HealthcareAnalytics

September 16, 2025

[1]:
```
# 1. Install required packages
# Run this once in your environment
# pip install kaggle pandas sqlite3

import os
import pandas as pd
import sqlite3
```

[8]:
```
#print(os.getcwd())
```

[1]:
```
# 3. Load CSV into pandas
#df = pd.read_csv('Hospital ER.csv')
```

[4]:
```
# 4. Load DataFrame to SQLite
conn = sqlite3.connect('HospitalData.db')
df.to_sql('HospitalData', conn, index=False, if_exists='replace')
```

[4]:
```
9216
```

[5]:
```
# 5. Example SQL Queries and Analysis
def query_and_print(sql):
    result = pd.read_sql(sql, conn)
    print(result, '\n')

analysis_queries = {
    'How are patients distributed among various departments based on referrals':
 ↪ '''
        SELECT "department_referral", COUNT(*) AS patient_count
        FROM HospitalData
        GROUP BY department_referral
        ORDER BY patient_count DESC;
    ''',
    'What is the monthly breakdown of patient visits?': '''
        SELECT strftime('%m', date) AS MonthNumber, strftime('%m-%Y', date) AS␣
 ↪MonthYear, COUNT(*) AS Visits
        FROM HospitalData
        GROUP BY MonthNumber, MonthYear
        ORDER BY Visits DESC;
```

```
    ''',

    'What is the gender breakdown of patients?': '''
        SELECT
            "patient_gender",
            COUNT(*) AS gender_count,
            ROUND((COUNT(*) * 100.0) / (SELECT COUNT(*) FROM HospitalData), 2)␣
↪AS percentage
        FROM
            HospitalData
        GROUP BY
            patient_gender;
    ''',

    'What is the average satisfaction score by age group and gender?': '''
        SELECT
            CASE
                WHEN "patient_age" BETWEEN 11 AND 20 THEN '11-20'
                WHEN "patient_age" BETWEEN 21 AND 30 THEN '21-30'
                WHEN "patient_age" BETWEEN 31 AND 40 THEN '31-40'
                WHEN "patient_age" > 50 THEN 'Above 50'
                ELSE 'Other'
            END AS age_group,
                patient_gender,
                AVG("patient_sat_score") AS avg_satisfaction_score
        FROM
            HospitalData
        GROUP BY
            age_group,
            patient_gender
        ORDER BY
            avg_satisfaction_score DESC;

    ''',
    'How do patient visits vary by age group?': '''
        SELECT
            CASE
                WHEN patient_age <= 10 THEN '0-10'
                WHEN patient_age <= 20 THEN '11-20'
                WHEN patient_age <= 30 THEN '21-30'
                WHEN patient_age <= 40 THEN '31-40'
                WHEN patient_age <= 50 THEN '41-50'
                ELSE 'Above 50'
            END AS age_group,
            COUNT(*) AS visit_count
        FROM
```

```
        HospitalData
    GROUP BY
        age_group
    ORDER BY
        visit_count DESC;

''',
'Average Wait Time?': '''
    SELECT AVG(patient_waittime) AS avg_patient_waittime
    FROM HospitalData

''',

'Total Patient Visits by Weekday': '''
SELECT
    strftime('%w', date) AS day_number,  -- day of week 0-6 (Sunday=0)
    CASE strftime('%w', date)
        WHEN '0' THEN 'Sunday'
        WHEN '1' THEN 'Monday'
        WHEN '2' THEN 'Tuesday'
        WHEN '3' THEN 'Wednesday'
        WHEN '4' THEN 'Thursday'
        WHEN '5' THEN 'Friday'
        WHEN '6' THEN 'Saturday'
    END AS days_,
    COUNT(patient_id) AS total_patient_visits
FROM
    HospitalData
GROUP BY
    day_number, days_
ORDER BY
    CAST(day_number AS INTEGER) ASC;

''',

'Patient visits by Race': '''

SELECT patient_race, COUNT(patient_id) AS Total_Patient_visits
FROM HospitalData
GROUP BY patient_race
ORDER BY COUNT(patient_id) DESC
''',

'Patient visits by Day Type': '''
SELECT
CASE
```

```
        WHEN strftime('%w', date) IN ('0', '6') THEN 'Weekend'  -- Sunday=0,␣
↪Saturday=6
        ELSE 'Weekday'
    END AS day_type,
    COUNT(patient_id) AS total_patient_visits
    FROM
        HospitalData
    GROUP BY
        day_type
    ORDER BY
        total_patient_visits DESC;
    ''',

    'Patient by category': '''
    SELECT
    CASE
        WHEN department_referral IS NULL OR LOWER(TRIM(department_referral)) IN␣
↪('none', '') THEN 'Walk-in'
        ELSE 'Referral'
    END AS patient_category,
    COUNT(*) AS patient_count
    FROM
        HospitalData
    GROUP BY
        patient_category
    ORDER BY
        patient_count DESC;

    '''
}
```

[6]:
```
for desc, sql in analysis_queries.items():
    print(f'-- {desc} --')
    query_and_print(sql)

conn.close()
```

```
-- How are patients distributed among various departments based on referrals --
  department_referral  patient_count
0                None           5400
1    General Practice           1840
2         Orthopedics            995
3       Physiotherapy            276
4          Cardiology            248
5           Neurology            193
6    Gastroenterology            178
7               Renal             86
```

```
-- What is the monthly breakdown of patient visits? --
   MonthNumber MonthYear  Visits
0           08   08-2020     530
1           05   05-2020     519
2           01   01-2020     513
3           03   03-2020     506
4           06   06-2019     506
5           08   08-2019     494
6           10   10-2019     493
7           12   12-2019     489
8           07   07-2020     488
9           06   06-2020     485
10          05   05-2019     480
11          04   04-2019     479
12          10   10-2020     471
13          04   04-2020     469
14          09   09-2019     469
15          09   09-2020     466
16          07   07-2019     464
17          11   11-2019     464
18          02   02-2020     431

-- What is the gender breakdown of patients? --
  patient_gender  gender_count  percentage
0              F          4487       48.69
1              M          4705       51.05
2             NC            24        0.26

-- What is the average satisfaction score by age group and gender? --
   age_group patient_gender  avg_satisfaction_score
0      11-20             NC                7.000000
1      21-30              F                5.282209
2      Other              M                5.187500
3      31-40              M                5.106918
4      Other              F                5.091803
5   Above 50              M                5.008333
6      11-20              F                4.925926
7      21-30              M                4.920000
8      31-40              F                4.880000
9      11-20              M                4.826923
10  Above 50              F                4.773364
11  Above 50             NC                3.000000
12     Other             NC                0.000000
13     21-30             NC                     NaN
14     31-40             NC                     NaN

-- How do patient visits vary by age group? --
  age_group  visit_count
```

```
0  Above 50         3345
1     21-30         1207
2     31-40         1191
3      0-10         1176
4     11-20         1160
5     41-50         1137

-- Average Wait Time? --
   avg_patient_waittime
0             35.259874

-- Total Patient Visits by Weekday --
  day_number      days_  total_patient_visits
0          0     Sunday                  1310
1          1     Monday                  1377
2          2    Tuesday                  1318
3          3  Wednesday                  1314
4          4   Thursday                  1305
5          5     Friday                  1260
6          6   Saturday                  1332

-- Patient visits by Race --
                    patient_race  Total_Patient_visits
0                          White                  2571
1                African American                  1951
2               Two or More Races                  1557
3                          Asian                  1060
4            Declined to Identify                  1030
5                Pacific Islander                   549
6  Native American/Alaska Native                   498

-- Patient visits by Day Type --
  day_type  total_patient_visits
0  Weekday                 6574
1  Weekend                 2642

-- Patient by category --
  patient_category  patient_count
0          Walk-in           5400
1         Referral           3816
```

Insights

1. In August, there was a peak in patient visits for the year, with 1024 registered patients, indicating a possible seasonal trend or event-driven demand in healthcare services during this month. In contrast, January saw the lowest with 431 visits, suggesting varying healthcare utilization patterns throughout the year.

2. The distribution of patient visits between AM and PM is almost equal, with 4632 visits in the morning and 4584 in the afternoon, demonstrating a steady and balanced demand for healthcare services throughout the day.
3. A significant number of patients (5400) were not referred to any specific department, illustrating a significant volume of general or non-specialized healthcare needs. Among specialized referrals, General Practice and Orthopedics were the most frequented, pointing to prevalent general health and musculoskeletal issues among the patient population.
4. There is a slight female predominance in patient visits, with females making up 51% and males 49%, reflecting a marginally higher healthcare utilization by women in the observed patient population.
5. The highest satisfaction scores are seen in the 21-30 age group, particularly among females, suggesting higher satisfaction with healthcare services in younger adults. Conversely, other age groups such as 31-40, 11-20, and those over 50 reported lower satisfaction, indicating potential areas for service improvement.
6. The over 50 age group had the highest number of visits, underscoring a greater healthcare requirement in this demographic. Conversely, the 0-10 age group had the fewest visits, highlighting lower healthcare engagement or need among younger children.

[ ]: