

DATA MINING AND DISCOVERY ASSIGNMENT

Empowering Healthcare: Transformative Solutions for Hospital Management

Name: Aswini Sivakumar

Student Id: 22027143

GitHub Link: <https://github.com/Aswinisiva/Data-mining-and-discovery-Assignment1>

Introduction:

In the digital age of healthcare, an efficient Hospital Management System (HMS) is indispensable for optimizing medical operations. Leveraging Python and SQLite, we present a dynamic solution tailored to streamline hospital administration. This system encompasses doctors, patients, and visits, meticulously organized to ensure seamless data integration and accessibility. By populating the database with randomized datasets, we simulate realistic healthcare scenarios, fostering informed decision-making and resource allocation. The robustness of SQLite coupled with Python's versatility enables scalable solutions adaptable to evolving healthcare landscapes. Our HMS promises to revolutionize hospital management, offering a comprehensive platform for enhancing patient care and operational efficiency.

Data Generation:

```
import sqlite3
import random
import datetime

# Connect to SQLite database
conn = sqlite3.connect('hospital.db')
cursor = conn.cursor()

# Drop tables if they exist
cursor.execute("DROP TABLE IF EXISTS Doctors")
cursor.execute("DROP TABLE IF EXISTS Patients")
cursor.execute("DROP TABLE IF EXISTS Visits")
```

This Python code establishes a connection to an SQLite database named 'hospital.db' and creates a cursor object for executing SQL commands. It then drops three tables (Doctors, Patients, and Visits) if they exist in the database, preparing for potential schema updates or recreations. The code is a part of an initial setup or database reset process.

Record 1: Doctor's Table

- The initial report provides a foundation for a healthcare-related database, allowing for the management of information about doctors, their specialties, and relevant details such as years of experience and hospital department affiliations.
- The Doctors table consists of seven columns:
- DOCTOR_ID, FIRST_NAME, LAST_NAME, SPECIALTY, EMAIL, YEARS_OF_EXPERIENCE, and HOSPITAL_DEPARTMENT.
- DOCTOR_ID is set as the primary key, ensuring uniqueness for each doctor record.
- Random data for doctors' first names, last names, specialties, emails, years of experience, and hospital departments have been generated.

Code:

```
# Function to generate random data for Doctors table
def generate_doctors(num_rows):
    # List of medical specialties
    specialties = ['Cardiology', 'Neurology', 'Orthopedics', 'Pediatrics', 'Oncology', 'Dermatology']

    # Loop to generate specified number of rows
    for _ in range(num_rows):
```

```

# Randomly choose first name from a list
first_name = random.choice(["John", "Jane", "Bob", "Alice", "Charlie"])

# Randomly choose last name from a list
last_name = random.choice(["Doe", "Smith", "Johnson", "Brown"])

# Randomly choose a medical specialty
specialty = random.choice(specialties)

# Generate an email using the chosen first and last names
email = f"{first_name.lower()}.{last_name.lower()}@example.com"

# Generate a random number for years of experience (1 to 30 years)
years_of_experience = random.randint(1, 30)

# Randomly choose a hospital department
(Note: 'departments' list is not defined in the provided code)
hospital_department = random.choice(departments)

# Execute SQL INSERT statement to add a new record to the Doctors table
cursor.execute("INSERT INTO Doctors (first_name, last_name, specialty, email,
years_of_experience, hospital_department) VALUES (?, ?, ?, ?, ?, ?)",
(first_name, last_name, specialty, email, years_of_experience, hospital_department))

```

Table: Doctors							
	DOCTOR_ID	FIRST_NAME	LAST_NAME	SPECIALTY	EMAIL	YEARS_OF_EXPERIENCE	HOSPITAL_DEPARTMENT
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	Bob	Brown	Orthopedics	bob.brown@example.com	4	Cardiology
2	2	Alice	Brown	Oncology	alice.brown@example.com	14	Cardiology
3	3	John	Doe	Neurology	john.doe@example.com	29	Neurology
4	4	John	Doe	Orthopedics	john.doe@example.com	20	Dermatology
5	5	Bob	Johnson	Neurology	bob.johnson@example.com	21	Neurology
6	6	Bob	Doe	Neurology	bob.doe@example.com	25	Dermatology
7	7	Alice	Johnson	Orthopedics	alice.johnson@example.com	17	Orthopedics
8	8	John	Brown	Cardiology	john.brown@example.com	10	Oncology
9	9	Alice	Doe	Cardiology	alice.doe@example.com	27	Orthopedics
10	10	Charlie	Brown	Orthopedics	charlie.brown@example.com	17	Dermatology
11	11	Charlie	Smith	Neurology	charlie.smith@example.com	5	Orthopedics
12	12	Jane	Brown	Oncology	jane.brown@example.com	3	Neurology
13	13	Alice	Smith	Orthopedics	alice.smith@example.com	29	Pediatrics
14	14	Charlie	Smith	Cardiology	charlie.smith@example.com	24	Orthopedics
15	15	John	Smith	Pediatrics	john.smith@example.com	7	Neurology
16	16	John	Johnson	Cardiology	john.johnson@example.com	30	Cardiology
17	17	John	Doe	Neurology	john.doe@example.com	18	Oncology
18	18	Jane	Johnson	Neurology	jane.johnson@example.com	23	Neurology
19	19	Jane	Doe	Cardiology	jane.doe@example.com	17	Cardiology
20	20	Charlie	Smith	Cardiology	charlie.smith@example.com	1	Orthopedics
21	21	Alice	Smith	Pediatrics	alice.smith@example.com	16	Oncology
22	22	Bob	Johnson	Pediatrics	bob.johnson@example.com	29	Neurology
23	23	Jane	Doe	Dermatology	jane.doe@example.com	15	Oncology

Record 2: Patient's Table

- The Patients table has been successfully created, incorporating essential patient information and establishing a foreign key relationship with the Doctors table

- The Patients table includes the following columns:
- PATIENT_ID, FIRST_NAME, LAST_NAME, AGE, GENDER, EMAIL, ADMIT_DATE, BLOOD_TYPE, and DOCTOR_ID.
- PATIENT_ID serves as the primary key for uniquely identifying each patient..
- Random data for patients' first names, last names, ages, genders, emails, admit dates, blood types, and associated doctor IDs have been generated.
- The FOREIGN KEY constraint on DOCTOR_ID ensures that each patient's doctor is associated with a valid doctor ID from the Doctors table.

Code:

```
import random
import datetime # Import the datetime module for working with dates

# Function to generate random data for Patients table
def generate_patients(num_rows):
    # List of blood types
    blood_types = ['A+', 'A-', 'B+', 'B-', 'AB+', 'AB-', 'O+', 'O-']

    # Loop to generate specified number of rows
    for _ in range(num_rows):
        # Randomly choose first name from a list
        first_name = random.choice(["John", "Jane", "Bob", "Alice", "Charlie"])

        # Randomly choose last name from a list
        last_name = random.choice(["Doe", "Smith", "Johnson", "Brown"])

        # Generate a random age between 1 and 100 years
        age = random.randint(1, 100)

        # Randomly choose gender
        gender = random.choice(['Male', 'Female'])

        # Generate an email using the chosen first and last names
        email = f"{first_name.lower()}.{last_name.lower()}@example.com"

        # Generate a random admission date within the past year
        admit_date = datetime.date.today() - datetime.timedelta(days=random.randint(1, 365))

        # Randomly choose a blood type
        blood_type = random.choice(blood_types)

        # Randomly assign a doctor ID (assuming there are 20 doctors)
        doctor_id = random.randint(1, 20)

        # Execute SQL INSERT statement to add a new record to the Patients table
        cursor.execute("INSERT INTO Patients (first_name, last_name, age, gender, email,
            admit_date, blood_type, doctor_id) VALUES (?, ?, ?, ?, ?, ?, ?, ?)",
            (first_name, last_name, age, gender, email, admit_date, blood_type, doctor_id))
```

Table: Patients Filter in any column									
	PATIENT_ID ▼	FIRST_NAME	LAST_NAME	AGE	GENDER	EMAIL	ADMIT_DATE	BLOOD_TYPE	DOCTOR_ID
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	Charlie	Smith	96	Male	charlie.smith@example.com	2023-04-19	B+	10
2	2	Jane	Doe	8	Female	jane.doe@example.com	2024-03-01	B+	8
3	3	Charlie	Johnson	19	Male	charlie.johnson@example.com	2023-10-27	B+	5
4	4	Alice	Doe	64	Male	alice.doe@example.com	2024-01-01	B-	15
5	5	Charlie	Brown	10	Male	charlie.brown@example.com	2024-02-23	B+	12
6	6	John	Doe	99	Male	john.doe@example.com	2023-12-22	A-	14
7	7	Alice	Smith	81	Female	alice.smith@example.com	2024-01-01	AB+	7
8	8	Alice	Smith	86	Female	alice.smith@example.com	2023-09-27	O+	4
9	9	Bob	Brown	49	Female	bob.brown@example.com	2023-06-12	AB-	16
10	10	John	Brown	57	Female	john.brown@example.com	2023-10-13	AB-	10
11	11	Charlie	Doe	70	Female	charlie.doe@example.com	2023-12-11	O+	19
12	12	Alice	Johnson	59	Female	alice.johnson@example.com	2023-03-17	A+	19
13	13	Jane	Johnson	35	Female	jane.johnson@example.com	2024-02-26	A-	17
14	14	Charlie	Smith	90	Female	charlie.smith@example.com	2023-10-25	AB+	5
15	15	John	Smith	70	Female	john.smith@example.com	2023-06-07	O+	16
16	16	Jane	Johnson	23	Male	jane.johnson@example.com	2023-03-12	AB-	9
17	17	Jane	Brown	32	Female	jane.brown@example.com	2023-03-25	O-	15
18	18	Bob	Brown	59	Female	bob.brown@example.com	2023-09-08	AB+	12
19	19	Alice	Johnson	74	Female	alice.johnson@example.com	2023-05-15	B+	11
20	20	Jane	Johnson	87	Male	jane.johnson@example.com	2023-06-30	A-	11
21	21	Jane	Brown	38	Female	jane.brown@example.com	2023-07-25	O-	6
22	22	Charlie	Doe	41	Male	charlie.doe@example.com	2023-04-07	B+	5
23	23	John	Johnson	58	Male	john.johnson@example.com	2023-06-05	B+	8

1 - 24 of 1000 Go to: 1

Record 3: Visit History

- The Visits table has been successfully created with additional columns, including PATIENT_ID to establish a foreign key relationship with the Patients table
- The Visits table now includes columns:
- VISIT_ID, VISIT_DATE, PATIENT_ID, DOCTOR_ID, DIAGNOSIS, and TREATMENT_COST.
- VISIT_ID serves as the primary key for uniquely identifying each visit.
- Random data for visit dates, patient IDs, doctor IDs, diagnoses, and treatment costs have been generated.
- The FOREIGN KEY constraints on PATIENT_ID and DOCTOR_ID ensure that each visit is associated with valid patient and doctor IDs.

Code:

```
# Function to generate random data for Visits table
def generate_visits(num_rows):
    # List of single-word diagnoses
    single_word_diagnoses = ['Fever', 'Injury', 'Cold', 'Allergy', 'Hypertension', 'Migraine', 'Flu', 'Sprain', 'Asthma']

    # Loop to generate specified number of rows
    for _ in range(num_rows):
        # Generate a random visit date within the past year
        visit_date = datetime.date.today() - datetime.timedelta(days=random.randint(1, 365))

        # Randomly assign a patient ID (assuming there are 100 patients)
```

```

patient_id = random.randint(1, 100)

# Randomly assign a doctor ID (assuming there are 20 doctors)
doctor_id = random.randint(1, 20)

# Randomly choose a diagnosis from the list
diagnosis = random.choice(single_word_diagnoses)

# Generate a random treatment cost between $50 and $5000
treatment_cost = round(random.uniform(50, 5000), 2)

# Execute SQL INSERT statement to add a new record to the Visits table
cursor.execute("INSERT INTO Visits (visit_date, patient_id, doctor_id, diagnosis,
treatment_cost) VALUES (?, ?, ?, ?, ?)",
              (visit_date, patient_id, doctor_id, diagnosis, treatment_cost))
              (first_name, last_name, specialty, email, years_of_experience, hospital_department))

```

Table: Visits						
	VISIT_ID	VISIT_DATE	PATIENT_ID	DOCTOR_ID	DIAGNOSIS	TREATMENT_COST
	Filter	Filter	Filter	Filter	Filter	Filter
1	1	2023-08-02	20	15	Allergy	1696.71
2	2	2023-09-16	6	3	Flu	4384.88
3	3	2023-12-06	55	5	Cold	4434.15
4	4	2023-08-16	44	19	Cold	1358.95
5	5	2023-04-18	66	10	Allergy	534.27
6	6	2024-01-11	47	13	Asthma	633.89
7	7	2023-07-18	76	9	Sprain	470.93
8	8	2024-01-11	42	18	Asthma	927.74
9	9	2023-06-21	92	3	Flu	3187.61
10	10	2023-08-20	63	1	Fever	3240.01
11	11	2023-06-13	25	12	Fever	1005.46
12	12	2023-11-15	100	19	Cold	2686.36
13	13	2024-02-27	60	6	Asthma	722.57
14	14	2023-10-12	11	5	Sprain	4862.57
15	15	2023-05-11	14	10	Asthma	4207.38
16	16	2023-08-23	23	5	Allergy	4636.26
17	17	2023-04-20	28	15	Asthma	3952.9
18	18	2023-03-08	79	12	Allergy	4781.33
19	19	2023-12-02	95	4	Injury	4523.96
20	20	2023-11-03	92	5	Hypertension	1634.64
21	21	2023-06-08	25	4	Sprain	881.44
22	22	2023-10-12	54	13	Asthma	2055.2
23	23	2024-01-15	63	1	Allergy	1493.15

Schema:

Tables (3)

Name	Type	Schema
Doctors		CREATE TABLE Doctors (DOCTOR_ID INTEGER PRIMARY KEY, FIRST_NAME TEXT NOT NULL, LAST_NAME TEXT NOT NULL, SPECIALTY TEXT, EMAIL TEXT, YEARS_OF_EXPERIENCE INTEGER, HOSPITAL_DEPARTMENT TEXT)
DOCTOR_ID	INTEGER	"DOCTOR_ID" INTEGER
FIRST_NAME	TEXT	"FIRST_NAME" TEXT NOT NULL
LAST_NAME	TEXT	"LAST_NAME" TEXT NOT NULL
SPECIALTY	TEXT	"SPECIALTY" TEXT
EMAIL	TEXT	"EMAIL" TEXT
YEARS_OF_EXPERIENCE	INTEGER	"YEARS_OF_EXPERIENCE" INTEGER
HOSPITAL_DEPARTMENT	TEXT	"HOSPITAL_DEPARTMENT" TEXT
Patients		CREATE TABLE Patients (PATIENT_ID INTEGER PRIMARY KEY, FIRST_NAME TEXT NOT NULL, LAST_NAME TEXT NOT NULL, AGE INTEGER, GENDER TEXT, EMAIL TEXT, ADMIT_DATE DATE, BLOOD_TYPE TEXT, DOCTOR_ID INTEGER, FOREIGN KEY (DOCTOR_ID) REFERENCES Doctors(DOCTOR_ID))
PATIENT_ID	INTEGER	"PATIENT_ID" INTEGER
FIRST_NAME	TEXT	"FIRST_NAME" TEXT NOT NULL
LAST_NAME	TEXT	"LAST_NAME" TEXT NOT NULL
AGE	INTEGER	"AGE" INTEGER
GENDER	TEXT	"GENDER" TEXT
EMAIL	TEXT	"EMAIL" TEXT
ADMIT_DATE	DATE	"ADMIT_DATE" DATE
BLOOD_TYPE	TEXT	"BLOOD_TYPE" TEXT
DOCTOR_ID	INTEGER	"DOCTOR_ID" INTEGER
Visits		CREATE TABLE Visits (VISIT_ID INTEGER PRIMARY KEY, VISIT_DATE DATE, PATIENT_ID INTEGER, DOCTOR_ID INTEGER, DIAGNOSIS TEXT, TREATMENT_COST REAL, FOREIGN KEY (PATIENT_ID) REFERENCES Patients(PATIENT_ID), FOREIGN KEY (DOCTOR_ID) REFERENCES Doctors(DOCTOR_ID))
VISIT_ID	INTEGER	"VISIT_ID" INTEGER
VISIT_DATE	DATE	"VISIT_DATE" DATE
PATIENT_ID	INTEGER	"PATIENT_ID" INTEGER
DOCTOR_ID	INTEGER	"DOCTOR_ID" INTEGER
DIAGNOSIS	TEXT	"DIAGNOSIS" TEXT
TREATMENT_COST	REAL	"TREATMENT_COST" REAL

Explanation and Justification:

Doctor's Table

- **DOCTOR_ID:** This is a unique identifier for each doctor within the system. It serves as the primary key, ensuring that each doctor has a distinct identification number.
- **FIRST_NAME:** The first name of the doctor is stored as text. The "NOT NULL" constraint indicates that this information must be provided for every doctor entry, meaning it cannot be left blank.
- **LAST_NAME:** Similar to the first name, the last name of the doctor is stored as text, and it cannot be left blank.
- **SPECIALTY:** This column represents the medical specialty or field of expertise of the doctor. It can include values like 'Cardiology,' 'Neurology,' etc.
- **EMAIL:** The email address of the doctor is stored in this column. It provides a means of communication and contact for the hospital staff.
- **YEARS_OF_EXPERIENCE:** This column stores the number of years of experience the doctor has. It helps in assessing the level of expertise of the medical professionals.
- **HOSPITAL_DEPARTMENT:** Indicates the specific department within the hospital where the doctor works. Examples include 'Cardiology Department' or 'Orthopedics Department.'

Patient's Table:

- **PATIENT_ID**: Similar to the doctor's table, this column serves as a unique identifier for each patient and is the primary key of the table.
- **FIRST_NAME**: The first name of the patient, stored as text. The "NOT NULL" constraint ensures that this information is mandatory for each patient record.
- **LAST_NAME**: The last name of the patient, stored as text, and must be provided for every patient entry.
- **AGE**: Represents the age of the patient. It is stored as an integer value.
- **GENDER**: Indicates the gender of the patient, such as 'Male' or 'Female.'
- **EMAIL**: Stores the email address of the patient.
- **ADMIT_DATE**: This column records the date when the patient was admitted to the hospital. It is stored as a date data type.
- **BLOOD_TYPE**: Represents the blood type of the patient, such as 'A+', 'B-', etc.
- **DOCTOR_ID**: This is a foreign key that establishes a link between the Patients table and the Doctors table. It references the DOCTOR_ID column in the Doctors table, indicating the doctor responsible for the patient.

Visitor's Table:

- **VISIT_ID**: A unique identifier for each visit, serving as the primary key.
- **VISIT_DATE**: Records the date when the visit occurred.
- **PATIENT_ID**: A foreign key linking the visit to a specific patient by referencing the column in the Patients table.
- **DOCTOR_ID**: Another foreign key linking the visit to a specific doctor by referencing the column in the Doctors table.
- **DIAGNOSIS**: Stores the diagnosis made during the visit, providing information about the patient's health condition.
- **TREATMENT_COST**: Represents the cost of the treatment provided during the visit. It is stored as a real number, allowing for decimal values.

The Doctors, Patients, and Visits tables in the hospital management system database collectively serve as a comprehensive and organized framework for efficiently managing healthcare data. The Doctors table centralizes information about medical professionals, enabling categorization by specialty and department, tracking years of experience, and facilitating effective assignment to patients. The Patients table manages detailed patient records, including admission information, demographics, and associations with assigned doctors. The Visits table tracks patient interactions, capturing visit history, diagnoses, treatment costs, and establishing links to specific patients and doctors. The foreign key relationships ensure data integrity and enable seamless retrieval of information, contributing to efficient healthcare administration, comprehensive patient care, and informed decision-making. Together, these tables form a robust foundation for a holistic hospital management system.

Query Examples:

Query 1:

```
SELECT
    Patients.first_name || ' ' || Patients.last_name AS patient_name,
    Doctors.first_name || ' ' || Doctors.last_name AS doctor_name,
    Doctors.specialty
FROM
    Patients
JOIN
    Visits ON Patients.patient_id = Visits.patient_id
```



```

JOIN
  Doctors ON Visits.doctor_id = Doctors.doctor_id
WHERE
  Visits.diagnosis = 'Fever';
LIMIT 5;

```

patient_name	doctor_name	SPECIALTY
John Smith	Bob Brown	Orthopedics
Jane Brown	Jane Brown	Oncology
John Doe	John Doe	Neurology
Jane Smith	John Doe	Neurology
Jane Johnson	Charlie Smith	Cardiology

Explanation: This SQL query retrieves the names of patients and their treating doctors, along with the doctors' specialties, for visits with a diagnosis of 'Fever' from the hospital database. It involves joining the Patients, Visits, and Doctors tables based on patient and doctor IDs, filtering by the diagnosis, and limits the result to the top 5 records.

Query 2:

```

SELECT
  Patients.first_name || ' ' || Patients.last_name AS patient_name,
  SUM(Visits.treatment_cost) AS total_cost
FROM
  Patients
JOIN
  Visits ON Patients.patient_id = Visits.patient_id
WHERE
  Patients.age BETWEEN 20 AND 30
GROUP BY
  Patients.patient_id;

```

patient_name	total_cost
Jane Johnson	14003.84
Jane Smith	31398.87
Jane Brown	18110.94
John Johnson	7552.82
Jane Smith	26898.82
John Johnson	19422.61
Alice Smith	38366.11
Alice Doe	18628.82
Alice Brown	26089.81
Charlie Johnson	34973.28
Charlie Johnson	14900.18
Charlie Johnson	26302.03

Explanation: The query retrieves the concatenated names of patients and the total treatment cost for each patient aged between 20 and 30. The results are grouped by patient ID, and the sum of treatment costs is calculated for each patient.

Query 3:

```
SELECT
    Doctors.first_name || ' ' || Doctors.last_name AS doctor_name,
    Doctors.years_of_experience,
    Doctors.hospital_department
FROM
    Doctors
WHERE
    Doctors.years_of_experience > 27
    AND Doctors.hospital_department = 'Cardiology';
```

doctor_name	YEARS_OF_EXPERIENCE	HOSPITAL_DEPARTMENT
John Johnson	30	Cardiology
Alice Johnson	29	Cardiology
Charlie Brown	28	Cardiology
Jane Smith	28	Cardiology
John Doe	29	Cardiology
Charlie Johnson	28	Cardiology
John Johnson	30	Cardiology
John Smith	29	Cardiology
Bob Johnson	29	Cardiology
Alice Smith	30	Cardiology
Alice Doe	28	Cardiology
Charlie Smith	30	Cardiology
Alice Brown	30	Cardiology
John Brown	28	Cardiology
Alice Smith	30	Cardiology

Explanation: This SQL query selects the names, years of experience, and hospital department of doctors specializing in Cardiology with more than 27 years of experience from the hospital database. The WHERE clause filters based on the specified conditions.

Query 4:

```
SELECT
    blood_type,
    AVG(age) AS average_age
FROM
    Patients
GROUP BY
    blood_type;
```

BLOOD_TYPE	average_age
A+	53.2105263157895
A-	51.8947368421053
AB+	47.4180327868853
AB-	51.9645390070922
B+	47.8090909090909
B-	53.140625
O+	54.4864864864865
O-	52.1914893617021

Explanation: This SQL query calculates the average age for each blood type category in the Patients table, grouping the results by blood type. The SELECT statement retrieves blood types along with their corresponding average ages.

Query 5:

```
SELECT
    Patients.patient_id,
    Patients.first_name || ' ' || Patients.last_name AS patient_name,
    COUNT(Visits.visit_id) AS visit_count
FROM
    Patients
LEFT JOIN
    Visits ON Patients.patient_id = Visits.patient_id
GROUP BY
    Patients.patient_id, patient_name
ORDER BY
    visit_count DESC
LIMIT 5;
```

PATIENT_ID	patient_name	visit_count
73	John Johnson	19
10	John Brown	17
27	Jane Smith	17
46	Bob Brown	17
65	Jane Brown	17

Explanation: This SQL query retrieves the patient ID, concatenated patient name, and the count of visits for each patient, ordering the results by visit count in descending order. It uses a LEFT JOIN to include patients without visits and limits the output to the top 5 records.

Conclusion:

Our Python-based hospital management system, integrated with SQLite, redefines operational standards. By optimizing data management and fostering adaptability, it empowers healthcare professionals with actionable insights. With scalability and reliability at its core, it heralds a new era of streamlined operations and elevated patient care.