,
        test_results,
        COUNT(*) AS cases,
        CONCAT(
            LEFT(
                ROUND(COUNT(*) * 100.0 / SUM(COUNT(*)) OVER (PARTITION BY medical_condition), 2),
            5),
        '%') as percentage
    FROM healthcare_data
    GROUP BY medical_condition, test_results)
SELECT * FROM corr
ORDER BY medical_condition, test_results, cases DESC;


-- Test Result Variation by Age Group
WITH variation AS (
    SELECT
        medical_condition,
        test_results,
        CASE
                WHEN age <=25 THEN '18-25'
                WHEN age <=35 THEN '26-35'
                WHEN age <=45 THEN '36-45'
                WHEN age <=55 THEN '46-55'
                WHEN age <=65 THEN '56-65'
                WHEN age <=75 THEN '66-75'
                ELSE 'Above 75'
            END AS age_group,
```

```
            COUNT(*) as result_count
    FROM healthcare_data
    GROUP BY medical_condition, test_results, CASE
                WHEN age <=25 THEN '18-25'
                WHEN age <=35 THEN '26-35'
                WHEN age <=45 THEN '36-45'
                WHEN age <=55 THEN '46-55'
                WHEN age <=65 THEN '56-65'
                WHEN age <=75 THEN '66-75'
                ELSE 'Above 75'
            END
)
SELECT
    medical_condition,
    age_group,
    test_results,
    result_count,
    LEFT(
        ROUND(result_count * 100.0 / SUM(result_count) Over(PARTITION BY medical_condition, age_group),2),
    5) as percentage
FROM variation
GROUP BY medical_condition, age_group, test_results, result_count
ORDER BY medical_condition, age_group, test_results, result_count, percentage DESC


-- -----------------------------------------------------
-- ------------ Operational Analysis -------------------
-- What are the most common admission types?
SELECT admission_type, COUNT(*) as cases
FROM healthcare_data
GROUP BY admission_type;

-- Which hospitals and doctors handle the most patients?
SELECT
    hospital,
```

```sql
    doctor,
    COUNT(*) as total_patients
FROM healthcare_data
GROUP BY hospital, doctor
ORDER BY total_patients DESC;

-- ------------------------------------------------------
-- ------------ DATA QUALITY CHECKS --------------------
-- Identifying data anomalies
-- How long do patients typically stay in the hospital?
SELECT
    date_of_admission,
    discharge_date,
    DATEDIFF(DAY, date_of_admission, discharge_date) as length_of_stay,
    CASE
        WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 7 THEN '1 week'
        WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 14 THEN '2 weeks'
        WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 21 THEN '3 weeks'
        WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 28 THEN '4 weeks'
        ELSE '1 Month'
    END AS 'No_of_days'
FROM healthcare_data;

-- Total No of Patients with their length of stay
SELECT No_of_days, COUNT(*) as cases FROM
    (SELECT
        date_of_admission,
        discharge_date,
        DATEDIFF(DAY, date_of_admission, discharge_date) as length_of_stay,
        CASE
            WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 7 THEN '1 week'
            WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 14 THEN '2 weeks'
            WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 21 THEN '3 weeks'
            WHEN DATEDIFF(DAY, date_of_admission, discharge_date) <= 28 THEN '4 weeks'
            ELSE '1 Month'
```

```sql
        END AS 'No_of_days'
    FROM healthcare_data) AS t1
GROUP BY No_of_days
ORDER BY cases DESC;

-- Patients with high billing amounts
SELECT
    name,
    medical_condition,
    Hospital,
    doctor,
    ROUND(billing_amount, 2) as billing_amount
FROM healthcare_data
WHERE billing_amount > (SELECT AVG(billing_amount) FROM healthcare_data)
ORDER BY billing_amount DESC;

-- Trends in Test Result
SELECT
    medical_condition,
    LEFT(DATENAME(MONTH, date_of_admission),3) as monthly_trend,
    test_results,
    COUNT(*) as result_count,
    LEFT(ROUND(
        COUNT(*) * 100.0 / SUM(COUNT(*)) OVER(PARTITION BY medical_condition, DATENAME(MONTH, date_of_admission)),
    2),5) as percentage
FROM healthcare_data
GROUP BY medical_condition, DATENAME(MONTH, date_of_admission), test_results
ORDER BY medical_condition, DATENAME(MONTH, date_of_admission), result_count DESC;

-- ------------------------------------------------------
-- ------------ Financial AnalysisS --------------------
-- What is the average billing amount by medical condition?
SELECT
    medical_condition,
    ROUND(AVG(billing_amount),2) as average_billing_amount
```

```sql
FROM healthcare_data
GROUP BY medical_condition

-- Which insurance providers cover the most patients?
SELECT
    insurance_provider,
    COUNT(*) as total_patients
FROM healthcare_data
GROUP BY insurance_provider
ORDER BY total_patients DESC;

-- How does billing amount correlate with length of stay?
SELECT
    medical_condition,
    AVG(DATEDIFF(DAY, date_of_admission, discharge_date)) as average_length_of_stay,
    AVG(billing_amount) as average_billing_amount
FROM healthcare_data
GROUP BY medical_condition
ORDER BY average_billing_amount DESC;
```